American Society of Human Genetics 66th Annual Meeting
October 18–22, 2016 VANCOUVER, CANADA

POSTER ABSTRACTS

Session Topic/Title	Abstract/Poster Board Numbers	
	Start #	End #
Statistical Genetics and Genetic Epidemiology	328W	591F
Cardiovascular Phenotypes	592W	723F
Molecular and Cytogenetic Diagnostics	724W	1032F
Evolution and Population Genetics	1033W	1178T
Complex Traits and Polygenic Disorders	1179F	1649T
Bioinformatics and Computational Approaches	1650F	1956F
Epigenetics and Gene Regulation	1957W	2157F
Developmental Genetics and Gene Function	2158W	2254W
Mendelian Phenotypes	2255T	2661F
Cancer Genetics	2662W	2989W
Genetic Counseling, ELSI, Education, and Health Services Research	2990T	3133W
Omics Technologies	3134T	3227T
Prenatal, Perinatal, and Reproductive Genetics	3228F	3303F
Genome Structure and Function	3304W	3397W

Posters should remain on the board for all three days (Wednesday through Friday)

The program and abstract/poster board number next to each listing is followed by an **W** (Wednesday), **T** (Thursday), or **F** (Friday) to indicate the day on which authors must be present at their poster boards.

328W

Genetic determinants of glycated hemoglobin levels in the Greenlandic Inuit population. *E.V.R. Appel[1], I. Moltke[2], M.E. Jørgensen[3], P. Bjerregaard[4,5], A. Linneberg[6,7,8], O. Pedersen[1], A. Albrechtsen2[2], T. Hansen[1], N. Grarup[1].* 1) NNF-CBMR: Section for Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Copenhagen Ø, Denmark; 2) The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark; 3) Steno Diabetes Center, 2820 Gentofte, Denmark; 4) National Institute of Public Health, University of Southern Denmark, 1353 Copenhagen, Denmark; 5) Greenland Centre for Health Research, University of Greenland; 6) Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark; 7) Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark; 8) Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

We recently showed that a common genetic variant leads to a remarkably increased risk of type 2 diabetes (T2D) in the small and historically isolated Greenlandic population. Motivated by this, we aimed at discovering novel genetic determinants for glycated hemoglobin (HbA$_{1c}$) and at estimating the effect of known HbA$_{1c}$ associated loci in the Greenlandic population. We analyzed genotype data from 4,049 Greenlanders, which have previously been generated using the Illumina Cardio-Metabochip. We performed the discovery association analysis by applying an additive linear mixed model. To estimate the effect of known HbA$_{1c}$ associated loci we modeled the effect in the European and Inuit ancestry proportions of the Greenlandic genome (EPG and IPG). We found no novel significant associations after correcting for multiple testing. When we investigated loci known to associate with HbA$_{1c}$ levels, we found that the lead variant in the *GCK* locus associated significantly with HbA$_{1c}$ levels in the IPG (P$_{IPG}$=4.8·10^{-6}, β$_{IPG}$=0.13 SD). Furthermore, for 10 of 15 known HbA$_{1c}$ loci the effect was similar to previously reported effects. Interestingly, the *ANK1* locus showed a significant ancestral population specific effect, with opposing directions of effect in each ancestral population. We found no evidence of additional independent association signals in the known HbA$_{1c}$ associated loci. In conclusion, we found only 1 of the 15 known HbA$_{1c}$ loci to significant associate within the IPG and that 10 of 15 loci showed similar effects as found in European and East Asian populations. Our results shed light on the genetic effects across ethnicities.

329T

The Type 2 Diabetes Knowledge Portal: A paradigm for the democratization of human genetic information. *M.C. Costanzo, Accelerating Medicines Partnership for Type 2 Diabetes Knowledge Portal.* Medical and Population Genetics, Broad Institute, Cambridge, MA.

Discovery of the association of human genetic variants with diseases and traits can spur novel insights in health and medicine. But most association data are warehoused in disparate silos and are difficult to access. Even in a specific disease area, results from different consortia are often not integrated with each other. Thus it can be prohibitively difficult for researchers even to access the data, let alone use them to generate hypotheses that could catalyze the clinical translation of genetic discoveries. The Accelerating Medicines Partnership in Type 2 Diabetes (AMP T2D) addresses this problem by forging collaboration between the National Institutes of Health and academic, industrial, and nonprofit stakeholders in order to produce and aggregate data, develop analysis methods and tools, and build infrastructure for data storage and presentation. The results of this collaboration are available via the Type 2 Diabetes Knowledge Portal (http://www.type2diabetesgenetics.org), an open-access resource that aggregates genetic and phenotypic data in a framework that facilitates dynamic analysis across disparate data sets while protecting patient privacy. The Portal knowledgebase contains T2D association studies from all the large-scale T2D consortia and projects, covering more than 260,000 individuals, along with harmonized genome-wide association study (GWAS) meta-analysis data from over 600,000 samples for an additional 25 relevant traits. We have recently integrated variant data from 1000 Genomes, 2,500 whole genomes from GoT2D, and 60,000 exome sequences from the Exome Aggregation Consortium. By the end of 2016 we will release results for an additional 9,000 exomes and GWAS data for over 30,000 T2D samples: 4,000 multiethnic samples from a clinical pre-diabetic study (CAMP); 14,000 African-American samples from BioMe; and 11,800 Finnish samples from METSIM and FUSION. Portal data are currently stored in a Data Coordinating Center (DCC) at the Broad Institute. Soon, they will also be stored at connected federated nodes, including the European Bioinformatics Institute, enabling each node to protect patient data in accordance with local regulations while facilitating global access to analyses. To spark insights from these data, we have developed tools for custom analysis of variant associations. Software and methods developed within the AMP T2D project will be publicly available and are generally applicable to variant associations with any disease or trait.

330F

Genotype-by-multivariate dietary environment interaction effects on cardiometabolic risk factors in Mexican American children: The SAFARI study. *V.P. Diego[1], J. Vanamala[3,4], V.S. Farook[2], L. Reddivari[5], S.P. Fowler[6], S. Puppala[7], G. Chittor[8], R.G. Resendez[2], J.C. Alvarenga[2], S. Mummidi[2], L. Almasy[1], J.E. Curran[1], A.G. Comuzzie[7], D.M. Lehman[6,9], C.P. Jenkinson[2], J.L. Lynch[10], R.A. DeFronzo[6], J. Blangero[1], D.E. Hale[10], R. Duggirala[2], R. Arya[2].* 1) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX; 2) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Edinburg, TX; 3) Department of Food Science, Penn State University, University Park, PA; 4) The Pennsylvania State Hershey Cancer Institute, Penn State Milton S Hershey Medical Center, Hershey, PA; 5) Department of Plant Science, Penn State University, University Park, PA; 6) Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX; 7) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 8) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC; 9) Department of Cellular & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX; 10) Department of Pediatrics, University of Texas Health Science Center at San Antonio, San Antonio, TX.

   Obesity, type 2 diabetes (T2D), and the metabolic syndrome (MetSyn)—defined as a cluster of obesity, pre-diabetes, and other cardiometabolic risk factors (CMRFs)—are worldwide epidemics, and are increasingly prevalent in children of minority ethnicity. As part of the San Antonio Family Assessment of Metabolic Risk Indicators in Youth (SAFARI) Study on healthy Mexican American children, we studied genotype-by-environment interaction (GEI) effects on CMRF variation in 216 boys and 208 girls with a mean age of 11.6 years. We studied GEI in relation to dietary multivariate environments (MEs), defined as scores from the first 4 factors of a principal component analysis (PCA) of 15 dietary variables followed by varimax rotation. The 15 dietary variables consist of 13 survey variables—intake of carbohydrate; protein; total, trans-, saturated, monounsaturated, and polyunsaturated fats; cholesterol; sugar; fiber; ω-3 and ω-6 fatty acids; and total kilocalories—and serum α- and β-carotene, measured using an ultra pressure liquid chromatography-photo diode array detector. The first 4 factors, denoted by F1, F2, F3, and F4, explain > 90% of the total system variance. By likelihood ratio tests, we tested the two null hypotheses of GEI effects, namely additive genetic variance homogeneity across a given ME (VH) and an across-ME genetic correlation equal to 1 (GC). Our most interesting results are from analysis of the F2 and F4 MEs, where F2 had significant loadings by protein (0.71), saturated (0.56) and monounsaturated fat (-0.36) intake, and F4 had significant loadings for carbohydrate (0.71), cholesterol (-0.42) and total fat (-0.38) intake. A loading is significant if its absolute value is $\geq$ 0.3. For the F2 ME, we found significant GEI effects for C-peptide (p=0.04 and p<0.01 respectively from the VH and GC tests), and for fasting serum insulin (FSI) (p=0.01), homeostasis model assessment 2 of insulin resistance (p=0.01), HDL cholesterol (p=0.01), % fat mass (%FM) (p<0.001), and waist circumference (p=0.02) from the GC test. For the F4 ME, we found significant GEI effects for C-reactive protein (p=0.03) from the VH test, and for FSI (p<0.001), %FM (p=0.01), and a MetSyn variable (p=0.01) from the GC test. Our results show that dietary variables play critical roles in influencing the genetic determinants of obesity, T2D, and the MetSyn, and are novel in that they show an improved efficacy in detecting GEI effects by multivariate modeling of the environment.

331W

Type 2 diabetes and risk of stroke subtypes: A Mendelian randomization analysis based on individual participant data of 105,000 Chinese adults from the China Kadoorie Biobank. *W. Gan[1,2], M. Holmes[3], R. Walters[3], I. Millwood[3], F. Bragg[3], Y. Guo[4], J. Vaucher[3], Z. Bian[4], A. Mahaja[1,2], R. Clarke[3], L. Lee[4,5], M. McCarthy[1,2,6], Z. Chen[3].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology, and Metabolism (OCDEM), University of Oxford, Oxford, United Kingdom; 3) Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU),Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom; 4) Chinese Academy of Medical Sciences, Dong Cheng District, Beijing, China; 5) School of Public Health, Peking University Health Sciences Center, Beijing, China; 6) National Institute of Health Research, Oxford Biomedical Research Centre, Oxford, United Kingdom.

   Background: Observational studies report that type 2 diabetes (T2D) is associated with increased risk of both ischaemic and haemorrhagic stroke, but the underlying causal relationships are unclear. Given the increasing prevalence of T2D, clarifying the nature of these relationships could have important repercussions for stroke incidence and highlight potential opportunities for prevention. **Method:** We developed genetic risk scores (GRSs) based on 49 established T2D risk variants, weighting individual genetic variants using the most recent effect size estimates (Gan et al). We used data on 105,696 participants from the population-based China Kadoorie Biobank (CKB) prospective study, 93,119 randomly-selected subjects supplemented with 12,580 from a nested stroke case-control study (n= 8,284 combined prevalent/incident T2D cases and 12,122 stroke cases). We conducted Mendelian randomization (MR) analyses of individual participant data, using the GRS as an instrumental variable to investigate the causal role of T2D for overall stroke and stroke subtypes. Both observational and MR analyses were adjusted for age, sex, geographic region, BMI, smoking, alcohol intake, educational level and household income. **Results:** In observational analyses of the CKB study (512,891 participants), individuals with T2D had higher risks of total stroke (OR [95% CI] = 1.78 [1.72 - 1.84]), ischaemic stroke 1.88 (1.81 - 1.95) and haemorrhagic stroke 1.21 (1.10 - 1.32). In MR analysis, genetically-instrumented T2D was causally associated with an increased risk of stroke (OR = 1.81 [1.25 - 2.46], P = 3.31 × 10^{-3}), mainly with ischaemic stroke (n=6,859 cases; OR = 2.20 [1.50-3.04], P = 3.56 × 10^{-4}) but not haemorrhagic stroke (n=4,733 cases; OR = 1.00 [0.24 - 2.0], P = 0.40). For both stroke subtypes, the observational and causal estimates were similar ($P_{heterogeneity}$ > 0.46). No differences in risks of total or ischaemic stroke were identified when restricting the GRS to T2D genetic variants related primarily to either beta cell function or insulin sensitivity ($P_{heterogeneity}$ > 0.33). **Conclusions:** This MR study supports the causal relationship between T2D and risk of ischaemic stroke and highlights the potential importance of targeting diabetes for prevention of stroke.

332T

Genome-wide interaction with the insulin secretion locus *MTNR1B* reveals *CMIP*, a novel type 2 diabetes susceptibility gene in African Americans. *J.M. Keaton[1,2,3], C. Gao[1,2,3], J.N. Hellwege[2,3], N.D. Palmer[2,3,4,5], J.S. Pankow[6], M. Fornage[7], J.G. Wilson[8], A. Correa[8], L.J. Rasmussen-Torvik[9], J.I. Rotter[10], Y.D.I. Chen[10], K.D. Taylor[10], S.S. Rich[11], L.E. Wagenknecht[5,12], B.I. Freedman[3,5,13], M.C.Y. Ng[2,3], D.W. Bowden[2,3,4].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 6) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 7) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 8) University of Mississippi Medical Center, Jackson, MS; 9) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 10) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA; 11) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 12) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 13) Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC.

Type 2 diabetes (T2D) is the result of metabolic defects in insulin secretion and insulin sensitivity, yet most of the T2D loci identified to date are related to insulin secretion. We hypothesize that T2D loci, particularly those impacting insulin sensitivity, can be identified through interaction with insulin secretion loci. To test this hypothesis, an intronic SNP, rs10830963, in *MTNR1B* associated with acute insulin response to glucose (AIR_g; P=1.2x10^{-5}), a dynamic measure of first-phase insulin secretion, was identified in African Americans (AfA) from the Insulin Resistance Atherosclerosis Family Study (IRASFS). rs10830963 was tested for interaction and joint (main + interaction) effects in AfA from five cohorts (ARIC, CARDIA, JHS, MESA, WFSM) comprising 2,452 cases and 3,772 controls. Genome-wide genotype data was derived from the Affymetrix HumanGenome 6.0 array imputed to a 1000 Genomes reference panel. Logistic regression with T2D as the outcome was modeled including rs10830963 dosage, age, sex, and principal component covariates. Joint effects were captured using the Kraft 2 degree of freedom test. Genome-wide significant (P<5x10^{-8}) interaction and joint effects were detected for the intronic SNP rs17197883 in *CMIP* ($P_{interaction}$=1.43x10^{-8}; P_{joint}=4.70x10^{-8}). *CMIP* variants have previously been nominally associated with T2D, fasting glucose, and adiponectin in individuals of East Asian (EAS) ancestry, and with waist-to-hip ratio adjusted for body mass index (BMI) in Europeans (EUR). In follow up cross-ethnic analysis, 54 SNPs in linkage disequilibrium (r^2>0.3) with 7 CMIP SNPs from the AfA interaction analysis (P_{joint}<10^{-6}) were tested for association with AIRg and BMI-adjusted insulin sensitivity (SI) in IRASFS Hispanic subjects (n=1,016). One SNP (rs880795) was nominally associated with SI (P=0.02), while none were associated with AIR_g. These data support the hypothesis that additional genetic factors contributing to T2D risk, including insulin sensitivity loci, can be identified by interactions with insulin secretion loci.

333F

Deep whole genome sequencing empowers and refines traditional eQTL studies. *A. Liu[1], H.M. Kang[1], M. Kretzler[1], V. Nair[1], R. Nelson[2].* 1) Biostatistics, University of Michigan, 1415 Washington Heights, ANN ARBOR, MI; 2) NIDDK, Phoenix Indian Medical Center, 4212 North 16th Street, Phoenix, AZ 85016.

Traditional expression quantitative trait loci (eQTL) studies based on microarrays have successfully identified tremendous numbers of cis-acting associations between genetic variants and expression levels. However, because microarrays estimate gene expression levels from a limited number of exon-overlapping probes, it is well known that reference bias in probe sequences containing genetic variants create false positive cis-eQTLs. However, it is impossible to distinguish such false positive cis-eQTLs completely without knowing all genetic variants polymorphic in the samples. This has been a challenge in our eQTL studies of Pima diabetic nephropathy cohorts with renal expression data from glomerular and tubular tissues collected at two time-points over 8 years. Because Pima Native Americans are genetically distant from populations with European and African ancestry, there are many population-specific variants that would not have been discovered without deep genome sequencing. Recently we deeply sequenced 97 subjects to perform a comprehensive cis-eQTL association analysis across nearly all genetic variants. Here we demonstrate that the availability of deep genome sequence data empowers and refines existing eQTL studies, allowing for correction of reference-bias in variant-overlapping probes. With over 10 million variants genotyped from 97 deeply sequenced genomes, we comprehensively identified a subset of probes overlapping with variants that are polymorphic in the Pima study cohort, and re-normalized the expression data excluding 26,743 probes across 12,970 genes and performed eQTL studies. Surprisingly, we observed that even recent Affymetrix Expression (HuGene 2.1 ST) array platforms have substantial overlap (12.4%) between probe sequences and variants polymorphic in the Pima cohort which is more than the overlap (6.1%/7.7%) of earlier-generation arrays (HG-U133A/HG-U133 Plus 2.0). When performing eQTL studies excluding these probes, we were able to eliminate 63 potentially spurious cis-eQTL genes affected by variants-in-probe effect, confirmed by heavily imbalanced direction of effect sizes consistent to reference bias. Moreover, we were able to identify additional 67 cis-eQTL genes that are largely tissue-specific. As the cost of deep sequencing rapidly decreases, re-visiting microarray-based eQTL studies with deeply sequenced genomes will refine our understanding of the shared and tissue-specific architecture of eQTLs in more population-specific manner.

334W

A T2D case-control study in Africans of Zulu origin. *M. Sun[1], J. Chen[2], A. Morris[1], F. Pirie[3], A. Motala[3], M. Sandhu[3], M. McCarthy[1], I. Barroso[2], A. Mahajan[1], E. Wheeler[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 3) University of KwaZulu-Natal, Nelson R Mandela School of Medicine, 419 Umbilo Road, Congella, Durban 4001, South Africa.

Genome-wide association studies (GWAS) and follow up meta-analyses have identified nearly 100 susceptibility loci that show robust evidence of association with type 2 diabetes mellitus (T2D). However, these studies are predominantly in samples of European or Asian ancestries. African ancestry populations are still under-studied and need more representation (i) to find variants specifically associated with African populations; and (ii) to facilitate fine mapping of the causal variant(s). To support these aims, we conducted a GWAS of 1602 T2D cases and 976 controls (Durban diabetes case control study) in African subjects of Zulu origin collected in Durban, South Africa. Samples were genotyped using a customized Illumina Multi-Ethnic Genotyping Array which leverages content discovered in several sequencing consortia and databases e.g. the 1000 Genomes Project, CAAPA Consortium, PAGE, and includes European, Asian, Hispanic, African and multiethnic exome content. Samples were imputed up to the 1000 Genomes phase 3 reference panel (October 2014 release). We tested variants for association with T2D using a linear mixed model to account for cryptic relatedness and population structure. We found two loci reaching conventional genome-wide significance ($p < 5x10^{-8}$) and a further four loci with $p<5x10^{-7}$. One genome-wide significant locus mapped to a region outside an established T2D loci [rs79276081 intergenic variant, $p=2.5x10^{-8}$, minor allele frequency (MAF) =1.4% in our data]. The other mapped to the *INS-IGF2* region for which previous association has been seen in African-Americans: the lead variant in our analysis, rs6578990 ($p=2.5x10^{-8}$, MAF=39%) was located 8.4kb from the previously-reported T2D GWAS index SNP rs3842770 (p= 0.01, MAF=29% in our data). In African individuals, the two variants are in weak LD ($r^2<0.2$), indicating either multiple signals or refined localization of the causal variant by our study. Both variants in are rarer (MAF<2.0%) in non-African populations. In addition, seven lead GWAS variants from established T2D susceptibility loci were associated with T2D (p<0.05) and show directionally consistent effects, the strongest signal being observed at *TCF7L2* ($p=1.0x10^{-6}$). Despite the modest sample size of this study, our results highlight the importance of performing GWAS in African ancestry populations which have not been widely studied to date. However, further expansion of sample sets is needed to support replication and meta-analyses.

335T

Estimation of the prevalence of classic galactosemia based on the gene frequency in next-generation sequencing data. *D. Wang[1], J. Johnston[2], A. Cashion[1], L. Biesecker[2].* 1) NINR/NIH, Bethesda, MD; 2) NHGRI/NIH, Bethesda, MD.

Purpose Classic galactosemia is a rare autosomal recessive disorder caused by deficiency of galactose-1-phosphate uridyltransferase (GALT), due to mutations in the *GALT* gene. To date, more than 300 mutations in *GALT* have been associated with GALT deficiency. In US, the incidence of classic galactosemia determined by newborn screening is about 1/50,000, though the exact prevalence is unknown. We analyzed the *GALT* variants in three cohorts of exome sequence data to estimate the prevalence of the disorder. Methods We evaluated the *GALT* variants from the ClinSeq cohort (n=1001) and the NHLBI GO Exome Sequencing Project (ESP) cohort (n= 6,503) or the Exome Aggregation Consortium (ExAC) cohort (n= 60,706). Pathogenicity of each variant was determined based on the latest ACMG Variant-Interpretation Guidelines (2015). The disease prevalence was estimated based on the frequency of the pathogenic/likely pathogenic variants using Hardy-Weinberg equilibrium. A second method was based on the allele frequency of the p.Q188R variant which accounts for 62.8% of all disease-causing alleles in European Americans (EA) patients. Results We identified 16 pathogenic/likely pathogenic variants in the ClinSeq and the GO ESP cohorts. The disease prevalence was estimated to be 1/ 53,120 (95% CI 1/29,901-1/99,012) in EA and 1/59,896 (95% CI 1/24,044 – 1/170,330) in African Americans (AA). Based on the allele frequency of the p.Q188R variant in the two cohorts, the prevalence in EA was estimated to be 1/60,790 (95% CI 1/33,556 -1/115,929). A total of 47 pathogenic/likely pathogenic variants were identified in the ExAC cohort and the majority of the variants were population-specific. The estimated disease prevalence varies greatly in different populations with the highest in AA (1/33,057, 95% CI 1/19,837-1/56,689) and the lowest in East Asian Americans (1/2,993,259, 95% CI 1/550,326-1/28,293,345). Based on the allele frequency of the p.Q188R variant in the cohort, the prevalence in EA was estimated to be 1/92,018 (95% CI 1/70,695 – 1/120,898). Our results suggest that analysis of gene frequency using next generation sequencing data is a powerful and unbiased approach to estimate prevalence of autosomal recessive diseases based on precise categorization of variants. The data also provide an explanation for the widely varied incidences of classic galactosemia in different countries, suggesting that different mutations are mainly responsible for GALT deficiency in different populations.

336F

Detecting differences in longitudinal gene expression profiles in Cesarean section in the first year of life. *M. Laimighofer[1,2], R. Puff[4], R. Fürst[4], F. Theis[1,2], A. Ziegler[4], J. Krumsiek[1,3].* 1) Institute of Computational Biology, Helmholtz Zentrum München, Munich, Bavaria, Germany; 2) Department of Mathematics, TU Munchen, Garching, Germany; 3) German Center for Diabetes Research (DZD), Munchen-Neuherberg, Germany; 4) Institute of Diabetes Research, Helmholtz Zentrum Munchen and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universitat, Munchen, Germany.

   Background: Cesarean section (CS) is a surgical procedure that was originally introduced to diminish maternal and fetal birth complications. However, changes in the social, demographic, and economic characteristics of pregnant women as well as alterations in the health care systems have led to a continuous increase of CS over the last two decades. As previously reviewed, the mode of delivery affects the developing immune system by influencing the perinatal programming, thereby causing short- and long-term effects on the offsprings' health. In line with these immunological alterations, CS is associated with a higher risk for autoimmune disorders such as type 1 diabetes (T1D). We are interested in the difference between gene expression profiles during the first year of life after CS or vaginal delivery. **Data:** Peripheral blood mononuclear cell (PBMC) gene expression profiles were measured using microarrays in 108 children, 62 of them born by vaginal delivery and 46 by CS. These measurements were repeated in time and resulted in overall 450 genome-wide transcriptome samples. Restricting to the first year of life, 154 samples were used for this analysis. **Methods**: In order to investigate time-resolved gene expression differences after CS, the following generalized mixed model was fitted per gene: A time correction was performed by adding a spline effect for sample age, a random effect was introduced to account for repeated measurements per child, and a group term for CS versus vaginal delivery was added to capture the effect of interest. Including functional knowledge about genes, we performed a gene set enrichment analysis to identify CS-affected pathways. **Results:** At global transcriptome level and univariate gene level, we could not identify significant effects of CS. However, at pathway level, we obtained a series of significant associations. Pathway enrichment was performed on KEGG pathways and GO slim biological processes. We found significant associations with CS for the pentose phosphate pathway, pyrimidine metabolism and the generation of precursor metabolites. **Conclusion:** Our study shows that there is a systematic impact of CS on the early PBMC transcriptome. Future work should attempt to delineate how these molecular changes translate to an increased risk of autoimmune diseases such as T1D.

337W

Detection and correction of pleiotropic bias in multivariable Mendelian randomization. *C. Chen[1,2,3], M. Verbanck[4], R. Do[4], B. Neale[1,2,3].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Icahn School of Medicine, Mount Sinai Hospital, New York, NY.

   Inferring the causal relationships between a disease outcome and associated traits is essential to understanding disease etiology and the development of treatments. Mendelian randomization (MR) is a widely used method that uses genetic variants as instruments to obtain estimates of causal effects. With the increasing number of disease-associated loci identified by genome-wide association and sequencing studies, methods have been developed to aggregate causal effect estimates from multiple MR analyses, such as Multivariable MR (Do et al. 2013 Nat. Genet.) and MR-Egger regression (Bowden et al. 2015 Int. J. Epidemiol.).   A key assumption of MR analysis is exclusion restriction, which states that the genetic instrument solely acts on the disease outcome through the risk factor. However, pleiotropy is a pervasive phenomenon in human genetics. Consequently, Multivariable MR approaches are increasingly subject to violations of the exclusion restriction assumption. We propose a diagnostic test for detecting unaccounted pleiotropic effects across the set of genetic variants that are included in a Multivariable MR analysis. Specifically, we test for excess variance in the residuals of the Multivariable MR regression analysis. To evaluate the sensitivity of this approach, we performed a series of simulations with and without pleiotropic effects of the genetic variants. Our simulations showed that inverse-variance weighted meta-analysis, which assumes all genetic variants have no pleiotropic effect, gives biased results. Both MR-Egger regression and Multivariable MR showed similar ability to control for pleiotropic bias. These simulations also demonstrated that our method can identify the presence of SNPs with pleiotropic effects. The key parameters that drive the power to detect such pleiotropy are number of instruments, variance explained of the intermediate risk factor and the degree of pleiotropy. Following the detection of pleiotropic variants, outlier identification and robust regression were performed to correct for pleiotropic bias. We applied our method to a multivariable MR analysis of the causal effects of plasma lipids on coronary artery disease and type 2 diabetes, and found no significant pleiotropic bias. Our method improves on existing approaches to multivariable MR analysis by reducing the chance of incurring pleiotropic bias.

338T

Genome-wide association study of quantitative liver fat from the UK Biobank identifies loci that may contribute to fatty liver disease. *M.R. Miller[1], T. Tuthill[1], K.A. Loomis[1], O. Dahlqvist Leinhard[2,3], C.L. Hyde[1].* 1) Human Genetics and Computational Biomedicine, Pfizer, Inc, Cambridge, MA; 2) Linköping University, Sweden; 3) Advanced MR Analytics AB, Linköping, Sweden.

  Introduction: The UK Biobank is a large population based resource with clinical information, medical records and genotype information on 500,000 subjects. Whole body magnetic resonance (MR) images have been collected on ~9000 subjects, and measurements of adiposity, liver fat and muscle volume have been quantified. The imaging data, in conjunction with clinical data and medical records can help to better define obesity-related phenotypes, such as non-alcoholic fatty liver disease (NAFLD) and to identify genetic loci that may contribute to these phenotypes. NAFLD is associated with liver fat accumulation, and can lead to inflammation and fibrosis (non-alcoholic steatohepatitis; NASH). Both NAFLD and NASH have a high burden of disease and no approved treatments. Because of the difficulty in obtaining liver fat biopsies or quantitative measures of liver fat, very few GWAS of liver fat or NAFLD have been performed. Identification of new loci for liver fat based on MR imaging may point to mechanisms of disease.   **Methods**: A genome-wide association study of quantitative liver fat was performed via linear regression on 1467 European subjects from the UK Biobank with MRI and genotyping data available. The model was adjusted for age, sex, BMI, type 2 diabetes status, and daily alcohol consumption.   **Results:** We confirmed GWAS associations of liver with *PNPLA3* (rs738409; $p = 6.31 \times 10^{-15}$) and *TM6SF2* (rs58542926; $p = 3.71 \times 10^{-10}$), both established liver fat/NAFLD genes. While no additional genome-wide signals were identified, one suggestive signals was also identified on chr10 in the region of EMX2 and EMX20S (rs74743905, $p = 6.213 \times 10^{-07}$).   **Conclusions**: A genome-wide association study on a relatively small sample of subjects was able to confirm two previously identified GWAS loci for liver fat, demonstrating that MR imaging of liver fat can be used to identify genetic loci that may contribute to metabolic diseases. As the size of the imaging cohort in the UK biobank increases, the UK Biobank will be a valuable resource to identify additional genetic targets for liver fat. In addition, the UK Biobank will continue to be a valuable resource for scientific discovery and precision medicine by enabling integration of individual level genetic and multi-dimensional phenotypic data.

339F

Association study of rare variants on serum urate level by using whole genome cohort study samples. *K. Misawa[1], T. Hasegawa[1,2], K. Kojima[1], Y. Kawai[1], A. Hozawa[1], T. Takai[1], N. Fuse[1], N. Minegishi[1], J. Yasuda[1], S. Ogishima[1], S. Kure[1], K. Kinoshita[1], M. Yamamoto[1], M. Nagasaki[1].* 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Health Intelligence Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan.

  Gout is one of the most common inflammatory arthritides, and is caused by hyperuricemia. A twin study showed that the heritability of serum uric acid levels (SUL) and gout are approximately 50 %. However, only 6 % of genetic factors for SUL and gout can be explained by known genetic variation. Unknown low-frequency and rare variants msut contribute to the missing heritability. The Tohoku Medical Megabank Organization (ToMMo) has sequenced 1,070 whole genomes of Community Resident Cohort participants. To identify low-frequency and rare variants that affect SUL, we utilized this data. A linear regression model as well as a sequence kernel association test (SKAT) was used to analyze the associations between serum uric acid levels and SNP genotypes by incorporating age, sex, and body mass index (BMI). We identified novel rare variants with genome-wide significance associated with SUL. Fifteen percent of genetic variation in SUL was explained by these novel and known variants.

340W

Integration of whole genome sequence and epigenomic data highlights regulatory activity in the genomic architecture of glycemic traits. *A. Manning[1], S. Lent[2], M. Jones[3], CHARGE, GoT2D/T2D-GENES.* 1) Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA; 2) School of Public Health, Boston University, Boston, MA; 3) Cedars-Sinai Medical Center, Los Angeles, CA.

Studies of common genetic variants have identified many loci associated with type 2 diabetes (T2D) and glycemic traits that fall outside of protein-coding regions, implicating gene regulation as the causal mechanism. We hypothesize that by integrating diabetes physiology in appropriately designed association tests for whole genome sequence (WGS) data, we will reveal novel underlying molecular mechanisms pointing to new therapies for T2D and insulin resistance. We analyzed WGS data (6-fold average depth) from 2,854 European, non-diabetic individuals from 3 CHARGE cohorts (FHS, CHS and ARIC). With the hypothesis generating nature of the approach in mind, we limited analysis to 69 loci previously associated with glycemic trait levels. We performed targeted replication of CHARGE findings in WGS data with 1,400 non-diabetic individuals from GoT2D. Our genetic association models accounted for family structure and used inverse-normalized covariate-adjusted fasting glucose (FG) and log-transformed fasting insulin (FI) levels, with additional refinement to account for possible interaction by obesity. We aimed to (1) fine-map common variant associations, (2) discover independent common (minor allele frequency [MAF]>5%), low-frequency (0.5%<MAF<5%) and rare (MAF<0.5%) genetic associations in these regions, and (3) using publicly available data from the Roadmap Epigenomics Project, perform rare variant associations using SKAT of active regulatory elements in pancreas, liver, adipose and skeletal muscle tissue. We found 267 loci with nominal (P<0.0001) single variant associations with FG and/or FI levels, within which 34 independent variant associations were observed (14 common, 6 low-frequency and 14 rare). Only 2 of the common variants replicated with P<0.01 in GoT2D. In rare variant tests of active transcription marks (H3K27ac, H3K4me1 and H3K4me3), 25 loci showed nominal (P<0.001) associations with either the same trait as previously reported for the locus or both FG and FI. Replication in GoT2D showed rare variant associations of non-coding variants (SKAT P<0.01) at the *SLC10A6*, *ADCY5* and *PCSK1* loci (with epigenomic peaks observed in multiple tissues). Although sample sizes are small in the current analysis, we have increased our understanding of the genomic architecture of glycemic traits and demonstrated the validity of performing rare variant tests for FG and FI in the non-coding genome using epigenomics data in tissues implicated in T2D and insulin resistance.

341T

Analysis of population specific pharmacogenomic variants using next generation sequencing data. *T. Park[1], E. Ahn[2], Y. Kim[2,3], B-J. Kim[3], T2D Genes Consortium.* 1) Department of Statistics, Seoul National University, Seoul, South Korea; 2) Interdisciplinary Program of Bioinformatics, Seoul National University, Seoul, South Korea; 3) Division of Structural and Functional Genomics, Center for Genome Science, Korean National Institute of Health, Osong, Chungchungbuk-do, South Korea.

The analysis of population differentiation (PD) in pharmacogenomic variants provides an insight on the ethnic sensitivity of drug responses. The analysis of PD for common variants has been successfully performed by using Fst or chi-square test. However, the analysis of PD for rare/less-common variants has rarely been performed. In this study, we aim to perform PD analysis of rare variants to identify the drug-related genes (pharmacogenes). We first show that Fst and chi-square test used for common variants are not suitable for rare variants. We then suggest a new gene-based PD method for handling rare variants inspired by the Generalized Cochran-Mantel-Haenszel statistics. Through simulation studies, we reveal that the proposed PD method adequately summarizes the rare and common variants for PD over a gene. We also apply the proposed method to the real whole exome sequencing dataset which consists of 10K data from Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) and 3K data from Genetics of Type 2 diabetes (Go-T2D) samples. Among the 48 genes that are annotated with very important pharmacogene summary in PharmGKB database, our PD method identified 7 novel candidate genes including CYP2B6, ACE, and SCN5A. These three genes were previously reported as being highly differentiated in expression levels, but PD of them has not been identified via Fst, which supports the validity of our proposed method through literature review.

342F

Genome-wide association analysis identifies multiple loci associated with kidney function-related traits in Korean populations. *N. Heo[1], Y. Lee[1], J.E. Lee[2], E.S. Shin[2], M.S. Jang[2].* 1) Healthcare System Gangnam Center, Seoul National University Hospital , Seoul, South Korea; 2) DNA Link, Inc., Seoul, South Korea.

   Background/Aims: Chronic kidney disease is a significant public health problem, and the assessment of genetic factors influencing kidney function has substantial clinical relevance. Here, we report an analysis of genome-wide association studies for kidney function–related traits in Korean. **Methods:** We performed a genome-wide association study for kidney function–related traits in large Korean population based samples of 7,999 subjects. We investigated SNPs that are associated with kidney function–related traits, including the concentrations of blood urea nitrogen and estimated glomerular filtration rate based on serum creatinine level. **Results:** We identified 5 previously known loci and 6 new loci associated with kidney function–related traits at the genome-wide significance level (P<10[-7]). After adjusting for age, sex, hypertension and diabetes, four loci, Rs12702509 (UNCX), rs12776722 (METTL10), rs10175462 (PAX8), and rs 9687753 (ANKHD1) were newly identified as loci associated with blood urea nitrogen. And our analysis identified three loci, Rs2390793 (LRP2), rs12702509 (UNCX), and rs144503008 (SSBP3) which were newly associated with glomerular filtration rate after adjusting for age, sex, hypertension and diabetes. Rs12702509 is associated with both blood urea nitrogen and estimated glomerular filtration rate. **Conclusions:** We identified associations among these loci with kidney function–related traits. These findings provide new insights into the susceptibility to CKD and genetics of kidney function. In future studies, we will further examine the association of these loci with CKD trait in the other cohort for validation.

343W

Increased power for sequence-based case-control genome-wide association studies by including external controls. *D. Ray, P. Yajnik, M. Boehnke.* Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

   Genome-wide association studies (GWAS) routinely employ the case-control study design for binary traits like disease status. For diseases with low/moderate prevalence, perfectly ascertained controls may be replaced or supplemented with publicly available datasets that have dense genotype measurements but no (or poor) information about disease status. Inclusion of external controls increases sample size but may misclassify some affected individuals as controls. Edwards et al (2005) determined power for association analytically when such misclassification is present. As publicly available datasets become larger, inclusion of external controls matched for ancestry and genotyping quality may substantially increase power without affecting type I error (Ho et al, 2010). Keeping the financial cost fixed at the cost of a case-control study with perfectly ascertained n cases and n controls (reference design), we consider two study designs that make use of N external controls in addition to either (i) n cases and n controls (design A) or (ii) $2n$ cases and 0 controls (design B). We make the simplifying assumptions that the external controls were sampled randomly from the same population as the study individuals (so that they have the same genetic ancestry, and the proportion of external controls who are truly affected equals the disease prevalence) and that there are no systematic differences in batch genotype effects. We evaluated the performance of these study designs compared to the reference design. For our simulations, we fixed $n=2500$, the odds-ratio to 1.15, the minor allele frequency to 0.2, and varied N to be 2500, 3500, 5000 or 7000 and prevalence to be 0.1, 0.2, 0.3 or 0.4. Our initial results focus on power; eventually we plan to focus on effective sample size. In general, for a fixed N, powers of the studies decrease with increasing disease prevalence (equivalently, increasing mislabeling). For prevalence 0.2 or less, design A is more powerful than the reference design regardless of the value of N, and greater power gain is possible by sequencing more cases. Greatest power gain is achieved using design B if N is at least as large as the number of study individuals. In summary, we show that one can substantially increase power for association by including external controls when disease prevalence is not too large. The choice of the external control size is governed by disease prevalence and the number of available internal and external study individuals.

344T

iFunMed: Integrative functional mediation analysis of GWAS and eQTL.
C. Rojo[1], Q. Zhang[3], S. Keles[1,2]. 1) Department of Statistics, University of Wisconsin-Madison, Madison, WI; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI; 3) Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE.

Genome wide association studies (GWAS) are widely used to elucidate genetic variation associated with disease and traits. When combined with genome-wide expression quantitative loci analysis (eQTL), these studies have the potential to identify genetic variants that modulate disease or trait phenotypes through their impact on expression of specific genes. Two challenging aspects of these studies are: high dimensionality of the variant space and that most candidate genetic variants are in non-coding regions making their interpretation challenging, if not impossible. We develop a novel mediation analysis framework, named iFunMed, to integrate genome-wide association studies and the expression quantitative loci (eQTL) studies with the effective utilization of publicly available functional annotation data.   iFunMed extends the scope of standard mediation analysis framework by incorporating information from multiple genetic variants at a time by utilizing variant-level summary statistics. Furthermore, this integrative framework incorporates the regulatory information encoded in functional annotation data to elucidate how top-ranking association SNPs might be modulating gene expression. A key aspect of iFunMed is its ability to select annotations that are most relevant to the data and cluster them appropriately into a reduced number of groups with a more homogenous biological context . iFunMed has the power to identify mediation potential of genes, estimate direct and indirect effect of each SNP for the trait, and identify relevant functional annotations impacting direct and indirect effects of the SNPs.   We applied our method to Framingham Heart Study data of diabetes for 1,663 subjects focusing on regions with 153 known GWAS hits and within 1MB from each transcript resulting in a total of 525,168 SNPs and 17,873 transcripts from whole genome sequencing. We used 642 epigenomic annotations from the ROADMAP Epigenomics Project and 10 annotations derived in a gene-centric way using GENCODE data. This large scale analysis identified many genes with significant mediator effects. Among these is *PPT1* which is highly expressed in mouse pancreatic islets.

345F

Empirical heritability estimates for body mass index in American Indian sibships. *R.L. Hanson, W.C. Hsueh, S. Kobes, W.C. Knowler, L.J. Baier.* DECRS, NIDDK, Phoenix, AZ.

Genome-wide association studies (GWAS) have identified numerous variants associated with body mass index (BMI), but these associations only explain a small fraction of the currently estimated genetic heritability. One potential explanation is that classical heritability estimates, which are based on expected allele sharing in relatives, can be confounded by shared environmental factors, whereas estimates which employ empirically observed measures of the proportion of alleles shared identical by descent (IBD) within sibships are robust to such confounding. We employed this empirical approach to estimate heritability for BMI in 4462 American Indians, constituting 1508 sibships and 5855 sibling pairs, who had participated in a longitudinal study. A GWAS was performed using a custom Axiom array and IBD allele sharing was estimated for 506,627 autosomal SNPs using the program IBDLD. Actual IBD allele sharing in sibling pairs was taken as the average of observed IBD allele sharing across all SNPs. The maximum age-sex standardized BMI Z-score observed in the longitudinal study was analyzed with adjustment for age at maximum score, sex, birth year and the 1st 5 principal components from the GWAS. Variance components methods were used to obtain the classic heritability estimate (h^2, based on expected IBD sharing), and to further partition familial resemblance into a genetic component (a^2 reflecting actual IBD sharing) and a common sibling component (c^2, presumably reflecting shared environment). Mean IBD allele sharing among all sibling pairs was 0.50 with SD=0.04. BMI was highly familial with h^2=0.68 (P=2.0×10[-106]) by the classic estimate; when this familial resemblance was partitioned by analyzing actual IBD sharing, the genetic component was significant (a^2=0.49, P=0.027) while the common sibship effect was not (c^2=0.09, P=0.24). When analysis was restricted to individuals who had been examined when they were not diabetic (n=4183), similar results were obtained: h^2=0.71 (P=1.2×10[-103]), a^2=0.56 (P=0.015), c^2=0.07 (P=0.31). These analyses are consistent with the hypothesis that familial resemblance of BMI largely reflects the effect of genetic factors; the fact that established variants account for little of the heritability may be indicative of the polygenic nature of obesity.

346W

Multiple lysosomal storage disease newborn screening in a Mexican cohort of patients: Results and follow-up from the Pemex Health System Newborn Screening Program. *J. Navarrete[1], D. Cervantes[1], A. Limon[2], G. Wakida[2], R. Delgado[3], C. Cantu[4], H. Cruz[4], M. Gaytan[1].* 1) Dept Gen, Hosp Sur PEMEX, Mexico City, Distrito Federal, Mexico; 2) Dept Pediatrics, Hosp Sur PEMEX, Mexico City, Distrito Federal, Mexico; 3) Dept Pediatrics, Hosp Reg Villahermosa, Tabasco, Mexico; 4) Dept Gen, Genomi-K Laboratories, Monterrey, Mex.

Lysosomal storage disorders (LSDs) are a diverse group of over 50 inherited metabolic disorderes, that result in cellular disfunction and multiple organ damage due to a progressive accumulation of metabolic precursors within the lysosomes, secondary to reduced or lack of a lysosomal enzyme. The prevalence of these conditions had been shown to be more frequent than previously expected. With a collective estimated incidence of 1 in 7,000 to 9,000 live births, pilot newborn screening studies revealed unexpected LSD higher frequencies. Petroleos Mexicanos is a big governmental institution with approximately ten thousand workers and their families. Since 2005 a larger screening test has been done to all newborns in this institution through all the country. We test for most aminoacidpathies including acidurias, hemoglobinopathies, G6PD deficiency, adrenal hyperplasia, cystic fibrosis and biotinidase deficiency; since August 2 012 we included one immunodeficiency and 6 lysosomal storage diseases, Gaucher, Pompe, Fabry, Hurler, Niemann-Pick type A and B and Krabbe diasease. We analyzed our results from August 2012 to May 2015, we have found 5 newborns wtih Fabry disease, 5 patients with Pompe disease, 1 with Krabbe disease and 2 with Hurler disease. We describe our findings and compared with other populations, and we make a close follow up of all newborns with pathological mutations. and study and analize their families. We also look for biomarkers of the disease so we can start treatment depending of the mutation and the course of the disease as soon as possible. We present all our LSD patients, their mutations found and their families. We also present the patients that have started ERT and their follow up status.

347T

Population-level variability in adipose tissue cell-type composition and its link to obesity. *C.A. Glastonbury[1], K.S. Small[1], M. Civelek[2], A.J. Lusis[3].* 1) Twins Research & Genetic Epidemiology, King's College London, United Kingdom; 2) University of Virginia, United States; 3) UCLA School of Medicine, United States.

Adipose tissue biology impacts multiple complex traits, particularly those associated to obesity. Adipose tissue is a complex mixture of adipocytes, immune and non-immune cells. Obesity affects the cellular composition and inflammation of adipose tissue but due to the difficultly of flow-sorting adipose tissue, adipose cellular makeup and its relationship to phentoypic traits has been poorly explored. We sought to address this by computationally deconvolving the cell-fractions of primary adipose tissue biopsies from 766 healthy TwinsUK females using RNA-seq. First, we constructed a basis matrix from 7 purified adipose resident cells. Top basis matrix marker genes recapitulate well-known cell-specific genes such as *ADIPOQ*-Adipocytes, *VWF*-endothelial and *CXCR4*-Macrophages. We then estimated the proportion of these 7 cell types in the 766 biopsies using support vector regression. The dominant cell proportions were Adipocytes (μ=0.46, σ=0.09), Fibroblasts (μ=0.35, σ=0.07), Mesenchymal stem cells (μ=0.08, σ=0.01), Endothelial cells (μ=0.04, σ=0.01), M2 macrophages (μ=0.03, σ=0.01) CD4+ T-cells (μ=0.02, σ=0.006) and M1 macrophages (μ=0.007, σ=0.009). We derived cell-type estimates in two independent adipose expression datasets. METSIM samples from healthy Finns (N=200) recapitulated the cell-proportion distributions of TwinsUK but post-mortem GTEX samples (N=230) had poor deconvolution accuracy and highly variable cell-type prediction, potentially reflecting tissue fibrosis and prolonged ischemia. Macrophage infiltration into adipose tissue is known to correlate with obesity. We confirm this finding, showing that BMI and DXA derived visceral fat measurements correlate with M2 macrophages (r = 0.30, P = 2.2 x 10^{-16}) and the overall immune fraction (r = 0.29, P = 2.2 x 10^{-16}). Previous work has shown BMI has a large effect on the adipose transcriptome. To assess mechanistically BMI's association to gene expression, we ran a transcriptome-wide association study for BMI, with and without correcting for macrophage proportion. Without adjustment, 85% of genes are associated to BMI (FDR 5%), but 40% of expression associated with BMI is lost after correcting for Macrophage proportion, and those genes are enriched for the immune response (P=2.4 x 10^{-4}). Finally, we uncover 134 additional *cis*-eQTLs after correcting for macrophage estimates (FDR5%), showing that power to detect local genetic effects can be enhanced by accounting for cellular heterogeneity in tissues.

348F

Genetic and environmental overlap between uncontrolled eating and obesity. *U. Vainik[1,2], R. Mõttus[3,2], C. Kandler[4], A. Dagher[1].* 1) Neurology & Neurosurgery, Montreal Neurological Institute, Montreal, Quebec, Canada; 2) Institute of Psychology, University of Tartu, Tartu, Estonia; 3) Department of Psychology, University of Edinburgh, UK; 4) Department of Psychology and Sports Sciences Bielefeld University, Bielefeld, Germany.

Uncontrolled Eating (UE) is a core component in many obesity-related psychological traits. Longitudinal data has shown that increases or decreases in weight are often mirrored by concurrent changes in UE, and targeting UE with cognitive interventions might reduce BMI. Given this covariation, we assessed the genetic and environmental overlap between UE and obesity in the Bielefeld Longitudinal Study of Adult Twins (975 monozygotic and 493 dizygotic twins) using twin ACE models (A = additive genetic effects, C = shared environment effects, E = unshared environment effects). UE was assessed by a single question "I tend to eat too much of my favorite food."; this question has been previously shown to relate strongly to other measures of UE. Heritability was high for obesity (A=0.83 [0.81, 0.85], C=0 [0.13, 0.13], E=0.18 [0.19, 0.15]) and considerably lower for UE (A=0.25 [0.14, 0.4], C= 0.07 [0, 0.3], E=0.69 [0.61, 0.76]). Still, there was a large genetic correlation between UE and BMI (r= 0.81 [0.71, 0.91]). Therefore, UE does seem to share considerable genetic component with obesity. As a next step, other methods, such as narrow sense heritability should be used to clarify the extent of overlap between UE and obesity.

349W

Multi-association analysis identifies T2D variants in the Korean population-based cohort. *J. Lee, M. Go, Y. Lee, B. Kim.* Structural & Genomics, Centers Disease Control, Osong, South Korea.

Type 2 diabetes mellitus (T2D) is a multi-factorial polygenic disease. To date, T2D Genome-wide association studies (GWAS) and meta analysis loci (~ 80) are insufficient to dissect the complex genetic structure underlying etiological heterogeneity. Epistasis has long been suggested to be important to understanding of the missing heritability problem. To identify epistatic effects on T2D, we conducted a multi-association analysis as complementary to the current single SNP-based GWAS. By using community-based cohort, we selected several variables for examining multi traits-unit SNP association; glycated hemoglobin (HbA1C), Fasting plasma glucose (GLU0), plasma glucose 2-hour after ingestion of 75g oral glucose (GLU120). We composed trait set and examined multi traits-unit SNP association with age and sex as the covariates. Our study confirmed previously reported T2D loci (CDKAL1, C2CD4A etc). These results could provide additional insight into the genetic regulation of T2D but require further replicative efforts to confirm the findings.

350T

Genetic variation underlying self-reported weight-loss success in the presence of a dietary intervention. *J.F. Shelton, N.A Furlotte, R. Smith, C. Wheldon, M.N. Johnson, C. Wilson, G. Benton, R. Gentleman. 23andMe, Mountain View, CA.*

   Genetic variation underlying obesity and adiposity are well established, but recent studies have also identified variants that confer different response to diet and exercise. For example, variants in the FTO and MTIF3 genes are associated with differential weight loss in response to caloric restrictions, but variants in the ADRB [2 and 3], FABP2, PPARG, and APOA [2 and 5] genes may be associated with differential weight loss in response to reductions in macronutrients such as fat and carbohydrates. Using a web-based question-naire, we asked 100,000 23andMe consented research participants to detail their dieting (lifetime and past-year) experience. Participants answered questions on restriction of certain food and beverage groups, exercise frequency, duration of diet, adherence to diet, and dietary outcomes based on self-report-ed goals achieved and weight lost. Participants were selected without regard to previously reported body mass index. Of those who responded (n=9,858), 85 percent reported that they had dieted, and 62 percent reported dieting in the past year. Of those who had ever dieted, 12 percent reported never being able to lose weight. Among those who had dieted in the past year, 66 percent reported not being able to achieve their weight-loss goals. Using nine SNPs (rs801282, rs1799883, rs1885988, rs662799, rs9939609, rs5082, rs4994, rs1042713, and rs1137101) previously identified to have a role in differential response to weight-loss interventions, we evaluated the relationship between genotype and the percentage of body weight lost in the past year. For SNPs associated with carbohydrate or fat metabolism, we built models with interaction terms, controlling for other factors affecting the percent of weight lost from baseline. Stricter adherence and longer duration of the diet were predictive of greater weight loss. Preliminary findings show that among people with strict adherence to their diet — and controlling for exercise, duration of the diet, age, sex, and population stratification variables — the GG variant of rs662799 (APOA5) was associated with a five percent greater weight loss in the presence of a fat restriction among women but not men. An internal replication study will be conducted prior to publication of findings. These results indicate that while behavioral changes are the main drivers of weight loss, genetic variation may explain some differences across individuals.

351F

Whole genome sequencing identifies functional founder effect variants associated with CVD traits in the Norfolk Island isolate. *L.R. Griffiths[1], M.B. Benton[1], H.G. Sutherland[1], D.E. Eccles[1], L.M. Haupt[1], S. Stuart[1], J. Blangero[2], J. Curran[2], R.A. Lea[1]. 1) Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia; 2) The University of Texas Rio Grande Valley, Harlingen, Tx.*

   Background: Genetic isolates have distinct advantages for mapping gene variants that influence complex traits. Founder effects in particular can amplify the frequency of ancestrally rare variants, which in turn, can contribute to an increased prevalence of disease in the isolated population. The Norfolk Island (NI) population is a genetic isolate located off the east coast of Australia. Founded by 11 British mutineers of the HMS Bounty and 6 Polynesian women in the late 1700s the NI population is comprised of a unique 6000-member pedigree spanning 11 generations. We have previously shown that the existing NI population exhibits higher rates of several cardiovascular disease risk traits compared to mainland Australia. In this study we aimed to identify functional founder effect variants associated with CVD traits in the NI isolate.   **Methods:** We performed whole genome sequencing (WGS) of 108 individuals selected from the core founder pedigree using the Illumina HiSeq X10. We first parsed the sequence data through an in-house pipeline using BOWTIE2, SAMTOOLS, VCF, VEP. Secondly we prioritised variants based on being a) probably damaging according to 5 in-silico functional tests b) founder effect variants based on MAF >5% in NI and MAF <1% in the 1000G cohort c) associated with CVD traits.   **Results:** WGS analysis showed an average of >25X coverage. In total we identified 13.4M variants of which 96% are SNVs and ~5.3M variants are unique to NI. Variant prioritisation identified 11 functional founder effect variants. The top ranked variant was rs77408762 that codes a TYR-to-ASN change in the Acyl-CoA Thioesterase 4 (*ACOT4*) gene. The ASN variant was present at 26% in the NI cohort but was rare or absent in the ancestral cohorts European (0.9%) and East Asian/Polynesian (0%). This suggests that this founder effect allele originated specifically from England and was transferred to NI via the Bounty Mutineers. Interestingly *ACOT4* is involved in lipid metabolism and we found evidence that the *ACOT4*-ASN variant is significantly associated with altered HDL, LDL levels as well as body fat indices in the NI cohort.   **Conclusions:** This study has used a customised WGS approach to identify a panel of functional founder effect variants in the NI isolate. Furthermore, we provide evidence for a novel association between *ACOT* and CVD risk traits in NI highlighting the potential of this genetic isolate for uncovering previously undetected variants for CVD using this approach.

352W

A novel mutation in *HEXB* that causing Infantil Sandhoff Disease. *N.H. Akcakaya[1], O. Ozdemir[1], S.A. Ugur-Iseri[1], E. Yucesan[1], P. Tekturk[2], G. Gokcay[3], U. Ozbek[1], Z. Yapici[2].* 1) Istanbul University, Aziz Sancar Institute of Experimental Medicine, Department of Genetics, Istanbul, Turkey; 2) Istanbul University, Istanbul Medical Faculty, Department of Neurology, Division of Child Neurology, Istanbul, Turkey; 3) Istanbul University, Istanbul Medical Faculty, Department of Pediatrics, Division of Metabolic Disorders, Istanbul/Turkey.

   Introduction: Hexosaminidase beta subunit (*HEXB*) gene encodes a lysosomal enzyme. The mutations in *HEXB* leads neuronal damage due to the accumulation of GM2 ganglioside. This lysosomal storage disease named Sandhoff disease (SD) and has three forms (infantile, juvenile, and adult forms). Infantile SD starts between 3-6 months of life and causes death within a few years. In the course of SD developmental delay, progressive decline of motor functions, hearing loss, cherry red spots, macrocephaly, and seizures observed.Methods and Results: Eighteen month of age male patient was admitted to neurology clinic with developmental delay, progressive decline of motor functions, and seizures. He had consanguineous parents. He was hypotonic since birth. He wasn't able to sit until six months of age and within two months mental and motor deterioration starts. Seizures was added at 15 month of age. He had doll-like face, cherry red spots, and macrocephaly. Brain MRI revealed low signal intensity symmetrical at bilateral thalamus in FLAIR and T2-weighted images and high signal intensity at white matter. Total hexosaminidase activity was very low but hexosaminidase A activity was normal.Whole exome sequencing revealed novel homozygous c.1538T>C; p.Leu513Pro (RefSeq. NM_000521) *HEXB* mutation. We confirmed the mutation by Sanger sequencing. Healthy parents were heterozygous carriers. The variant was absent in 350 unrelated healthy individuals from Turkey, who were exome- sequenced by TUBITAK-BILGEM. In silico tools were used and the amino acid substitution was predicted to have a damaging effect on the structure and function of the protein.Discussion: SD has enzymatic diagnostic test, usually genetic testing is not performed and so there are few known mutations. The *GM2A, GNPTG, GNPTAB, HEXA* genes are the related genes with Sandhoff phenotype but these genes did not carry any additional pathogenic variant in our patient. Our patients mutation was in one of the HEXB alpha helix structure that plays an active role in enzymatic function. It has been shown to be compatible with early involvement of radiological and clinical rapid progression. Identifying genetic variations that causes SD could enable to made genotype-phenotype correlation.This work was supported by the grant of Scientific Research Projects Coordination Unit of Istanbul University, Project Number: 51985 & Istanbul Development Agency, Project Number: TR10/15/YNK/0093. .

353T

Multi-ethnic 1000G analysis of TG:HDL ratio in over 78,000 individuals from the CHARGE Consortium. *J.E. Huffman, The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Lipids Working Group.* The Framingham Heart Study, NHLBI Population Sciences Branch, Framingham, MA.

   The ratio of triglyceride (TG) to high density lipoprotein (HDL) cholesterol has been shown to be highly correlated with insulin resistance and a strong predictor of type 2 diabetes. A glucose clamp procedure or frequently sampled intravenous glucose tolerance test provide a measurement of insulin resistance but are not routinely done in large numbers of subjects due to invasiveness, time and cost of these tests. TG:HDL ratio is calculated based on values from routine lipid profiles and can be readily measured in a large number of individuals. The goal of this study is to perform a genome-wide association study (GWAS) of TG:HDL ratio in order to identify novel genetic regions associated with insulin resistance. The largest previous GWAS of this trait was completed in 2009 using HapMap imputed genotypes, was half the sample size, and identified 2 loci, *APOA1* & *CETP*, both of which were also associated with classical lipids. Meta-analysis was performed for 22 cohorts contributing data from 78,382 individuals, including 54,586 of European ancestry, 8,637 of African ancestry, 1,016 East Asians, and 14,143 Hispanics. Individual cohorts performed a GWAS of TG:HDL ratio using 1000Genomes imputed dosages. The outcome was fasting, natural log-transformed TG:HDL ratio adjusted for sex, age, age^2 and cohort-specific covariates. Inverse variance weighted meta-analysis was performed in METAL within each ethnicity and then combined for a trans-ethnic meta-analysis. Significant associations (p<5E-08) were found at 36 loci in regions previously associated with classical lipids. All have been associated with TG and/or HDL. Both hits from the previous TG:HDL ratio GWAS were replicated. In addition, 4 novel loci were discovered. The first was located at 5q31.1 with the lead SNP found ~2kb 3' of *HSPA4* and the majority of the association peak spanned this gene. The second was found at 8p23.1 with the top SNP mapping ~2kb 3' of *DEFB136*, in a region with extended LD in Europeans. A novel intronic variant within *PRR25* at 16p13.3 reached significance. The final novel association was at 19p13.11 ~5kb 5' of *JUND*. We are undertaking replication of our results in independent samples using TG:HDL ratio as well as in samples with measures of insulin-sensitivity. In conclusion, TG:HDL ratio can be measured in large numbers of individuals and thus may provide a surrogate marker of insulin resistance for potential discovery of novel contributing regions of the genome.

354F

Identification of important genes associated with high-density lipoprotein cholesterol using bioinformatics analysis. *X. Mo[1,2,3], H. Zhang[1,2], S. Lei[1,3], Y. Zhang[1,2].* 1) Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Suzhou University, Suzhou, China; 2) Department of Epidemiology, School of Public Health, Medical College of Soochow University, Suzhou, China; 3) Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Suzhou, China.

Background: High-density lipoprotein cholesterol (HDL-C) concentration is affected by both lifestyle and genetic factors. The aim of this study was to identify more related genes for HDL-C and evaluate the functional relevance to provide evidences for prioritizing these genes by integrating different types of data. **Methods:** We performed an initial gene-based association study in about 188,578 individuals (a publicly available data from the Global Lipids Genetics Consortium). Furthermore, we performed Gene Relationships Across Implicated Loci (GRAIL), differential expression, protein-protein interaction, coronary artery disease association analysis and other bioinformatics analyses to support the significance of the identified genes. **Results:** About 22,096 genes on the human genome were analyzed for HDL-C levels in gene-based association analysis. A total of 331 genes were found to be statistically significant after Bonferroni correction ($P < 2.3 \times 10^{-6}$). The evidence obtained from the analyses of this study suggested the importance of some well-known genes, e.g. *LPL*, *ABCA1*, *SCARB1*, *LCAT*, *APOE*, *VEGFA*, *LIPC*, *CETP* and *HNF4A*, as well as some unreported genes such as *TGOLN2* at 2p11.2, *TP53INP2* at 20q11.22, *RAC1* at 7p22.1, *IKZF3* at 17q12, *NR1I2* at 3q12-q13.3, *NR1H3* at 11p11.2, *ERBB2* at 17q12, *BCL3* at 19q13.1, *NOTCH4* at 6p21.3, *PACSIN1* at 6p21.3, *MMP9* at 20q13.12 and *BAZ1B* at 7q11.23, and so on. These genes involved in important lipid related pathways or have been reported to be associated with many related diseases or traits, and should be suggested as important candidates for further studies. **Conclusion:** The present study found some important HDL-C-associated genes and the findings might provide more insights into the genetic basis of lipid metabolism.

355W

Association analyses of up to 72,043 East Asian individuals and trans-ancestry analyses with up to 186,265 European individuals reveal new loci associated with cholesterol and triglyceride levels. *C.N. Spracklen[1], Y.J. Kim[2], P. Chen[3,4], X. Sim[3], H. Cai[5], S. Li[6], J. Long[5], Y. Wu[1], X. Wang[7], Y.X. Wang[8], F. Takeuchi[9], K.J. Jung[10], J.Y. Wu[11,12], T.A. Johnson[13], H. Li[14], R. Dorajoo[15], M. He[16], K. Akiyama[9], Y. Zhang[17], X. Hou[18], C. Hu[18], S. Moon[2], E. Salfati[19], AGEN Consortium.* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 2) Center for Genome Science, Korean National Institute of Health, Osong, Chungchungbuk-do, South Korea; 3) Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; 4) Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, AZ, USA; 5) Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA; 6) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; 7) Life Science Institute, National University of Singapore, Singapore, Singapore; 8) Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab, Beijing, China; 9) National Center for Global Health and Medicine, Japan; 10) Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea; 11) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 12) School of Chinese Medicine, China Medical University, Taichung, Taiwan; 13) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Japan; 14) Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai, China; 15) Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore; 16) Department of Occupational and Environmental Health and State Key Laboratory of Environmental Health for Incubating, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; 17) Department of Epidemiology, School of Public Health and Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Medical College of Soochow University, Suzhou, China; 18) Shanghai Diabetes Institute, Shanghai Jiaotong University Affiliated Sixth People's Hospital, Shanghai, China; 19) Department of Medicine, Stanford University School of Medicine, Palo Alto, CA, USA.

Large-scale meta-analyses of genome-wide association studies (GWAS) have identified >180 loci associated with fasting lipid levels, including total cholesterol (TC), HDL-C, LDL-C, and triglycerides (TG). With differences in linkage disequilibrium (LD) structure and allele frequencies among ancestry groups, genetic studies in additional large samples may detect additional loci. We conducted two GWA meta-analyses in up to 72,043 East Asian individuals from 24 studies performed in Japan, the Philippines, Korea, China, Singapore, and Taiwan, and a third trans-ethnic meta-analysis in up to 211,678 East Asian and European individuals. Included studies were genotyped and imputed to the HapMap Phase 2 reference panel (~2.4 million SNPs). Lipid levels were adjusted for age, age[2], sex, BMI, and other study-specific covariates, and the residuals were then inverse normalized. In the first and second meta-analyses, summary statistics of ~2.4 million study-specific lipid-SNP association analyses were combined using a fixed effects sample-size weighted meta-analysis approach assuming an additive genetic model. In the first meta-analysis, lead signals from the discovery phase (11 studies) were meta-analyzed with 5 *in silico* and 6 *de novo* genotyping studies of selected variants. In the second meta-analysis, results from the discovery phase were meta-analyzed with 2 additional genome-wide studies. With these analyses, we validated ($P<5\times10^{-8}$) 50 known lipid loci and identified 3 novel loci associated with HDL-C near *CD163* ($P=7.4\times10^{-9}$), *NCOA2* ($P=1.6\times10^{-8}$), and *NID2-PTGDR* ($P=4.2\times10^{-8}$), and 1 novel locus associated with TG near *WDR11* ($P=2.7\times10^{-10}$). Additionally, we identified 3 novel lipid associations at loci previously associated with a different lipid trait; HDL-C (*KLHL8*, *GPAM*) and TC (*METAP1*). In the trans-ethnic meta-analysis, summary association statistics from the East Asians were combined with data from 186,265 European individuals from the Global Lipids Genetics Consortium using MANTRA. In this analysis, we identified (logBF >6.1) 30 additional novel lipid loci. Lead signals for each trait were found at *FOLH1* (HDL-C), *MICA* (LDL-C), *KLF12* (TG), and *HLA-B* (TC). The novel HDL locus *NID2-PTGDR* could prove clinically useful, as *PTGDR* is a G protein coupled receptor, a common drug target. Taken together, in the largest East Asian genome-wide analysis of lipids to date, we identified multiple novel lipid loci, indicating new potential therapeutic targets.

356T

Genome-wide association study of abdominal fat depots in the UK Biobank identifies adiponectin as a candidate gene for visceral fat distribution. *A.M. Yiorkas[1], E.L. Thomas[7], M. Borga[4,5,6], J.D. Bell[7], O. Dahlqvist Leinhard[3,5,6], A.I. Blakemore[1,2].* 1) Section of Investigative Medicine, Imperial College London, London, United Kingdom; 2) Department of Life Sciences, Brunel University, London, United Kingdom; 3) Department of Medical and Health Sciences, Linköping University, Sweden; 4) Department of Biomedical Engineering, Linköping University, Sweden; 5) Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden; 6) Advanced MR Analytics AB, Linköping, Sweden; 7) Department of Life Sciences, University of Westminster, London, United Kingdom.

Introduction: Abdominal fat distribution has been recognized as a major determinant of metabolic health. In particular, visceral adipose tissue (VAT) is associated with increased risk of type 2 diabetes and cardiovascular disease, independently of BMI. It is also appreciated that the pattern of fat distribution is influenced by genetic factors, generally, to a larger extent than overall body adiposity. Whole-body magnetic resonance imaging (MRI) was used to measure abdominal fat depots on ~9000 UK Biobank (UKBB) participants. To our knowledge, this is the first GWAS using MRI measurements to identify genetic loci associated with abdominal adipose tissue distribution. **Methods**: A genome-wide association study of VAT, abdominal subcutaneous adipose tissue (ASAT) and visceral-to-subcutaneous ratio (VSR) was performed on European UKBB participants with available genotyping and MRI data (n = 1,493). An additive genetic model was considered, adjusting for age, gender, BMI, type 2 diabetes and the first five principal components. **Results:** We identified a novel locus (chr3; rs182052) in *ADIPOQ* (adiponectin) associated with VAT at the genome-wide significance level (p = 5.00×10^{-8}). The lead SNP was previously associated with circulating adiponectin levels and metabolic phenotypes. While no SNPs reached genome-wide significance in ASAT and VSR analyses, rs117388882 (p = 1.50×10^{-7}) in *CAV1* (caveolin 1) and rs74393452 (p = 1.54×10^{-6}) in *SCTR* (secretin receptor) were the lead SNPs respectively. These potential contributors to fat distribution are important key players since *ADIPOQ* is related to T2D and metabolism, *CAV1* is linked to lipodystrophy and *SCTR* knock-out mice (SctR$^{-/-}$) have decreased lipolysis. **Conclusions**: A genome-wide association study using MRI data on abdominal fat depots in UKBB participants has identified adiponectin SNPs as a contributor to VAT levels. Although replication is required to validate the suggestive hits for SAT and VSR, there is compelling biological evidence to support their involvement in abdominal adipose tissue homeostasis.

357F

Detecting and prioritizing triglycerides-associated genes by using gene-based method and bioinformatics analysis. *H. Zhang[1,2], X. Mo[1,2,3], S. Lei[1,2,3], Y. Zhang[1,2].* 1) Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, University, Suzhou, China; 2) Department of Epidemiology, School of Public Health, Medical College of Soochow University, 199 Renai Road, Suzhou, China; 3) Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, 199 Renai Road, Suzhou, China.

Background: Plasma concentration of triglycerides (TG) is reemerging as an important cardiovascular disease risk factor. The aim of this study was to identify more related genes for TG and evaluate the functional relevance to provide evidences for prioritizing these genes. Methods: We performed an initial gene-based association study in about 188,578 individuals (a publicly available data from the Global Lipids Genetics Consortium). Further, we performed Gene Relationships Across Implicated Loci (GRAIL), protein-protein interaction, coronary artery disease association analysis and other bioinformatics analyses to support the significance of the identified genes. Results: About 22,084 genes on the human genome were analyzed for TG levels in gene-based association analysis. A total of 298 genes were found to be statistically significant after Bonferroni correction ($P < 2.3 \times 10^{-6}$). These genes involved in important lipid related pathways or have been reported to be associated with many related diseases or traits. The evidence obtained from the analyses of this study signified the importance of many genes, e.g., *APOB*, *LPL*, *NR1H3*, *PLA2G6*, *TNF*, *VEGFA*, *APOA5*, *CETP*, *LIPC*, *MET*, *PEPD*, *PLTP*, *TNXB*, *LACTB*, *PARP6*, *BPTF*, and so on. These genes should be suggested as important candidates for further studies. Conclusion: Taken together, the present study found some important TG-associated genes and the findings might provide more insights into the genetic basis of lipid metabolism.

358W

Meta-analysis of quantitative pleiotropic traits at gene level with multivariate functional linear models. *C. Chiu[1], J. Jung[2], W. Chen[3], D. Weeks[3], H. Ren[4], J. Mills[1], M. Ting Lee[5], M. Xiong[6], R. Fan[1].* 1) Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892; 2) National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20892; 3) The University of Pittsburgh Medical Center, Pittsburgh, PA 15224; 4) DataParadise Inc., Belle Mead, NJ 08502; 5) University of Maryland, College Park, MD 20740; 6) University of Texas School of Public Health Houston, Texas 77225.

For a meta-analysis of multiple studies, multivariate functional linear models are developed to connect genetic variant data to multiple quantitative traits adjusting for covariates. The goal is to take the advantage of both meta-analysis and pleiotropy analysis in order to improve power and to carry out a uni ed association analysis of multiple studies and multiple traits of complex disorders. Three types of approximate F-distributions based on Pillai-Bartlett trace, Hotelling-Lawley trace, and Wilks's Lambda are introduced to test association between multiple quantitative traits and multiple genetic variants. Simulation analysis is performed to evaluate the false positive rates and power performance of the proposed models and tests. The proposed methods were applied to analyze lipid traits in eight European cohorts.

359T

Bi-ethnic GWAS refines genetic architecture of membranous nephropathy. *N. Mladkova[1], J. Xie[2], C. Sidore[3], M. Zoledziewska[3], R. LeDesma[1], J. D'Addario[1], N. Lester[1], M. Bodria[1], F. Scolari[4], A. Pani[5], F. Cucca[3], N. Chen[2], K. Kiryluk[1].* 1) Dept of Medicine, Columbia University , NYC, NY; 2) Institute of Nephrology, Department of Nephrology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy; 4) Division of Nephrology, Azienda Ospedaliera Spedali Civili di Brescia, Montichiari Hospital, University of Brescia, Brescia, Italy; 5) Department of Nephrology and Dialysis, G. Brotzu Hospital, Cagliari, Italy.

   Background: Idiopathic membranous nephropathy (iMN) represents the leading cause of nephrotic syndrome worldwide. iMN is characterized by autoimmune deposits in the glomerular basement membrane. The clinical course of iMN is highly variable, but up to one third of cases progress to end stage renal failure. Common variants at *PLA2R1* and *HLA* loci have previously been associated with iMN in Europeans, but these signals have not yet been fine-mapped. Notably, PLA$_2$R (encoded by *PLA2R1*) represents a known target of pathogenic antibodies in up to 70% of iMN cases. **Methods**: We performed a new GWAS discovery in 1,465 East Asians and 1,685 Sardinians (654 biopsy-diagnosed iMN cases and 2,496 ethnically-matched controls). The Asian cohort was genotyped with the HumanOmniZhongHua-8 chip while the Sardinian cohort with Illumina OmniExpress. Following stringent QC, genotype data were imputed using 1000-G reference (Minimac3). Association testing was performed using a dosage method in PLINK with adjustment for significant PCs of ancestry. Imputation of classical HLA alleles was performed with SNP2HLA. Ethnicity-specific results were meta-analyzed using METAL. **Results**: We replicated strong signals within the *HLA* region on chr.6p21 (OR=5.56, P=9.6E-65) and *PLA2R1* locus on chr.2q23 (OR=3.23, P=2.1E-24). We also confirmed a significant genetic interaction between these two loci (P=0.0014). We next performed conditional haplotype analyses to define at least 3 independently associated HLA signals and a single genome-wide significant *PLA2R1* haplotype. The analysis of classical HLA alleles revealed that the HLA signal is explained by *HLA-DRB1*1501* (OR=5.34, P=7.3E-47), as well as Arg at position 233 of *DRB1* (OR= 5.24, P=1.6E-07) and Arg at position 74 of *DRB1* (OR= 9.0, P= 2.2E-19). After conditioning on the *HLA* and *PLA2R1* loci, we identified a number of new suggestive loci across the genome. These are presently being tested for replication in >1,200 additional cases. **Conclusions**: In our bi-ethnic GWAS for iMN, we confirmed the previously implicated common variants with large effects, fine-mapped the MHC signal to a single gene, demonstrated that the association at the *PLA2R1* locus can be explained by a single risk haplotype, and identified several novel suggestive signals. Our genetic results confirm that the interaction between the antigen (PLA$_2$R) and *HLA-DRB1* is critical for the development of iMN, confirming strong autoimmune component to the disease pathogenesis.

360F

Dosage contribution of a non-classical HLA gene, *HLA-DOA***, to the risk of rheumatoid arthritis.** *Y. Okada[1,2,3], A. Suzuki[2], K. Ikari[4], C. Terao[5,6], Y. Kochi[2], K. Ohmura[6], M. Akiyama[2], K. Ashikawa[2], M. Kanai[2], J. Hirata[1,3], N. Suita[1], Y.Y. Teo[7], H. Xu[8], S.C. Bae[9], A. Takahashi[2], Y. Momozawa[2], K. Matsuda[10], S. Momohara[4], A. Taniguchi[4], R. Yamada[6], T. Mimori[6], M. Kubo[2], M. Brown[11], S. Raychaudhuri[5], F. Matsuda[6], H. Yamanaka[4], Y. Kamatani[2], K. Yamamoto[2,12].* 1) Osaka University Graduate School of Medicine, Osaka, Japan; 2) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Tokyo Medical and Dental University, Tokyo, Japan; 4) Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan; 5) Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 6) Kyoto University Graduate School of Medicine, Kyoto, Japan; 7) National University of Singapore, Singapore; 8) The Second Military Medical University, Shanghai, China; 9) Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea; 10) Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 11) University of Queensland Diamantina Institutem, Brisbane, Queensland, Australia; 12) Graduate School of Medicine, the University of Tokyo, Tokyo, Japan.

   Despite the progress in human leukocyte antigen (HLA) causal variant mapping, independent localization of major histocompatibility complex (MHC) risk from classical HLA genes is challenging. Here, we conducted a large-scale MHC fine-mapping analysis of rheumatoid arthritis (RA), an autoimmune disease with chronic destruction of synovial joints, in Japanese (6,244 RA cases and 23,731 controls) using the HLA imputation method. We further conducted a multi-ethnic validation study by including east Asians and Europeans (*n* = 7,097 and 23,149, respectively). Our study identified a risk of a synonymous mutation at *HLA-DOA*, a non-classical HLA gene, on anti-citrullinated-protein-autoantibody (ACPA)-positive RA risk (*P* = 1.4 × 10^{-9}), independently from the classical HLA genes (*HLA-DRB1*, *HLA-DPB1*, and *HLA-B*). The *HLA-DOA* risk variant demonstrated a cis-expression quantitative trait loci (cis-eQTL) effect on *HLA-DOA* expression levels, demonstrating its dosage expression effect on RA risk. Trans-ethnic comparison revealed different linkage disequilibrium (LD) patterns between *HLA-DOA* and *HLA-DRB1*, which explains the observed *HLA-DOA* variant risk heterogeneity among ethnicities; which was most evident in Japanese but not in Europeans. Whilst the previous HLA fine-mapping studies have identified amino acid polymorphisms of the classical HLA genes as driving genetic susceptibility of the diseases, our study additionally identifies the dosage contribution of a non-classical HLA gene to disease etiology. Our study contributes to understanding of HLA immunology in human diseases, and suggests the value of incorporating additional ancestry in MHC fine-mapping.

361W

Genotypic variability based association of seronegative rheumatoid arthritis identifies *DHCR7* **as a key non-additive locus without significant additive effects.** *W. Wei[1,2], S. Viatte[1], A. Barton[1,3], S. Robertson[2], J. Worthington[1,3].* 1) Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PT, UK; 2) Department of Women's and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin 9016, New Zealand; 3) NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

Seronegative rheumatoid arthritis [MIM 180300] is a complex and heterogeneous disorder where patients have no antibodies against citrullinated peptides (RA[-]). Genome-wide association studies (GWAS) of RA[-] have identified only two genome-wide significant loci *HLA-DRB1* [MIM 142857] and *ANKRD55* [MIM 615189]. It is therefore of interest to examine if there are any non-additive signals. We recently adapted genotypic variability based genome-wide association study (vGWAS) to dichotomous disease phenotypes, to allow identification of potentially interacting loci without prior knowledge of the interacting factors in two steps: using a mixed model approach to partition dichotomous phenotypes into additive risk and non-additive environmental residuals on the liability scale and then the Levene's (Brown-Forsythe) test to assess equality of the residual variances across three genotype groups per marker. We performed a vGWAS of a combined population of six RA[-] cohorts recruited from the UK, US, Sweden (2), Netherland and Spain, respectively, all genotyped with Immunochip. After quality control and the mixed model step, 19,108 unrelated samples (3323 cases and 15,785 controls) and 107,144 autosomal SNPs were used in the analyses, with an estimate of polygenic heritability of 0.045. At the genome-wide cutoff of 5.0e-08, vGWAS identified *HLA-DRB1* (lead SNP rs9275428, P = 2.0e-12) and *DHCR7* [MIM 602858] (lead SNP rs2852853, P = 1.3e-08). In contrast, the GWAS P values of the lead SNPs rs9275428 and rs2852853 were 1.5e-24 and 5.1e-05 respectively, suggesting *HLA-DRB1* can be detected from both analyses but *DHCR7* only from vGWAS. We further performed cross-validation iteratively by randomly splitting the combined population into two halves without changing any phenotypes, doing vGWAS in both sub-populations, and recording any discovery and/or replication successes. In a total of 10 iterations, *HLA-DRB1* was discovered and replicated every time whereas *DHCR7* narrowly missed discovery only once (P = 8.2e-08) but was replicated in all instances. *DHCR7* has an important role in vitamin D metabolism and *DHCR7* mutations could have been adaptive for habitation of early humans at Northern latitudes (e.g. Sweden) with insufficient sunlight and hence risk of vitamin D deficiency. Indeed we detected a significant (P = 0.005) interaction between the *DHCR7* lead SNP and cohorts fitted as an environmental factor. We conclude vGWAS is useful to discover novel non-additive signals.

362T

Joint prediction of gene expression using multimodal functional genomic readouts. *D. Lituiev, J. Ye.* Institute for Human Genetics, University of California, San Francisco, San Francisco, CA.

Although there has been much success in mapping genetic determinants of complex human traits, the prediction of traits using genetic data alone has proved challenging. One way to better understand the predictive capacity of genetic variation is by studying intermediate phenotypes such as gene expression and chromatin states. The advantage of this approach is that gene expression is less likely to be confounded by environmental factors than disease state. Here we explore several models incorporating multimodal genomic and genetic data to predict gene expression. In the first analysis, we develop a sparse generalized regression approach to jointly predict gene expression (GE) from genetic data and chromatin accessibility data. We use elastic net regression with distinct penalty for chromatin accessibility (CA) and genotype (GT) features. We apply our model to CD4+ T cells, where we have obtained ATAC-seq (CA) and RNA-seq (GE) data, and lymphoblastoid cell lines (LCLs), where published DNase-seq (CA) and RNA-seq (GE) data is readily available. Using an unbiased estimate of out-of-sample prediction performance, we estimated an average R^2 of 10% and 2% for those genes with positive R^2 (440/19518 in CD4s) and (199/12213 in LCLs) using GT alone. Including both GT and CA features increased the number of genes with expression predictable at a positive R^2 by 6% and 7% respectively, raising mean out-of-sample R2 by 3.6% and 9.1% respectively. These results suggest that although chromatin accessibility data carries predictive information about gene expression, the experimental noise and cost-limited cohort sizes (N~100 in both datasets) restricts its predictive potential. In the second analysis, we model interactions of *cis* and *trans* elements by utilizing transcription factor expression and genotypes to predict gene expression in ImmVar Nanostring data from dendritic cells. Including expression of 39 transcription factors as covariates raised the number of genes predictable with positive R^2 by 10% (58 to 64/405), improving out-of-sample prediction R^2 on average by 9% for genes with positive R^2. Additionally, we discovered 335 novel transcription factor — genotype interaction using linear regression at FDR < 0.01. Thus we demonstrate that sparse regression models involving genetic and functional genomic inputs can lead to improvements in prediction of gene expression.

363F

Common *HLA* alleles in patients candidates for hematopoietic stem cell transplant in western Mexico. *C. Ortega-de-la-Torre[1], H.J. Pimentel-Gutiérrez[1], A. Corona-Rivera[1,3], A. Márquez-Mora[1], F. Sánchez-Zubieta[2], S.A. Brukman-Jiménez[3], J.R. Corona-Rivera[4], L. Bobadilla-Morales[1,3].* 1) CYTOGENETICS UNIT, HOSPITAL CIVIL DE GUADALAJARA "DR. jUAN I MENCHACA", Guadalajara, Jalisco, Mexico; 2) PEDIATRIC HEMATO-ONCOLOGY SERVICE, PEDIATRICS DIVISION, HOSPITAL CIVIL DE GUADALAJARA "DR. JUAN I MENCHACA", Guadalajara, Jalisco, México; 3) cYTOGENETICS, GENOTOXICITY AND BIOMONITORING lAB, INSTITUTO DE GENETCA HUMANA "DR. ENRIQUE CORONA RIVERA", UNIVERSITY OF GUADALAJARA, Guadalajara, Jalisco, México; 4) PEDIATRICS DIVISION, HOSPITAL CIVIL DE GUADALAJARA "DR. J MENCHACA", Guadalajara, Jalisco, México.

INTRODUCTION. Human leukocyte antigens (HLA) genes are the most polymorphic in the genome and they are in the Major Histocompatibility Complex (MHC) located in 6p21.3 locus. For choosing hematopoietic progenitors (HP) donor should be considered HLA genes. PURPOSE. To determine the most common allele frequencies of HLA genes class I and class II in DNA samples of HP transplant candidates and their potential donors. MATERIAL AND METHODS. This is a descriptive, retrospective and transversal study. 134 transplant candidates and 411 potential donors, collected between 2012 and 2015 in the New Civil Hospital of Guadalajara were included in this study. DNA extraction was performed using purification columns (Qiagen) and analysis of HLA class I (A , B , and C) and class II (DQB1 and DRB1) was performed by PCR-SSO (One Lambda), and / or SSP (One Lambda), and / or SBT (Protrans) . RESULTS AND CONCLUSIONS. Here are mentioned the most common three alleles for each locus and their frequency and percentage. HLA-A: *02 with 311(29%), *24 with 171(15.9%) and *68 with 136 (12.7%); HLA-B: *35 with 197(18.8%), *40 with 102 (9.7%), and *44 with 99 (9.4%); HLA-C: *04 with 229 (21.7%), *07 with 225 (21.3%) and *03 with 150 (14.2%); HLA-DQB1: *03 with 510 (47.8%), *04 with 149(14%), and *06 with 134 (12.6%). HLA-DRB1 *04 with 310 (29%), *08 with 149 (14%) and *07 with 92 (8.6%). In this study was determined the HLA alleles frequency for this population of patients eligible for stem cell transplantation and their potential donors, they correspond to population of western Mexico. The three most common alleles for each locus were identified. The distribution of HLA alleles varies according to the population, so it is necessary to know the allelic frequencies in populations .

364W

Novel shared genomic segment analysis identifies *PARK2* and *ARID1A* as potential risk genes for myeloma. *R.G. Waller[1], T. Darlington[2], K. Curtin[3], D. Atanackovic[3,4], N.J. Camp[1,3,4].* 1) Biomedical Informatics, University of Utah, Salt Lake City, UT; 2) Psychiatry, University of Utah, Salt Lake City, UT; 3) Internal Medicine, University of Utah, Salt Lake City, UT; 4) Hematology, University of Utah, Salt Lake City, UT.

Identifying genetic variants involved in common disease remains a challenge due to multiple genes and sources of heterogeneity, reduced penetrance and sporadic cases. Variant prioritization often relies heavily on rarity and functional impact. Few studies have integrated formal statistical evidence for familial segregation, which could be especially useful in non-coding regions. Shared Genome Segment Analysis (SGS) provides statistical evidence that a genomic region is segregating in a high-risk pedigree (HRP) and indicates regions involved in disease risk. In SGS, chromosomal segments shared longer than expected by HRP cases are likely inherited from a common founder. Here, we use a modified heterogeneity-SGS algorithm that specifically assesses shared segments for all possible subsets of HRP cases. Thresholds, fully-corrected for multiple testing, were determined by optimizing over subsets with the Generalized Extreme Value distribution, accounting for genome-wide (GW) scan with the law of large deviations and correcting for the number of HRPs. We performed SGS using high-density SNP array data on 7 Myeloma (MM) HRPs, a highly heritable cancer of plasma cells. Each pedigree had at ≥ 3 genotyped MM cases with 16-23 meioses between cases. Six regions were identified with statistical evidence at the fully-corrected, GW suggestive level (Table 1). Three overlapped (24-855kb) with regions identified in a 2nd MM HRP. Of note, the overlapping chr18 region is within *PARK2*, a gene previously shown to harbor germline risk for lung cancer (shown to cluster in MM HRPs). To discover additional regions with multiple pedigree support, we added 3 smaller HRPs (13 meioses between 3 cases each) and considered regions that were pedigree-specific, GW suggestive. Four regions overlapped in ≥ 2 HRPs (Table 2). Of note, the chr1 region contains *ARID1A*, a gene somatically mutated in many cancers and associated with accelerated tumor growth. In sum, this study illustrates SGS can direct the search for the genetic basis of complex traits.

SITE	LENGTH (MB)	CASES	P-VALUES
Table 2-Regions significant in >2 HRPs at the GW suggestive level.			
1p36.11	1.16	3,3	3x10-4,2x10-4
10q23.33	0.02	3,4,3	2x10-4,2x10-4,7x10-4
16q12.2	0.24	3,3	4x10-4,4x10-4
18q22.2	0.21	3,4	4x10-5,9x10-5

365T

A novel statistical framework for meta-analyzing overlapping samples prevents a striking phenomenon that gold standard approach severely loses power when a subset of controls are shared. *E. Kim[1,2], B. Han[3]. 1)* Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea; 2) Seoul National University, Seoul, South Korea; 3) Department of Convergence Medicine, University of Ulsan College of Medicine and Asan Institute for Life.

   In order to maximize statistical power of genome-wide association studies (GWAS), current studies frequently augment control samples from public database or other GWAS. As a result, many different studies contain shared control samples in their analysis, which makes the samples of these studies dependent to each other. This dependency can cause problems in meta-analysis where we combine multiple independent studies into one. To solve this challenge of overlapping samples in meta-analysis, Lin and Sullivan proposed a novel strategy that calculates and accounts for the correlation between statistics induced by overlapping samples (AJHG 2009), which has become the gold standard (LS method). Here, we report a striking phenomenon regarding the power of this gold standard approach. We observed that when not all but a subset of controls are shared between studies, the gold standard approach shows severe power drop compared to splitting, where splitting refers to a strategy that splits genotype data of overlapping samples into individual studies. For example, in simulations performing meta-analysis of ten studies that share a half of controls, LS strikingly showed 10% power while splitting achieved 91% power (simulation codes available at the web address below). Intuitively, the power drop is caused by heterogeneity of information that each sample contains; it can be shown that shared samples have less information than study-specific samples towards the final summary statistic. Inspired by this phenomenon, we propose a novel power-preserving meta-analysis framework for overlapping samples, called FOLD (Fully-powered method for Over-Lapping Data). To prevent power drop, we obtain a set of summary statistics from each study where each statistic is calculated using a subset of controls that is homogeneous in terms of their information, and combine them into one final statistic. FOLD does not suffer from the power loss phenomenon of LS and achieves similarly higher power to splitting in all conducted simulations. We applied our method to the Wellcome Trust Case Control Consortium data. In the analysis combining three autoimmune diseases with partially overlapping shared controls, LS attenuated the statistical significances of previously reported loci compared to splitting, while FOLD preserved the significances. FOLD and our simulation codes are publicly available at http://software.buhm-han.com/FOLD.

366F

Efficient exact tests for longitudinal microbiome studies with application to HIV lung microbiome studies. *J. Zhou[1], J. Zhai[1], T. Hu[3], H. Zhou[2]. 1)* University of Arizona, Tucson, AZ; 2) University of California, Los Angeles, CA; 3) North Carolina State University, NC.

   Recent improvements in high-throughput sequencing make the longitudinal sampling of a relatively large number of individuals possible. Longitudinal studies that record the temporal variation of microbial communities in a wide range of environments allow us to observe temporal dynamics and its links to changes in human health status. For cross-sectional designs, many statistical methods and softwares analyzing association between microbiome and clinical outcomes have beed developed recently in addition to the traditional distance-based analysis. Among them, versatile regression based methods that directly regresses the outcome on the microbiome profiles while allowing for easy covariate adjustment enjoys great popularity. However, methods and softwares for longitudinal micriobiome studies that can deal with temporal variability are limited. In this paper, we propose efficient exact tests (score, likelihood ratio, and restricted likelihood ratio) for longitudinal microbiome studies (exVCLong). Our method is based on linear mixed effect models and combine with semi-parametric kernel association test to test the association of overall longitudinal microbiome profile and outcome. Additionally, our methods can also be used to detect association of one specific microbiome cluster while adjusting for the effects from related clusters. This can be particularly useful to find causal micribiome clusters. Longitudinal studies tend to have smaller sample sizes. Our 'exVCLong' method does not rely on asymptotic theory of large samples, which can significantly improve power to detect association for sample sample microbiome studies. Our simulation studies present correctly controlled type I error and superior power in detecting association in longitudinal studies. Finally, we applied our method to a longitudinal pulmonary microbiome datasets of Human Immunodeficiency Virus (HIV) patients. We identified two interesting genera Prevotella and Veillonella associated with force vital capacity, which shed lights on biological mechanisms. The software package is implemented in the open source, high-performance technical computing language Julia and is freely available at https://github.com/Tao-Hu/VarianceComponentTest.jl.

367W

A profound effect of common *HLA* polymorphisms on the survival of the patients with HIV infection in the pre-antiviral era in Thailand. *M. Yasunami[1,2], N. Wichukchinda[3], R. Miyahara[2], N. Tsuchiya[2], H. Nakamura[2], M. Mori[2], A. Rojanawiwat[3], P. Pathipvanic[4], P. Sawanpanyalert[3], K. Ariyoshi[2].* 1) Medical Genomics, Life Science Institute, Saga-ken Medical Centre Koseikan, Saga, Japan; 2) Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 3) Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand; 4) Day Care Center, Lampang Hospital, Lampang, Thailand.

   Background: *HLA class I* polymorphisms are known to play a crucial role in the control of HIV and progression to AIDS. We previously reported that *HLA-B*35:05* and *B*57:01* were associated with lower viral load in the patients with chronic HIV infection who were enrolled to a longitudinal observational study in Lampang Hospital, a government referral hospital in Lampang, Thailand. In addition to the *HLA class I*-mediated cytotoxic T lymphocytes, a variety of immune/inflammatory responses are regulated by the genes in the *HLA* region of human chromosome 6, so that the other *HLA* polymorphisms may associate with disease outcomes of the patients with HIV infection. **Methods:** A total of 556 HIV-1 CRF01_AE infected individuals were recruited from July 2000 to October 2002, and census data of survival was obtained on October 15, 2004 before the national program of antiviral treatment was introduced. CD4 count, viral RNA load and other information on clinical status were collected upon enrollment. Genotypes of *HLA-A, -B, -C, -DRB1* and *TNF* were determined. **Results:** Unlike *class I HLA* alleles, no effect of *TNF* polymorphisms on chronic phase viral load was detected. *TNF -308A* allele exhibited dominant unfavorable effect on the survival (*TNF -308AA+AG* vs *-308GG*, Cox hazard ratio: 1.46, p=0.034), and the effect was strengthened in HIV-1 positive patients with low CD4 count at enrollment (Cox HR: 1.52, p=0.024). Similar dominant unfavorable effect on the survival in the patients with low CD4 count was observed with *DRB1*03:01* (*DRB1*03:01*-positive vs *DRB1*03:01*-negative, Cox HR: 1.55, p=0.018) The effect of *TNF -308A* was enhanced by co-existence of *DRB1*03:01* allele (Cox HR: 1.65, p=0.010), suggesting a haplotype effect. Because of tight linkage between *TNF -308A* and *DRB1*03:01* alleles forming ancestral haplotype *AH58.1* commonly found in Asian population, it could not be distinguished whether the effects of *TNF/ HLA-DRB1* polymorphisms are primary or secondary to other genetic factor(s) carried by *AH58.1*. Interestingly, *AH58.1* shares the same alleles in *TNF-HLA-DRB1* interval with *AH8.1* which carries a *complement factor 4 null allele*. Shorter survival of *AH58.1*-positive patients might be elucidated by defective complement functions which are important to cope with opportunistic infectious agents such as *Pneumocystis*, etc. **Conclusion:** Unfavorable effect on the survival of the patients with chronic HIV infection in pre-ART era was associated with a common *HLA* haplotype *AH58.1*.

368T

NOD2 effect on inflammatory bowel disease is mediated by commensal microbiota. *H. Aschard[1,2], E. Tchetgen Tchetgen[1], D. Knight[3,4,5,6], P. Seksik[7], N. Zaitlen[8], J. Cosnes[7], R. Xavier[4,5,9], L. Beaugerie[7], D. Skurnik[10,11], H. Sokol[7,12,13].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 2) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 3) Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA; 4) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Center for Computational and Integrative Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 6) Biotechnology Institute, University of Minnesota, St. Paul, MN USA; 7) Department of Gastroenterology, Saint Antoine Hospital, Paris, France; 8) Department of Medicine, University of California, San Francisco, CA, USA; 9) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 10) Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 11) Massachusetts Technology and Analytics, Brookline, Massachusetts, USA; 12) Sorbonne University—UPMC Univ Paris 06, INSERM ERL 1157, Avenir Team Gut Microbiota and Immunity, UMR 7203, Saint-Antoine Hospital, AP-HP, UPMC Univ Paris 06, Paris, France; 13) Micalis Institute, INRA, AgroParis-Tech, Université Paris-Saclay, Jouy-en-Josas, France.

   Most genetic analyses and twin studies published to date support a genetic component to the inflammatory bowel disease (IBD). The most recent genome-wide scan identified a total of 163 loci associated with IBD phenotypes. The identification of these loci has enhanced our understanding of the pathogenesis of IBD, providing perspective on key pathways, and highlighting an essential role for host defense against infection. In parallel, recent work pointed out a potential role of the intestinal microbiota in initiating and maintaining IBD-related phenotypes. While it is generally accepted that the commensal microbiota can induce inappropriate activation of intestinal mucosal immunity, the precise role of the microbiota in the etiology of IBD in regards of the host genetic susceptibility remains unclear. In this study we used 182 well phenotyped IBD cases to explore the relationship between the intestinal microbiota and genetic variants from five of the most established IBD-associated genes (*NOD2, CARD9, LRRK2, ATG16L1* and *IL23R*). We performed a systematic screening for association between the four IBD-increasing genetic variants and bacterial quantification using a novel multivariate approach we recently developed. We identified 7 significant associations, 6 involving *NOD2*, and 1 involving *CARD9*. This includes in particular a negative association between *F. prausnitzii* and a genetic risk score for *NOD2*. *F. prausnitzii* is of particular interest in this context as we previously demonstrated its anti-inflammatory effect on cellular and colitis models. Focusing on this result, we first replicated the *NOD2-F. prausnitzii* association using an independent cohort of 474 IBD cases. We then demonstrated that this association remains after controlling for the established effect of *F. prausnitzii* and *NOD2* variants on IBD. Finally, using theory on causal inference and simulation, we showed that the increased IBD risk associated with *NOD2* is likely mediated through a decrease of *F. prausnitzii* in the gatstro-intestinal tract, showing for the first time an a three-way interdependency between bacteria, human genes and IBD. These results are of major importance for the understanding of IBD pathogenesis and for the discovery or new therapeutic targets on both host and microbiota sides. Moreover, this work has general implications as the concepts pointed out here could be applied to the growing number of other human diseases in which both microbiota and genetics are involved.

369F

Genome-wide analysis of vitamin D response elements implicates vitamin D regulation of antigen presentation in multiple sclerosis risk: Results from the Kaiser Permanente MS Research Program. *A. Mok[1,2], X. Shao[2], B. Rhead[1,2], L. Shen[3], H. Quach[2], C. Schaefer[3,4], T. Olsson[5], I. Kockum[5], L. Alfredsson[6,7], L.F. Barcellos[1,2,3,4].* 1) Computational Biology Graduate Group, UC Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genetics Laboratory, UC Berkeley, Berkeley, CA; 3) Kaiser Permanente Division of Research, Oakland, CA; 4) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA; 5) Department of Clinical Neuroscience and Center for Molecular Medicine, Karolinska Institutet at Karolinska University Hospital, Stockholm, Sweden; 6) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 7) Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden.

   Recent studies have demonstrated a causal role of vitamin D serum levels in multiple sclerosis (MS [MIM 126200]); however, the mechanisms underlying this association are not yet known. The vitamin D receptor is expressed in nearly all tissue types, but its target genes are specific to cell type and function. Thus, genome-wide analysis of vitamin D signaling may elucidate the downstream factors driving MS pathogenesis. We investigated whether genetic variants in putative vitamin D response elements (VDREs) are associated with MS in non-Hispanic White members of Kaiser Permanente Northern California (1,098 cases; 10,329 controls). Genotypes were obtained through whole-genome profiling and imputation. We identified 6,250 SNPs within 4,764 VDREs from six ChIP-seq datasets for analysis. To detect the effect of VDRE variants on MS susceptibility independent of well-established risk factors, we performed logistic regression for each SNP and controlled for sex, genetic ancestry, a weighted genetic risk score for 110 non-HLA risk alleles, and four independent SNPs in the MHC, the region with the strongest genetic contribution to MS risk. Results show significant disease associations for 19 SNPs in 16 VDREs (FDR q<0.05), including rs2213584 (OR=1.34, p=5.34e-07) and rs2395182 (OR=1.47, p=1.38e-06) in a VDRE 5kb downstream of *HLA-DRA*. We additionally restricted analyses to VDRE SNPs within 10kb of genes established through recent MS GWAS and found significant associations for rs1335532 (OR=0.83, p=1.97e-04) in *CD58* and rs9831894 (OR=0.76, p=5.96e-04) in *CD86*. As HLA-DRA, CD58, and CD86 are all cell surface proteins involved in antigen presentation, findings suggest that vitamin D regulation of antigen presentation may mediate MS risk. A majority (82%) of the MS-associated VDREs demonstrated specificity to B-lymphocytes, and none of the identified VDREs contained the canonical DR3 binding motif. Replication in an independent dataset of Swedish individuals (6,335 cases; 5,762 controls) is underway. We are performing high-throughput sequencing of each VDRE in a subset of samples to confirm findings, and conducting ChIP-seq analysis to identify candidate VDREs specific to T-lymphocytes for further study. We describe the first results from a comprehensive genomics approach to characterizing the contribution of vitamin D metabolism in MS. Our findings also have important implications for cancer and other autoimmune conditions more broadly, for which vitamin D plays a role.

370W

Shared effect modeling reveals that a fraction of autoimmune disease associations are consistent with eQTLs in three immune cell types. *S. Chun[1,2,3], A. Casparino[4], N.A. Patsopoulos[3,5,6], D. Croteau-Chonka[2,7], B.A. Raby[2,7], P.L. De Jager[2,3,5,6], S.R. Sunyaev[1,2,3], C. Cotsapas[3,4,8].* 1) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Neurology, Yale School of Medicine, New Haven, CT; 5) Department of Neurology, Brigham and Women's Hospital, Boston, MA; 6) Ann Romney Center for Neurological Diseases, Brigham and Women's Hospital, Boston, MA; 7) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 8) Department of Genetics, Yale School of Medicine, New Haven, CT.

   The majority of autoimmune disease risk effects identified by genome-wide association studies (GWAS) localize to open chromatin with gene regulatory activity. GWAS loci are also enriched for expression quantitative trait loci (eQTLs), suggesting that most disease risk variants exert their pathological effects by altering gene expression. However, because causal variants are difficult to identify and *cis*-eQTLs occur frequently, it remains challenging to translate this bulk observation into specific instances of a disease risk variant driving changes to gene regulation. We developed a new method, JLIM, to directly evaluate the joint likelihood that a disease association and an eQTL are due to the same underlying causal variant. We assess its performance by benchmarking it on simulated data against *coloc*, a well-calibrated Bayesian framework that considers spatial similarities of association data across windows of markers. Both methods have excellent performance when two distinct causal variants are not in LD (AUC=0.99 for both when $r^2 < 0.5$), but JLIM maintains higher specificity as LD between distinct causal variants increases (AUC = 0.92 vs 0.70 for *coloc* when $0.7 < r^2 < 0.8$). We next compared association signals to six autoimmune diseases in 188 loci densely genotyped on the ImmunoChip to eQTLs from LCLs, CD4+ T cells and CD14+ macrophages. All but 11 loci had at least one eQTL (p < 0.05) in at least one of the three cell types, with most found in all three tissues. We tested 9,268 pairs of disease and eQTL associations with JLIM and find evidence for shared effects for only 57/9,268 pairs (FDR < 0.05), with 43/57 surviving Bonferroni correction. Rather, JLIM strongly supports the alternative hypothesis for distinct effects for the other trait pairs. Overall, we find that only a fraction of disease-eQTL pairs in the same locus are due to the same underlying variant, whereas >75% are associated to distinct variants in the same locus. Thus, we uncover a fraction of gene regulatory changes as strong mechanistic hypotheses for disease risk, but conclude that most risk mechanisms do not involve changes to basal gene expression.

371T

Identifying substructure in genetic risk sharing between diseases. *L. Jostins, G. McVean.* Statistical Genetics, Wellcome Trust Centre Human Genetics, Oxford, Oxfordshire, United Kingdom.

While genetic risks for complex human diseases are often shared across multiple phenotypes, such as disorders of the immune system, this risk sharing shows complicated substructure. Risk alleles can be associated with multiple diseases with similar effect sizes or with smaller or opposite effects, or uniquely with one or a few diseases. Such complexity, along with incomplete power, makes inferring the exact sharing model for any given locus, and hence structure in sharing between loci, difficult.To address these problems we present a statistical approach which models the underlying biology driving genetic risk substructure across multiple loci and phenotypes. Motivated by a network model of shared and potentially pleiotropic risk components, we use a reduced rank representation of genetic covariance in which each risk allele belongs to one of a small number of pathways, where each pathway varies in complexity from simple perfectly correlated effects to more complex covariance models. The method adjusts the number and rank of pathways to infer the simplest model that explains observed association summary statistics. Individual loci are probabilistically assigned to inferred pathways, allowing us to examine the genes and functions associated with each pathway.Using simulations, we show the power and accuracy of the method to estimate underlying parameters and choose the correct pathway structure using the Bayesian information criterion. For example, we have high power to recover the structure of a 3-pathway model under realistic scenarios (100 loci, 5 diseases, 10000 total samples). We apply the method to association data for LDL cholesterol and cardiovascular disease (CVD), where we expect a direct causal relationship. We recover two pathways, a low rank pathway of loci that act on CVD via LDL, and a pleiotropic pathway that acts on CVD in an LDL-independent fashion. The genes in these two pathways are overrepresented in differing tissues (liver for the first, adipose and muscle for the second). We also apply the method to autoimmune diseases associations to recover significant substructure, including a low-rank pathway with opposite effects on type 1 diabetes and multiple sclerosis.The difficulties of cross-disease analysis scale exponentially with the number of diseases. In contrast, by providing simplified, biologically interpretable models of genetic risk sharing, our method could potentially analyse shared pathways across hundreds of diseases. .

372F

Trait agnostic identification of candidate genes associated with noncoding variation. *A. Wuster, D. Chang, T. Bhangale.* Genentech Inc., South San Francisco, CA.

The majority of variants identified by genome-wide association studies are located in noncoding regions and likely act by affecting gene expression. Variants typically contain multiple genes in the region of linkage disequilibrium and identifying the causative one is challenging. Many studies assume that the gene closest to the risk variant is the best candidate, which may often not be the case. Another approach is to look for associations between the variant of interest and the expression of nearby genes (eQTLs). In isolation, this has limitations too as eQTLs are not available for many tissues and disease states and as most variants are not strongly associated with gene expression. We have therefore developed a method that identifies the most likely gene or genes through which variation at a given locus acts. Using a Bayesian approach, our method integrates eQTL data from the Genome-Tissue Expression Project (GTEx), recombination data from the HapMap project, and published Hi-C genome conformation data. For any given locus, our method returns a list of likely candidate genes together with their relative probabilities. We compare our method to a previously published approach that made use of a variety of trait-specific data types to identify candidate genes at 101 rheumatoid arthritis risk loci (Okada et al. 2013, Nature 506, 376). Even though our method is trait agnostic, it identifies the same candidate gene for 84% of loci for which such a determination is possible. We also apply our method to 13,000 loci in the NCBI/EBI GWAS catalog and find that the resulting candidate gene probabilities closely reflect the probability that a gene has been annotated as a candidate gene in the catalog. This suggests that our method can accurately identify candidate genes in the majority of cases. While it is not a substitute for in-depth experimental validation, it can quickly suggest candidate genes without the need to first collect trait or disease specific data. In addition to returning candidate gene probabilities for single variants, our method can identify candidate genes for sets of variants that are in linkage disequilibrium by integrating the individual probabilities of those variants, thus making the identification of the most likely causative variant less of a priority.

373W

KING 2.0: Relationship inference and integrated analysis in one million samples. *W.-M. Chen[1,2], A. Manichaikul[1,2], S.S. Rich[1,2].* 1) Center for Pubic Health Genomics, University of Virginia, Charlottesville, VA; 2) Department of Public Health Sciences, University of Virginia, Charlottesville, VA.

The kinship coefficient between a pair of individuals can be accurately estimated using their genome-wide SNP data, without estimating the allele frequencies at each SNP in the whole dataset. Together with implementation of bit-wise operations that are applied directly on binary coded genotype data, relationship inference in the first generation of KING has been fast for genetic datasets consisting of 100,000 or fewer individuals. In order to infer related-ness and further to provide integrated genetic analyses in the latest genetic data where the number of individuals can be beyond 1 million, we develop the second-generation of KING. Three levels of computational speed-up are implemented in KING 2.0, including: 1) bit-level parallelism; 2) multiple-core parallelism using OpenMP; and 3) faster algorithms involving a multi-stage procedure to eliminate unrelated or distantly related pairs of individuals, in which we first select a subset of highly informative SNPs to obtain first-pass relationship estimates for all pairs of individuals, and then follow-up using all SNPs on a small subset of pairs identified in the first stage analysis. One particular computing technique for further speed-up is to propose computationally simpler estimators that are upper bounds of the KING estimator. The efficient implementation in KING 2.0 performs relationship inference in a matter of seconds in a typical dataset (with 10,000 or fewer samples), or a few hours in very large data consisting of ~1 million samples. This computational efficiency also applies to integrated genetic analyses that are available in KING 2.0, including automated pedigree reconstruction, extraction of subset of unrelated individuals, family-based Q/C without relying on reported pedigree data, as well as association scans for a massive number of phenotypes (*e.g., trans-* eQTL analysis). Lastly, we implement a new relationship inference algorithm that has a similar performance as PC-Relate (AJHG 98:127-148) while still feasible for very large data. To demonstrate the performance of KING 2.0, we applied relationship inference on a dataset consisting of 1,002,375 samples (generated by duplicating a real dataset consisting of 30,375 samples 33 times) each typed at 168,749 autosome SNPs. It took 2 hour and 13 minutes to identify all 16,090,272 pairs of duplicates, and 6 hours and 16 minutes to identify all 36,151,830 pairs of first-degree relative and duplicates, using 19 CPU cores.

374T

Mapping abundance of 1,129 serum proteins provides functional insights for disease risk alleles. *J. Maranville, Y. Huang, J. McElwee, U-BIOPRED Consortium.* Genetics and Pharmacogenomics, Merck Research Laboratories, Boston, MA.

Although disease GWAS have successfully identified thousands of loci associated with multiple diseases, the molecular mechanisms that underlie these associations are often unclear. Genetic association studies of molecular traits, such as protein abundance, complement genetic studies of disease risk to provide valuable insights into pathobiology. To expand our understanding of allelic effects on protein abundance traits, we mapped protein quantitative trait loci (pQTL) using 414 individuals collected as a part of the UBIOPRED study. SomaLogic assays were used to measure the abundance of 1,129 proteins in serum samples from each participant. This was combined with Axiom array and imputed genotype data to perform pQTL mapping in cis (defined as variants <1MB from encoding gene) and in trans (defined as variants >1MB from encoding gene). We found 269 proteins had at least one significant cis pQTL at a threshold of FDR<10^{-4}. At a Bonferroni-corrected threshold of 5x10^{-11}, we found that 55 proteins had at least one significant trans pQTL. These included multiple previously described QTL for protein abundance traits assayed using other technologies (e.g. ELISA). Additionally, pQTLs overlapped with GWAS signals for asthma and related phenotypes (e.g. allergy). For example, we replicated a cis-pQTL for the soluble IL-6 receptor that also shows a genome-wide significant association with asthma. Specifically, asthma risk alleles were associated with higher levels of soluble IL-6 receptor. This provides additional support to previous suggestions that inhibitors of the IL-6 receptor would be effective in treating asthma. Our study also highlighted novel protein-disease connections. For example, we found a cis-pQTL for the soluble IL-18 receptor α that overlaps a GWAS signal for asthma risk, with disease risk alleles also associated with higher levels of this protein. This suggests that inhibition of IL-18 receptor α could be a promising strategy for the development of novel asthma therapeutics. Overall, our results massively expand on previous studies by identifying QTL for hundreds of proteins and provide important insights into proteomic effects of disease risk alleles.

375F

Identification of SLE-associated risk variants in the *STAT1-STAT4* locus and their effect on differential transcription factor binding. *Z.H. Patel[1], X. Lu[2], D. Miller[3], E. Zoller[3], M. Weirauch[3,4], K. Kaufman[3,4,5], J.B. Harley[3,5], L.C. Kottyan[3,5], SLEGEN Consoritum & Collaborators.* 1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Immunology Graduate Program, College of Medicine, University of Cincinnati, Cincinnati, OH; 3) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Biomedical Informatics, College of Medicine, University of Cincinnati, Cincinnati, OH; 5) Department of Veterans Affairs, Veterans Affairs Medical Center – Cincinnati, Cincinnati, OH, USA.

   Systemic Lupus Erythematosus (SLE or lupus) is a chronic autoimmune disease with debilitating inflammation that affects multiple organ systems. The *STAT1-STAT4* locus is one of the first and most highly replicated genetic loci associated with SLE risk. In this study, we aimed to identify all the SLE-associated common variants at this locus most likely to be causal and to further identify the biological mechanism mediating the increased disease risk. We genotyped 328 SNPs spanning the *STAT1-STAT4* locus in 13,581 subjects representing four ancestral groups. We performed imputation and applied frequentist and Bayesian statistical analyses to identify the individual variants statistically most likely to causally increase lupus risk. We further used a separate larger African-American study to generate an Ancestry Informed Credible Set (AICS) of four variants. We computationally predicted differential transcription factor (TF) binding of AICS variants and identified the AT-hook family of TFs as a strong candidate for three of the four AICS variants. After identifying AT-hook family member HMGA1 as binding to rs11889341 through DNA Affinity Precipitation Assay (DAPA) followed by mass spectrometry, we confirmed binding of HMGA1 to two of four AICS variants with genotype-dependent binding by DAPA and Electrophoretic Mobility Shift Assay. In summary, we used large genetic datasets to identify a set of variants that are most likely to be causal for the *STAT1-STAT4* association with increased lupus risk and identified a potential disease-risk mechanism in which HMGA1 differentially binds three genetic variants in a lupus-risk haplotype.

376W

Quantifying directional effects of transcription factor binding on polygenic disease risk using GWAS summary statistics. *Y. Reshef[1], H. Finucane[3], A. Gusev[2], J. Ulirsch[4], B. van de Geijn[2], S. Grossman[4], P. Loh[2], G. Bhatia[2], S. Gazal[2], P. Palamara[2], N. Patterson[4], L. Dicker[5], L. Pinello[6], R. Adams[1], A. Price[2].* 1) Computer Science, Harvard University, Cambridge, MA; 2) Epidemiology, Harvard School of Public Health, Boston, MA; 3) Mathematics, Massachusetts Institute of Technology, Cambridge, MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Statistics and Biostatistics, Rutgers University, New Brunswick, NJ; 6) Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA.

   Connecting complex traits to specific transcription factors (TFs) represents an important path toward better mechanistic understanding of human disease. Existing genetic approaches to this problem focus on functional enrichment, e.g. evaluating whether SNPs in or around binding sites of a TF are enriched for heritability of a trait or over-represented among top GWAS SNPs. However, these approaches a) have limited power to detect enrichment at the resolution required for TF binding sites, and b) have limited mechanistic interpretability because TF binding may simply mark areas of the genome, such as certain enhancers or gene sets, whose importance for the trait is not directly related to the TF in question. To overcome these limitations we define a signed population quantity, *functional correlation*, which quantifies the extent to which the signed effect of SNPs on disease risk is correlated with their signed effect on TF binding. We introduce a method, FCor, that estimates functional correlation from GWAS summary statistics and provides closed-form standard errors for assessing significance. FCor can be viewed as a maximum-likelihood estimate of the genetic correlation between total TF binding and the phenotype in question, in a model in which SNP effects are fixed rather than random. Our simulations show that FCor attains higher power than unsigned functional enrichment approaches given accurate information about the direction of SNP effects on TF binding, and that FCor provides unbiased estimates of functional correlation and controls type I error.   We applied FCor to analyze the relationship of the transcription factor NFκB to rheumatoid arthritis (RA). Using RA summary statistics (N=38,242) together with an annotation of 4,805 NFκB occupancy-modifying SNPs, we found that the polygenic burden of alleles increasing NFκB occupancy is positively correlated with RA risk ($p < 7.9 \times 10^{-4}$). This finding translates into a genetic correlation of 3.2% (s.e. 0.97%) between total NFκB binding and RA, and is consistent with biomedical literature suggesting that increased NFκB activity is a component of RA pathogenesis. To our knowledge this has not previously been demonstrated using GWAS data and so our result constitutes new evidence for a mechanistic link between NFκB and RA. Our work illustrates the potential of functional correlation for improving understanding of human disease, and we plan to apply this framework to tens of phenotypes and TFs.

377T

Effect of some genetic variants changes over time post transplant on tacrolimus concentrations. *B. Wu[1], W. Guan[1], D. Schladt[2], K. Sanghavi[3], C. Dorr[2], A. Muthuswamy[2], R. Remmel[3], R. Mannon[4], A. Matas[5], A. Israni[2], W. Oetting[3], P. Jacobson[3], DeKAF Genomics Investigators.* 1) Division of Biostatistics, University of Minnesota, Minneapolis, MN; 2) Minneapolis Medical Research Foundation, Minneapolis, MN; 3) Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN; 4) Division of Nephrology, University of Alabama, Birmingham, AL; 5) Department of Surgery, University of Minnesota, Minneapolis, MN.

Previously we identified three genome-wide significant CYP3A5 variants (*3, *6, and *7) associated with the tacrolimus troughs (TT) in a GWAS of African American (AA) kidney transplant (tx) recipients (Oetting et al., Am J Transplant, 2016). When we estimated the regression coefficients of genotype at each time point posttx for these three variants, overall we found that the variants have non-constant effects on the TT over time posttx. We observed the largest changes of effects in the early days (day 1 to around week 3). Therefore we analyzed in detail our GWAS data with a new statistical model allowing for potentially non-constant variant effects. We evaluated dose-normalized TT and 426,960 common variants (MAF>0.05) genotyped using an Affymetrix TxArray, adjusting for age, gender, study center, top four ancestry principal components, and a set of previously identified important clinical variables. TT from 318 AA kidney tx recipients from three study centers in the DeKAF Genomics study were in the discovery cohort. Each recipient had multiple TT measurements (range 1-24, median 18) posttx from week 1 to month 6. We then validated the variants in a cohort of 35 AA samples from another two study centers. We identified 23 variants that passed genome-wide significance ($P<5x10^{-8}$) in the discovery cohort, and had test p-values $P<0.01$ in the validation cohort. When assuming constant genotype effects, we only identified 3 (located in CYP3A5 on chromosome 7) of the 23 significant variants. We identified variants that have varying effects over time on TT in AA recipients. Understanding and accounting for variants with changing effect sizes over time will allow for more precise tacrolimus dosing.

378F

Integrative whole genome sequencing and metabolomics networks in asthmatics identifies sphingolipid metabolites, genes, and networks associated with asthma severity. *J. Lasky-Su[1], R. Kelly[1], J. Hecker[2], H. Fier[2], D. Prokopenko[2], J. Celedon[3], C. Lange[4], S. Weiss[1,5], IMPACT consortium.* 1) Channing Division of Network Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 2) Institute of Genomic Mathematics, University of Bonn, Bonn, Germany; 3) Children's Hospital of Pittsburgh and University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Harvard School of Public Health, Boston, MA; 5) Partner's Healthcare Personalized Medicine, Brigham and Women's Hospital, Boston, MA.

In the U.S., asthma affects 26 million people and leads to estimated annual costs exceeding $12.7 billion. Given its strong public health impact, identifying key molecular determinants of asthma remains an important priority in translational science. The genetic determinants identified for asthma to date explain only a fraction of the estimated heritability. Metabolomic profiling enables a functionally relevant exploration of the relationships between the underlying genetic and environmental disease risk factors. We studied a cohort of children with asthma from Costa Rica (CR), which is rich in environmental and clinical measurements relevant to asthma. We generated comprehensive metabolomic profiling in 381 asthmatic children and whole- genome sequence (WGS) data on 1100 individuals from these CR families. For the WGS family data, novel family-based analytic methods that incorporate both individual variant and "region-based tests" , identifying several asthma-associated regions that met genome-wide significance. The most significant finding is in the 5' region of *PDE4DIP* on chromosome 1, the interacting protein for *PDE4D*, one of the first validated genes for asthma ($p=5.9x10^{-8}$). A total of 8185 metabolite features that covered a broad range of the metabolome were measured and passed QC in the 381 plasma samples. A total of 583 (7%) of features, most of which were free fatty acids and bile acids, were associated with asthma severity, defined by airway hyper-responsiveness to methacholine ($p<0.01$). The inclusion of a summary score based on these features (AUC: 0.685 (95%CI 0.625, 0.747)) significantly outperformed a baseline model in discriminating asthma cases by their degree of severity ($p=0.004$). This profile was enriched for metabolites of the sphingolipid, linoleic acid, arginine and proline metabolism and fatty acid biosynthesis pathways. To assess the interrelationships between genes and metabolites, we used a systems biology approach to integrate WGS and metabolomic data, both on an individual variant and pathway levels to identify a set of metabolites and genes that influence asthma. Findings in the aforementioned metabolites, genes, and pathways were validated using 3 populations, including 2 asthma birth cohorts and an asthma clinical trial. Taken together, these findings demonstrate the power of integrating WGS and metabolomics data together in the development of biomarkers of asthma severity and in the understanding of disease pathogenesis.

379W

Admixture mapping of kidney traits in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *L. Brown[1], T. Thornton[1], N. Franceschini[2]*. 1) Biostatistics, University of Washington, Seattle, WA; 2) Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Chronic Kidney Disease (CKD) affects approximately 14.5% of the adult U.S. population, with a high burden occurring among racial/ethnic minorities, including Hispanics. Albumin–creatinine ratio (ACR) is a quantitative measure of kidney function. Admixture mapping is a method for disease gene mapping in admixed populations that leverages allele frequency differences in ancestral reference populations to identify an association with locus-specific local ancestry. We implement a linear mixed model framework to perform admixture mapping of ACR in subjects from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). We identify a strong Amerindian ancestry driven association on chromosome 2 spanning q11.2-14.1, that is not found by GWAS. Our findings suggest the presence of ancestry-specific variants influencing the variation of urine albumin in admixed populations such as Hispanics.

380T

Genetic polymorphisms of *PNPLA3* and *SAMM50* gene are associated with severity of nonalcoholic fatty liver disease in Koreans. *G. Chung[1], Y. Lee[1], J. Yim[1], M. Kwak[1], C. Lee[1], E. Choe[1], J. Lee[2], J. Kim[2], H. Jang[2]*. 1) Seoul National University Hospital Gangnam Healthcare Center, Seoul, South Korea; 2) DNA Link, Inc., Seoul, Korea.

Background/Aims: Nonalcoholic fatty liver disease (NAFLD) is the most common liver disease. The development of NAFLD is associated with multiple genetic factors, environmental factors and lifestyle. **Methods:** We performed a genome-wide association study to identify genetic factors related to NAFLD in large Korean population based samples of 1,593 subjects with NAFLD and 2,816 controls using the Affymetrix Axiom® Customized Biobank Genotyping Arrays. We replicated the data in another sample including 744 subjects with NAFLD and 1,137 controls. In addition, we investigated SNPs related to severity of ultrasonography-diagnosed fatty liver and NAFLD with increased levels of alanine aminotransferase (ALT). **Results:** After adjusting for age, sex and body mass index, rs738409 located in the *PNPLA3* gene, a locus previously reported, was validated in our population (p=1.7 x 10^{-15}; OR, 1.54 in the discovery set and p=2.6 x 10^{-8}; OR, 1.54 in the replication set). Rs12483959 and rs2281135 were in strong linkage disequilibrium (LD) (r2=0.978 and 0.936) and showed significant associations (p=5.8 x 10^{-15}; OR, 1.53 and p=9.7 x 10^{-15}; OR, 1.52, in the discovery set and p=1.3 x 10^{-8}; OR, 1.56 and p=1.3 x 10^{-8}; OR, 1.51, in the replication set). Rs2143571, rs3761472 and rs2073080 in the SAMM50 gene were in strong LD (r2=0.932 and 0.996) and also showed significant associations with NAFLD (p=1.9 x 10^{-11}; OR, 1.43; p=2.2 x 10^{-11}; OR, 1.43; and p=2.7 x 10^{-11}; OR, 1.43, respectively in the discovery set and p=5.9 x 10^{-7}; OR, 1.48; p=3.3 x 10^{-7}; OR, 1.49; and p=7.8 x 10^{-7}; OR, 1.47, respectively in the replication set). Additionally, these six SNPs showed significant associations with severity of fatty liver (p=2.9 x 10^{-20}; p=2.0 x 10^{-19}; p=3.3 x 10^{-19}; p=1.8 x 10^{-15}; p=1.6 x 10^{-15} and p=3.8 x 10^{-15}, respectively in the discovery set p=1.3 x 10^{-8}; p=3.7 x 10^{-9}; p=3.2 x 10^{-8}; p=5.7 x 10^{-7}; p=1.3 x 10^{-7} and p=6.1 x 10^{-7}, respectively in the replication set) and NAFLD with elevated levels of ALT (p=8.4 x 10^{-17}; p=3.3 x 10^{-16}; p=5.9 x 10^{-16}; p=1.7 x 10^{-12}; p=9.7 x 10^{-13} and p=2.3 x 10^{-12}, respectively in the discovery set p=1.4 x 10^{-7}; p=7.8 x 10^{-8}; p=1.1 x 10^{-6}; p=1.7 x 10^{-6}; p=1.1 x 10^{-6} and p=1.5 x 10^{-6}, respectively in the replication set). **Conclusions:** We demonstrated that the PNPLA3 gene and SAMM50 gene are strongly associated with the presence and severity of NAFLD in Korean population. These findings confirm important roles of genetic factors in the pathogenesis of NAFLD.

381F

Genome-wide SNP interaction analysis suggests polygenic regulatory effects on bone density among adult survivors of childhood cancer. *C. Im[1], C.L. Wilson[2], K.K. Ness[2], W. Moon[2], S.C. Kaste[3,4], W. Chemaitilly[5], M.M. Hudson[2,4], L.L. Robison[2], Y. Yasui[1,2].* 1) School of Public Health, University of Alberta, Edmonton, Alberta, Canada; 2) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN; 3) Department of Radiological Sciences, St. Jude Children's Research Hospital, Memphis, TN; 4) Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN; 5) Division of Endocrinology, Department of Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN.

Interaction between genetic regulatory elements that modulate gene expression may explain some of the heritability that remains uncharacterized for many complex diseases and traits. We explored this hypothesis among childhood acute lymphoblastic leukemia (ALL) survivors for whom bone mineral density (BMD) may be reduced following cancer treatments received in childhood. We analyzed genotype data from 856 adult survivors of ALL (≥10 years from cancer diagnosis) to find SNP interaction patterns associated with age- and sex-standardized lumbar spine BMD Z-scores measured via quantitative computed tomography. After restricting SNPs to exclusively retain those in the regions of promoters and enhancers using ChromHMM regulatory states, our final genome-wide SNP set included 115,793 SNPs. We used a novel algorithm for chromosome-wide learning based on logic regression, an adaptive regression methodology that stochastically searches the space of fixed size interactions, to select the best 3-SNP interaction patterns. These interaction patterns, or 3-SNP Boolean logic trees, represent biologically plausible necessary or sufficient interaction conditions linked to BMD levels, accounting for sex, ancestry, and cumulative cancer treatment dosages. Our algorithm was applied to each chromosome, sequentially selecting 3-SNP Boolean logic trees conditioned on previously selected trees. We used 1,000 permutations to assess the statistical significance of each tree considering its selection order. This algorithm yielded eight 3-SNP trees with permutation-based p-values<0.05, of which five trees were validated in an independent sample of 1,428 non-ALL childhood cancer survivors, either as main or treatment interaction effects. All validated trees featured enhancer and promoter SNPs, none of which achieved genome-wide significance as single SNPs. The validated tree with the strongest statistical evidence of interaction was on chromosome 12 (mean BMD deviation, conditioned on first tree=1.72 SD, P=2.87x10[-13]). This tree may modify methotrexate-related effects on BMD (P=0.02, non-ALL) and implicates interactions between regions regulating the activities of *PFDN5* (repressor of c-Myc), *ANO2* (encodes CaCCs), and *DYRK4* (dual specificity kinase family member) to jointly affect BMD. In summary, our study demonstrates the utility of interaction association analyses in revealing new regulatory-element/gene-gene interaction mechanisms associated with complex traits.

382W

Two amino acids in HLA-B explain majority of the associations between HLA and cutaneous adverse drug reactions induced by phenobarbital and phenytoin in Japanese population. *T. Ozeki[1], T. Mushiroda[1], A. Takahashi[2], M. Kubo[3].* 1) Laboratory for Parmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Laboratory for Stastical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Anticonvulsants phenobarbital (PB) and phenytoin (PHT) are known to show incidences of cutaneous adverse drug reactions (cADRs) including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug-induced hypersensitivity syndrome (DIHS). To identify a gene(s) susceptible to PHT and PB-induced cADRs, we conducted a genome-wide association study (GWAS) and subsequent HLA typing in 22 cases of PB and PHT-induced cADRs (consist of 10 PB, 10 PHT and 2 PB / PHT treated patients) and 880 subjects of a general population in Japanese (GWAS) or 751 subjects of PB/PHT tolerant (HLA typing). Among the SNPs analyzed in the GWAS followed by imputation, 5 SNPs showed significant association with PB and PHT-induced cADRs, and rs34341880 showed the smallest *P*-value for association with PB and PHT-induced cADRs (additive model, $P = 3.7 \times 10^{-11}$; dominant model, $P = 4.8 \times 10^{-10}$). These SNPs were located in the *HLA-B* loci. Thus, we genotyped the individual *HLA-B* alleles and found that HLA-B*13:01 was present in 31.8% (7/22) of the cADR cases, but in 2.0% (15/751) of the tolerant controls (OR = 22.9, 95% CI = 8.2-64.3, $P = 7.0 \times 10^{-7}$). By the conditional analysis for the HLA-B amino-acid sequences, the most significant association was observed at a haplotype Leu145-Ala158 (corresponding to HLA-B*13:01 and B*13:02; OR = 19.9, 95% CI = 7.5-52.5, $P = 3.2 \times 10^{-7}$), followed by another haplotype Arg145-Thr158 (corresponding to HLA-B*38:01, B*38:02, B*39:01, B*39:02, B*39:04 and B*67:01; OR = 6.2, 95% CI = 2.6-15.1, $P = 2.1 \times 10^{-4}$). These two haplotypes showed the independent associations with PB and PHT-induced cADRs and the sum of them possessed the largest effect size (OR = 14.8, 95% CI = 5.9-37.2, $P = 6.0 \times 10^{-9}$). Comparison of genotypes of the *HLA-B* and the marker SNP rs34341880 in all subjects revealed that the sum of these two haplotypes (corresponding to the 8 *HLA-B* alleles) was in strong linkage disequilibrium (LD) with the T allele of rs34341880 ($r^2 = 0.95$, $D' = 1.00$). This work was supported by JSPS Core-to-Core Program, A. Advanced Research Networks.

383T

GRIP1 and *SRGAP1* **as top candidate loci in extended multiplex autism families.** *P.L. Whitehead[1], S. Luzi[1], L. Gomez[1], A.J. Griswold[1], H.N. Cukier[1], J. Jaworski[1], S. Slifer[1], J.R. Gilbert[1,2], M.L. Cuccaro[1,2], E.R. Martin[1,2], J.I. Haines[3], M.A. Pericak-Vance[1,2].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) John T. MacDonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH.

Autism spectrum disorder (ASD) is a highly heritable and prevalent neurodevelopmental condition with a demonstrated high heritability but with a complex genetic architecture. While genetic studies to date, including common variant genome wide association studies and rare variant identification with whole exome sequencing, have identified more than 100 candidate genes and genomic loci, they have failed to explain all of the underlying genetic risk for ASD. In order to more fully investigate the genetic etiology of ASD, we have ascertained a unique resource of 86 families of European ancestry with multiple avuncular pairs of individuals with ASD across well characterized extended pedigrees. We hypothesize that in families of this type linkage studies will be a powerful tool to identify genomic risk loci. All individuals were genotyped on either the Illumina HumanHap550 or Illumina Human1M genome-wide genotyping platforms. Extensive parametric 2-point and parametric multipoint linkage was performed using MERLIN software followed by extensive data cleaning (Merlin and Pedwipe), LD pruning (r2=0.16), and manual inspection of potential double recombinants using the Progeny software. The two point parametric analysis indicated linkage at 31 autosomal loci with HLOD >3.3. Of these, highest significance was found at 12q14 and 15q21 loci with HLOD>5.4 and HLOD>3.5, respectively. Multipoint parametric analysis supported these findings. A 3Mb region of 12q14 showed an HLOD>2.5 and covers 26 genes[Office1] . A 1Mb region on 15q21 showed an HLOD>2.2 and covers eight genes. In order to refine these results, we used identity by descent sharing to narrow the linkage peaks to ~400kb on each chromosome. Among the genes contained in the linkage peaks are ASD candidate genes *GRIP1* and *SRGAP1* on chromosome 12 and a neuronal specific glycosidase enzyme *LCTL* on chromosome 15. Examination of existing whole exome sequencing data of both ASD and non-ASD individuals in these families did not reveal segregating coding variants that could explain the linkage signals. Therefore, follow-up custom targeted sequencing covering noncoding regions in the genome to identify segregating nonconding, regulatory variants is underway and will be presented[Office2] . Overall, this approach demonstrates the power of linkage in unique, extended ASD pedigrees and suggests a role for regulatory variants on chromosome 12 and 15 in the genetic etiology of ASD. .

384F

Predicting age of onset in multiple sclerosis patients: *RGS14*, *CD40*, **and** *HLA-DRB1*15:01* **carriers are four years younger.** *M.F. George[1], Y. Natanzon[2], F.B.S. Briggs[2].* 1) Public Health, Dominican University of California, San Rafael, CA; 2) Institute of Computational Biology and the Department of Epidemiology and Biostatistics, School of Medicine, CWRU, Cleveland OH.

Epidemiologic evidence suggests a genetic component to multiple sclerosis (MS) clinical presentation. Previous research shows *HLA-DRB1*15:01* (DR2), the primary MS risk allele, is associated with earlier onset age. However, other genetic determinants of MS onset may also determine age of onset; therefore we hypothesize that increasing copies of the *established* risk alleles for genic MS variants predict earlier age of onset. There were 1,113 non-Hispanic white participants in the Accelerated Cure Project for Multiple Sclerosis with a comprehensive medical history and genotypic data. After removing genetic outliers and related individuals, 1,088 were available for analyses; all of whom had adult onset MS with mean onset age 33.4 yrs (SD=9.8). The sample was 78% females, 11% with a family history of MS, 94% relapsing presentation, and 33% ever smokers before onset. There was ≥80% power to detect 1.75 year change for MAF≥10% and 1.25 years for MAF≥25%. Therefore 49 of the 63 known non-MHC SNPs (MAF≥10%) were tested. Given the hypothesis, we imposed a Bonferroni corrected 1-sided α (p≤0.002). Disease course and smoking history were associated with onset age, thus all models were adjusted for these variables and ancestry. Consistent with prior findings, DR2 (rs3135388A) was significantly associated with onset age (p=0.0055); cases were 1.8 years younger per risk allele. Cases were 1.3 years younger per *CD40* (p=0.0009) and *RGS14* (p=0.002) risk allele. A score summing DR2, *CD40* and *RGS14* significantly predicted earlier onset (p=9.6x10[-7]). Individuals were on average 1.3 years younger per risk allele. Those with at least one risk allele (89%) were four years younger than cases with zero copies (p=2x10[-5]; 33 vs 37 years). Genetic risk score of all 50 or the remaining 47 variants were not associated with onset age. Genotyping of 250 white and 100 African American ACP participants are underway for replication analyses. These results are intriguing, suggesting specific MS risk variants may be associated with earlier onset age. *CD40* is a costimulatory protein found on antigen presenting cells and is required for their activation. Conversely *RGS14* is enriched in CA2 pyramidal neurons and suppresses synaptic plasticity. The findings suggest both inflammatory and neurodegenerative processes may contribute to earlier onset of MS. Replication is warranted, however if these findings are confirmed, the SNPs may contribute to the development of MS diagnostic tools. .

385W

Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid lipofuscinosis. *J. Xing[1], E. Gedvilaite[1], Y. Zhang[1], P. Lobel[2], D. Sleat[2].* 1) Department of Genetics, Rutgers Univ, Piscataway, NJ; 2) Center for Advanced Biotechnology and Medicine and Department of Biochemistry and Molecular Biology, Rutgers Univ, Piscataway, NJ.

The neuronal ceroid lipofuscinoses (NCLs) are a group of fatal neurodegenerative lysosomal storage diseases. While clinically similar, they are genetically distinct and result from mutations in at least twelve different genes. Estimates of NCL incidence range from 0.6 to 14 per 100,000 live births but vary widely between populations and are influenced by whether patients are classified based upon clinical or genetic criteria. We investigated mutations in twelve NCL genes in ~61,000 individuals represented in the Exome Aggregation Consortium (ExAC) whole exome sequencing database. Variants were extracted from ExAC and pathogenic alleles differentiated from neutral polymorphisms using annotated variant databases and missense mutation prediction tools. Carrier frequency was dependent on ethnicity, with the highest (1/75) observed for *PPT1* in the Finnish. When data are adjusted for ethnic diversity within the USA, *PPT1*, *TPP1* and *CLN3* carrier frequencies were found to be the highest of the NCLs, each at ~ 1/400. Carrier frequencies calculated from ExAC correlated well with incidence estimated from numbers of living NCL patients in the US. In addition, the analysis identified numerous variants that are annotated as pathogenic in public repositories but have a predicted frequency that is not consistent with patient studies. These variants appear to be neutral polymorphisms that are reported as pathogenic without validation. Based upon literature reports, such alleles may be annotated in public databases as pathogenic and this propagates errors that can have clinical consequences.

386T

Epigenome-wide association study of delayed cerebral ischemia in stroke patients. *A. Arockiaraj, T.A. Koleck, J.R. Shaffer, P.R. Sherwood, E.A. Crago, S.M. Poloyac, D.E. Weeks, Y.P. Conley.* University of Pittsburgh , Pittsburgh, PA.

Aneurysmal subarachnoid hemorrhage (aSAH) is a form of stroke leading to severe patient outcomes including morality, cognitive impairment, functional disability, and emotional dysfunction. Variability in recovery and development of chronic complications is thought to be negatively impacted by delayed cerebral ischemia (DCI; insufficient blood flow to the brain), which may occur in patients 4 to 10 days following aSAH. Though little is known about the pathophysiology of DCI and its role in the development of chronic complications, dynamic methylomic changes occur during the period following brain injury, and DNA methylation is associated with ischemic tolerance in brain tissue. We hypothesize that epigenetic changes post-injury lead to differences in methylomic profiles between patients that go on to develop DCI and those that do not. Moreover, we hypothesize that these epigenetic changes occur in key genes and/or biological pathways related to recovery and chronic complications. To investigate these hypotheses, we performed parallel epigenome-wide association studies for DCI in 86 blood samples (from 43 DCI cases and 43 controls) and 66 cerebrospinal fluid samples (from 28 DCI cases and 38 controls) collected from patients 1 day after aSAH. Whole-genome methylation data were collected using the Illumina HumanMethylation450K BeadChip. Surrogate variable analysis was used to remove variation due to unknown confounding factors (e.g., distributions of cell types and batch effects). After correction for multiple testing, no significant results were obtained in the CpG site-specific tests for differential methylation in either of the two samples, suggesting that soon after aSAH, methylation differences may not be associated with future DCI events. Results of bump-hunting for differentially methylated regions spanning multiple CpG sites, which may be more powerful, will be presented. In conclusion, this study provides preliminary characterization of the methylomic profiles of cerebrospinal fluid in aSAH patients. Additional work is needed to understand how these methylomic profiles change across time post-injury, which specific regions are associated with DCI, and whether DNA methylation biomarkers may ultimately be useful for predicting aSAH recovery outcomes. R01NR013610.

387F

Whole exome sequencing of early-onset Alzheimer disease highlights candidate risk genes involved in endocytosis. *B.W. Kunkle[1], B.N. Vardarajan[2], A.C. Naj[3], P.L. Whitehead[1], S. Rolati[1], S. Slifer[1], E.R. Martin[1], J.R. Gilbert[1], C. Reitz[2], J.L Haines[4], G.W. Beecham[1], G.D. Schellenberg[3], R.P. Mayeux[2], M.A. Pericak-Vance[1], The Alzheimer's Disease Genetics Consortium.* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

Introduction. Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for only ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease (AD) unexplained. This extreme phenotype likely harbors highly penetrant risk variants however, and thus represents an excellent resource for discovery of novel risk genes and pathways for AD. **Methods.** We performed Whole-Exome Sequencing (WES) in 53 Non-Hispanic White (NHW) EOAD cases (Age-at-onset (AAO) < 65) previously screened negative for *APP*, *PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. Variant filtering for rare, damaging (CADD Phred Score ≥ 15) nonsynonymous or loss-of-function (LOF) variants (MAF<0.1%) was performed, using familial segregation data if available. Followup of a) variants in two or more NHW cases, b) genes with variants in the same gene for two or more NHW cases and c) candidate variants and genes present in 19 Hispanic EOAD WES families, was conducted in an Alzheimer's Disease Genetics Consortium (ADGC) cohort of 1,292 EOAD cases (AAO <= 65), 7,087 late-onset AD (LOAD) cases (AAO > 65) and 6,889 controls (Age-at-exam ≥ 65) genotyped with the Illumina exome chip. **Results.** We identified 108 rare, damaging nonsynonymous or LOF variants in two or more EOAD cases, of which 43 were available on the exome chip. Among these, we found a missense variant in the endosomal trafficking gene *RUFY1* associated with EOAD (P=0.0038). Gene-based testing (SKAT-O) revealed several genes associated with EOAD including *PSD2* (P=0.001), which was also associated with risk of LOAD in the ADGC cohort (P=6.2x10^{-6}). Additionally, a damaging missense variant in the gene *TCIRG1* was present in both NHW and Hispanic EOAD WES cohorts, was more frequent in cases than controls (MAF = 0.0032 and 0.0014, respectively) in the ADGC EOAD cohort (P=0.059), and was significant in the ADGC LOAD cohort (P=0.0065). *PSD2* and *TCIRG1* are also involved in endocytosis, a process known to be important to development of AD through the gene *SORL1*. Multiple brain expression datasets show all three genes to be differentially expressed in AD. **Conclusions.** WES of EOAD identified novel candidate endocytic genes for EOAD, two of which are shared risk genes between EOAD and LOAD, a phenomena previously discovered for genes such as *SORL1*, *PSEN2* and *TREM2*.

388W

High-resolution imputation of rare variants identifies novel rare variant candidate loci in late-onset Alzheimer's disease: The Alzheimer's Disease Genetics Consortium. *A.C. Naj[1], Y. Zhao[2], L. Qu[2], S.J. van der Lee[3], K.L. Hamilton-Nelson[4], B.W. Kunkle[4], A. Kuzma[2], O. Valladares[2], C. Reitz[5], G.W. Beecham[3], E.R. Martin[3], L.-S. Wang[2], J.L. Haines[6], R. Mayeux[5], L.A. Farrer[7], M.A. Pericak-Vance[4], G.D. Schellenberg[2], Alzheimer's Disease Genetics Consortium.* 1) Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Department of Epidemiology, Erasmus University Medical Centre, Rotterdam, Netherlands; 4) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 5) Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 6) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 7) Departments of Medicine, Neurology, Ophthalmology, Genetics & Genomics, Epidemiology, and Biostatistics, School of Medicine, Boston University, Boston, MA, USA.

The largest genome-wide association study (GWAS) to date for late-onset Alzheimer Disease (LOAD) identified 19 susceptibility loci in addition to *APOE*, the majority of which were common variant associations (minor allele frequency (MAF)>0.05). The Haplotype Reference Consortium (HRC) recently released a reference panel of 64,976 haplotypes with 39,235,157 SNPs allowing imputation down to an unprecedented MAF=0.00008 to test association of very low frequency variants. To identify novel rare variant associations, ADGC imputed 33 case-control, family, and prospective datasets including 14,743 cases and 15,871 non-cases to the HRC reference panel (8/15 release) using Minimac3 on the University of Michigan Imputation Server, with an average 39,216,773 genotyped or imputed SNVs per dataset. Logistic regression on common variants (MAF>0.01) was performed using imputed genotype probabilities in PLINKv1.9 (generalized linear mixed model in R were used for family-based datasets) and meta-analyzed in METAL, while rare variants (MAF≤0.01) were analyzed using score-based tests and meta-analysis in the SeqMeta/R package; both analyses adjusted for age, sex, and population substructure. Preliminary analyses identified five loci with P<5×10^{-8} in known GWAS candidate loci (*APOE*, *BIN1*, the *MS4A* region, *PICALM*, and *CR1*), consistent with previous associations in these data. An additional 13 loci demonstrated multiple single variant associations with P<10^{-5}, including variants at two prior GWAS loci, rs13155750 in *MEF2C* (OR(95% CI): 1.13(1.08,1.20); P=5.09×10^{-6}) and rs755951 in *PTK2B* (OR(95% CI): 1.11(1.06,1.16); P=5.61×10^{-6}). Novel associations observed include signals at *LILRA5* (19:54821819; OR(95% CI): 1.14(1.08,1.20); P=4.84×10^{-7}), known to be involved in innate immunity pathways; and at *SMOX* (rs1884732; OR(95% CI): 1.11(1.06,1.17); P=5.17×10^{-6}), which is involved in the catabolism of polyamines, levels of which are altered in AD brains. Score test analyses of rare variants identified a novel association at *CUTC* (chr10:101510468, MAF=0.002; OR(95% CI): 0.27(0.17,0.43); P=4.57×10^{-8}), a chromosome 10 copper homeostasis gene proximal to the AD-implicated gene *DNMBP*. Replication analyses and functional follow-up are underway; results will be presented. Several novel candidate loci for LOAD were identified using high-quality imputation of rare variants in the ADGC, demonstrating the utility of rare variant imputation using dense haplotype reference panels for gene discovery.

389T

Dissecting the shared genetic architecture of human communication: Insights into speech, language and reading. *H. Voss-Hoynes[1], C.M. Stein[1,5], B. Truitt[1], J. Tag[2], L. Freebairn[2], J. Vick[2], H.G. Taylor[3], B.A. Lewis[2], S.K. Iyengar[1,4].* 1) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Communication Sciences, Case Western Reserve University, Cleveland, Ohio; 3) Pediatrics, Case Western Reserve University, Cleveland, Ohio; 4) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio; 5) Center for Proteomics & Bioinformatics, Case Western Reserve University, Cleveland, Ohio.

Purpose. Speech sound disorders (SSD) are communication disorders that occur in 16% of three-year-olds. These disorders affect sound patterns (phonology) and sound production (articulation), and can be comorbid with impairments in language (LI) and reading (RD) likely due to shared genetic architecture. We aimed to explore the shared architecture through genome-wide association study of reading ability and to determine if the genetic-quantitative trait relationships are affected by RD and LI. **Methods**. Using a mixed linear model, we performed a family-based, genome-wide association study in 318 Caucasian individuals over 5 years old with Omni 2.5 Exome chip data imputed to Phase 3 of the 1000 Genome Project. The outcome was scores on the Woodcock Word Identification (WRID) test, a known word reading test. **Results**. We report evidence of the shared genetic architecture between LI, RD, and SSD. The most significant marker, rs193704, B_{unadj}=2.47 ± 0.44, p=3.63x10^{-7}, is a protective locus within *ATP2C2*, a gene previously associated with phonological processing in individuals with LI. Phonological processing, the ability to fraction words into component sounds, is the most common link between SSD, LI, and RD and is positively correlated with reading ability (R^2=0.40). The strength of the relationship between markers in *ATP2C2* and reading scores is slightly reduced when accounting for LI (B_{unadj}=2.47 ± 0.44 vs. $B_{adjlang}$=2.17 ± 0.45, Δ= -12%) and RD affection status ($B_{adjread}$=2.22 ± 0.44, Δ= -10%), but the marker remains genome wide suggestive (p_{lang}= 1.91x10^{-6}, p_{read}=4.46x10^{-6}). Our next most significant signal was a novel locus in *IQCE* (rs4721874,-2.09±0.44, p=3.12x10^{-6}). IQCE is expressed in the hippocampus and is involved in cilia mediated Shh signaling, a process which mediates adult neurogenesis in the same region. Because the hippocampus plays a role in learning, memory, and visual processing, aberrant neurogenesis could result in reading deficits. Accounting for LI and RD attenuates the relationship between rs4721874 and WRID (B_{lang}= -1.6 ±0.42, p_{lang}=1x10^{-4}; B_{read}=-1.4 ±0.4, p_{read}=3x10^{-4}). While IQCE may influence reading, it does not do so independently of LI and RD status. **Conclusions**. Our findings suggest that speech, language, and reading are intricately connected by pleiotropic gene networks. Perturbations in these networks and genes provide insights into how modern human communication has evolved. DC00528, DC012380, T32-HL007567.

390F

Deep genomic analysis of complex diseases and its potential application to precision medicine. *M. Xiong[1], P. Wang[4], N. Lin[1], Y. Zhu[2], J. Zhao[2], L. Jin[4], D. Bennett[3].* 1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX; 2) The Tulane University , New Orleans, LA; 3) Rush University Medical Center, Chicago, IL; 4) Fudan University, Shanghai, China.

Next-generation sequencing, imaging and sensing technologies will generate deeper genomic and phenotypic data with millions of features. There are multiple steps between genes and phenotypes. Each step may be influenced by genomic variation and can weaken links between genes and phenotypes. As a consequence, this will obscure the causal mechanism of the phenotype. Integrated and deep analysis of such increasingly larger and deeper biomedical datasets will provide invaluable information for holistic discovery of complicated genetic structures underlying human complex diseases. Precision medicine demands deep, systematic, comprehensive and precise analysis of genotype-phenotype – "and the deeper you go, the more you know". But, standard genetic analysis is to test the association of single variants with a single trait. However, these traditional approaches are less efficient in analyzing high-dimensional, heterogeneous types of correlated genomic and multiple phenotype data, and thus unable to reveal the deep causal structures underlying human complex disorders. Therefore, there is urgent need to develop novel statistical methods and computational algorithms for big and deep genomic analysis of complex diseases and precision medicine. To break the current impasse in progress to precision medicine, we use causal inference theory and chain graph models to develop an innovative analytic platform for deep and precise multilevel hybrid causal genotype-disease network analysis to increase breadth (the number of phenotypes which the genetic variants affect) and (2) depth (the number of steps which are taken by the genetic variants to reach the clinical outcomes). The proposed methods have been applied to Alzheimer Disease (AD) dataset with 53295 genes typed in 1,708 individuals. We identified 387 genes associated with AD, and 20 AD related traits including 7 cognitive function traits, 7 lipoprotein metabolism traits) and 5 environments. We also identified a genotype-disease network consisting genotype subnetwork, phenotype subnetwork and environment subnetwork. The network consists of 37 nodes and 139 edges. All phenotypes in the network were clustered into two groups: cognitive trait group and lipid metabolism trait group. Two groups were connected via MMSE and SBP connection. Five genes were directly connected with AD. Eight genes have > 9 connections. Most genes included in this network are involved in neurological and neuropsychiatric disorders.

391W

Heteroscedastic pleiotropy mapping on the biomedical image data of multiethnic schizophrenia patients. *H. Qin[1], W. Ouyang[1], Y.P. Wang[2].* 1) Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Biomedical Engineering, New Orleans, LA 70118, USA.

Heteroscedasticity (aka, dispersion heterogeneity of trait residuals) is a common phenomenon in biomedical image data mining. It is especially significant when massive clinical traits are involved. As a cause for cross-trait association, pleiotropy occurs when a genetic locus truly affects two or more distinct clinical disorders. Detection and interpretability of complex SNP(s)-trait(s) associations are complicated due to LD patterns between SNPs and correlations between traits. Background confounders, especially in multiethnic imaging studies, may have impact both mean and higher-order heterogeneities of clinical traits. However, existent multivariate methods neglect all the dispersion effects of both test biomarkers and background confounders. After calibrating confounding effects, leveraging dispersion effects of multiple genetically correlated traits is expected to localize not only trait-specific causal loci and genes but also those pleiotropic ones shared between the traits. Herein, we describe a double generalized heteroscedastic linear-mixed model based procedure for heteroscedastic pleiotropy mapping (HPM). In this multi-trait gene-wise association method, we scale the importance of individual traits by the *p*-values of marker-wise heteroscedastic effects. For each trait, causative SNPs and the flanking random markers are up-weighted, and others, down-weighted. By extensive simulations under published designs, the proposed method properly controlled type I error rates and appeared strikingly more powerful than existent prominent multivariate sequence association methods. The HPM is computationally efficient, integrates informative heterogeneities and diverse genetic effects while simultaneously adjusting for arbitrary background covariates and cryptic relatedness. It displayed particular utility in analyzing the biomedical image data of multiethnic schizophrenia patients. In summary, we have paved an effective avenue to integrate high-order informative heterogeneities while adjusting for heteroscedastic background confounders. Our results have demonstrated the tremendous potential gain of explicitly handling higher-order heterogeneities in big multiethnic image data. This work was supported by Innovative Programs Hub (I2PH) Grants Award of Tulane (632037) and Tulane's Committee on Research fellowship (600890).
Contact: Huaizhen Qin, Ph.D. Email: hqin2@tulane.edu.

392T

Novel genetic loci associated with relapse rate in multiple sclerosis. *Y. Zhou[1], S. Simpson,Jr.[1], J. Charlesworth[1], I. van der Mei[1], R. Lucas[2], A.L. Ponsonby[3], B.V. Taylor[1].* 1) Menzies Institute for Medical Research, Hobart, TAS, Australia; 2) National Centre for Epidemiology and Population Health, Canberra, Australia; 3) Murdoch Children's Research Institute, University of Melbourne, Melbourne, Australia.

Background: Due to the gradual accumulation of multiple sclerosis (MS) progression and the high cost of follow-up, no study has ever tried to identify genetic factors that predict relapse rate at a genome wide scale. **Material and Methods:** In the discovery stage, the genome wide association analysis for predicting relapse rate was done using the Southern Tasmania MS Longitudinal Study with 141 participants followed for 2.5 years. In the validation stage, SNPs with P<0.05 from discovery stage were further validated using another independent longitudinal study (the Ausimmune Longitudinal Study) with 279 participants followed for 5 years. Predictors of time to relapse were evaluated by Cox proportional hazards regression models. **Results:** We discovered one novel loci (*LRP2*) with multiple variations with genome wide significance that predict relapse (HR=2.25, P=4.02×10^{-8}). LRP2 is critical for the reuptake of ligands, including lipoproteins, sterols, vitamin-binding proteins and hormones. LRP2 is expressed on the surface of many CNS cells including neurons and oligodendrocytes and is a critical receptor in axonal guidance. Another loci (*N4BP2*) showed suggestive association with relapse (HR=2.31, P=8.82×10^{-8}). **Conclusion:** Our results provided novel insights into the genetic drivers of variation in risk of relapse rate in MS. The finding of a genome wide significant loci that has extensive effects on neuronal development and repair is of significant interest.

393F

Two‑component mixture modelling approach integrating genetic and clinical variables in analysis of time to remission in epilepsy. *B. Francis[1], A. Jorgensen[1], A. Morris[1], A. Ingasson[2], A. Marson[1], M. Johnson[3], G. Sills[1], EpiPGX.* 1) University of Liverpool, Liverpool, Merseyside, United Kingdom; 2) deCODE, Iceland; 3) Imperial College London, United Kingdom.

Time to twelve months seizure freedom (remission) after randomisation to an antiepileptic drug (AED) is an important outcome in the treatment of epilepsy with an established genetic component to variable response. A collaborative dataset of 1,706 patients of European ancestry has been assembled via the EpiPGX Consortium to investigate genetic risk factors for remission and other epilepsy-related outcomes. Genotype data were subject to quality control and individuals with <90% match with European reference data were removed from downstream association analyses. The cleaned data was then imputed up to the 1000 Genome reference panel (March 2012 release , all ancestries) using IMPUTE2. Two sub-populations exist for time to remission; those who enter remission and those not susceptible to remission. A traditional survival analysis approach was therefore inappropriate, as the assumption of a homogenous population is violated, and may lack power to detect SNP associations. To consider these two sub-populations, mixture modelling with cure fraction was implemented in a two component model. The two component model is computationally expensive and cannot be applied on the scale of the whole genome. Therefore, to test association, deviance residuals for susceptibility to remission and survival were derived, after adjustment for relevant clinical covariates. We performed multivariate association analysis of SNPs with "reverse regression" modelling in the PLEIOTROPY software. No SNP associations attained the threshold of genome-wide significance ($p < 5 \times 10^{-8}$). However, lead SNPs in several loci attained nominal significance ($p < 1 \times 10^{-5}$), including: IQCK (rs724614, $p = 9.36 \times 10^{-7}$) and PDE3A (rs7484691, $p = 7.62 \times 10^{-7}$). Further investigation of these loci will be undertaken by performing computationally expensive two component survival modelling, including SNP genotypes in the model alongside clinical covariates.

394W

Case-only investigation confirms multiple sclerosis risk conferred by tobacco smoke is modified by *NAT1.* *F.B.S. Briggs[1,2].* 1) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH; 2) Institute of Computational Biology, Case Western Reserve University, Cleveland, OH.

Background: Multiple sclerosis (MS) is a multi-factorial autoimmune and neurodegenerative disease of the central nervous system, with genetic and exogenous risk components. Of the exogenous risk factors, tobacco smoke confers the largest risk (OR~1.5). Unfortunately the underlying processes mediating MS risk in smokers remain unknown. A prior case-control study identified a gene-environment interaction between *NAT1* and tobacco smoke (Briggs *et al* 2014), suggesting tobacco smoke metabolism is relevant. MS risk among smokers was significantly increased among carriers of a rs7388368A-rs4921877T-rs6586711C haplotype. **Methods:** 1,088 non-Hispanic white MS participants in the Accelerated Cure Project for MS (ACP) with epidemiologic and genetic data were available for analyses. All participants had an age of onset ≥18 years. Participants reported detailed smoking history; they were classified as a smoker if they were an active smoker within 5 years prior onset. Participants were genotyped using a Human Illumina Custom BeadChip, including >200,000 exonic variants. 80 *NAT1* variants with MAF≥0.01 and genotype rate≥99% were available. Given the prior finding, this investigation focused specifically on variation within the respective haplotype block, defined similarly using solid spline (extend where D'>0.8). There were 4 haplotype blocks across the gene; the second block consisted of 6 SNPs including rs4921877 and rs6586711 (rs7388368 was not available) and spanned 13kb. Case-only analyses were conducted using smoking status before onset as the outcome for the individual SNP and haplotype analyses. All models were adjusted for gender, birth year, disease course, age of onset, and ancestry. A 1-sided Bonferroni-corrected (6 SNPs) alpha of 0.0083 determined significance. **Results:** *NAT1* rs56261729A, which alters a SRF motif, was significantly associated with smoking status (OR=3.3, p=0.0067; MAF=0.03). Smokers were more likey to have rs4921877T and rs6586711C (OR~1.3), but associations were not significant (p>0.1). There were 6 haplotypes, only one carried rs56261729A, and it was significantly associated with smoking status in MS cases (OR=2, p=0.0055). Both the haplotype and the SNP had the same frequency: 4.2% in smokers and 2.3% in non-smokers. **Conclusions:** This case-only analysis confirms a previously identified *NAT1* haplotype block locus as modifying MS risk conferred by tobacco smoke. Further analyses are needed to further characterize this *NAT1* haplotype block.

395T

Using LASSO regression to identify gene-gene and gene-environment interactions influencing cognitive function in a sample at increased risk for Alzheimer's disease. *B. Darst[1], M. Bilgel[2,3], R. Koscik[4], B. Hermann[4,5,6], B. Jedynak[7], S. Johnson[4,5,8], C. Engelman[1,4,8].* 1) Department of Population Health Sciences, University of Wisconsin, Madison, Madison, WI; 2) Department of Biomedical Engineering, Johns Hopkins University School of Engineering, Baltimore, MD, USA; 3) Laboratory of Behavioral Neuroscience, National Institute on Aging, NIH, Baltimore, MD, USA; 4) Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 5) Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI, USA; 6) Department of Neurology, University of Wisconsin School of Medicine and Public Health, WI, USA; 7) Department of Applied Mathematics and Statistics, Portland State University, Portland, OR, USA; 8) Alzheimer's Diseases Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

Alzheimer's disease (AD) is the 6th leading cause of death in the US and the only leading cause that cannot be prevented, cured, or treated. Although the heritability of AD is estimated to be 58%, most known genetic variants only account for 8% of the phenotypic variance of AD. Expanding investigations to include complex nonlinear relationships between genetic and environmental factors may explain some of this missing heritability. In this analysis, we explore gene-gene and gene-environment interactions that may be influencing cognitive function using participants from the Wisconsin Registry for Alzheimer's Prevention, a longitudinal family-based cohort of initially cognitively healthy adults (N=1,293) enriched for a parental history of AD. This analysis included 19 genetic variants indicated to be associated with AD by the International Genomics of Alzheimer's Project, plus *APOE* ε2 and ε4, and eight environmental factors that have been previously associated with AD: age, gender, education, physical activity, waist-hip ratio, total cholesterol, insulin resistance, and inflammation as measured by interleukin 6 (IL6). We performed variable selection using least absolute shrinkage and selection operator (LASSO) regression, inputting all possible pairwise interaction terms into the model and using the value of the parameter λ that gave the minimum mean cross-validation error. Potential interactions identified by LASSO were tested for significance in mixed linear models. Of the 14 gene-gene or gene-environment terms identified by LASSO, two were statistically significant. An interaction between rs6656401 (*CR1*) and rs9271192 (*HLA-DRB1*) (*P*=.004) suggested that being a carrier of either of these variants was associated with poorer cognitive function, whereas being a carrier of both of these variants was associated with better cognitive function. Evidence of this relationship was also provided by three of six other independent cognitive measures (all with *P*<.004). This suggests that while these variants are independently detrimental, collectively they may have a protective effect. The second interaction was between rs190982 (*MEF2C*) and IL6 (*P*=.006); however, we did not see evidence for this relationship in other cognitive measures. All three of these genes have been reported to be in the immune system pathway and are associated with several other autoimmune diseases. These findings support the early involvement of the immune system in AD.

396F

Mendelian randomization identifies ACE, APOC-1, APOE, clusterin, and GDF-15 as causal serum biomarkers of Alzheimer disease risk. *D. Carere[1,2,3], J. Sjaarda[1,2,4], S. Hess[5], H. Gerstein[1,6,7], G. Pare[1,2,3].* 1) Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, 237 Barton Street East, Hamilton, Ontario, Canada; 2) Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada; 3) Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Ontario, Canada; 4) Department of Medical Sciences, McMaster University, Hamilton, Ontario, Canada; 5) Sanofi Aventis Deutschland GmbH R&D Division Diabetes, Frankfurt, Germany; 6) Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; 7) Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

Background: Dozens of blood-based biomarkers have been linked to Alzheimer disease (AD), but findings are inconsistent, and likely biased by unmeasured confounding and reverse causation. We combined biomarker data from the Outcome Reduction within Initial Glargine Intervention (ORIGIN) trial with AD data from the International Genomics of Alzheimer's Project (IGAP), and used Mendelian randomization (MR) to identify blood serum biomarkers with a causal role in AD. **Methods:** Genotyping data and baseline serum measurements of 237 biomarkers were available from 5,078 ORIGIN trial participants. Summary statistics from a genome-wide association study of AD (17,008 cases and 37,154 controls) were obtained from IGAP. MR analyses were restricted to biomarkers directly encoded by an autosomal gene and SNPs within 300Kb of a biomarker's corresponding gene. MR associations were computed by regressing SNP-AD effect estimates on SNP-biomarker effect estimates. Conditional MR was used to evaluate independence of effect of proximally encoded biomarkers. **Results:** After quality control, 1,731 non-redundant SNPs associated with 203 biomarkers were available for MR. Five biomarkers were associated with AD, after adjusting for multiple hypothesis testing (α = 0.05/203). MR predicted a deleterious effect of apolipoprotein C-1 (APOC-1, 11 SNPs; OR=1.89 (1.65-2.16)) and growth differentiation factor 15 (GDF-15, 9 SNPs; OR=1.24 (1.14-1.35)); and a protective effect of angiotensin-converting enzyme (ACE, 15 SNPs; OR=0.91 (0.87-0.95)), apolipoprotein E (APOE, 20 SNPs; OR=0.33 (0.32-0.35)), and clusterin (4 SNPs; OR=0.67 (0.56-0.80)). In an adjusted analysis of APOE and APOC-1 (27 SNPs total), each biomarker remained significantly associated with AD. When we further removed from the joint analysis the *APOE* SNPs which determine epsilon allele status (rs7412, rs429358), and all SNPs in LD with these, the effects persisted: APOC-1 OR = 4.09 (3.48-4.81); APOE OR = 0.20 (0.18-0.23). **Conclusion:** Mendelian randomization suggests a causal role for serum ACE, APOC-1, APOE, Clusterin, and GDF-15 concentration in AD risk. The risk effects of increased serum APOC-1 and decreased serum APOE concentration appear to be independent of each other, and of *APOE* epsilon allele status. Our study takes advantage of the largest AD genetics dataset available, and in combination with data on hundreds of serum biomarkers, represents the first investigation of multiple blood biomarkers for a causal role in AD.

397W

Adaptive testing for multiple traits with applications to detect SNP-brain network associations. *J. Kim, W. Pan.* Division of Biostatistics, University of Minnesota, Minneapolis, MN.

There has been increasing interest in developing more powerful and flexible statistical tests to detect genetic associations with multiple traits, as arising from neuroimaging genetic studies. Most of existing methods treat a single trait or multiple traits as response while treating an SNP as a predictor coded under an additive inheritance mode. In this paper we follow an earlier approach in treating an SNP as an ordinal response while treating traits as predictors in a proportional odds model (POM). In this way, it is not only easier to handle mixed types of multiple traits, e.g. some quantitative and some binary, but it is also more robust to the common assumption of an additive inheritance mode. More importantly, we develop an adaptive test in a POM so that it can maintain high power across many possible situations. Contrary to the existing methods treating multiple traits as responses, e.g. in a generalized estimating equation (GEE) framework, the proposed method can be applied to a high dimensional setting where the number of phenotypes (p) can be larger than the sample size (n), in addition to a usual small p setting. The promising performance of the proposed method was demonstrated with applications to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data, in which either structural MRI driven phenotypes or resting-state functional MRI (rs-fMRI) derived brain functional connectivity were used as multiple phenotypes. The applications led to the identification of several top SNPs of biological interest. Furthermore, a simulation study showed competitive performance of the new method compared to several existing methods, including potential power gain of the new method in cases with a dominant inheritance mode.

398T

The rare variant generalized disequilibrium test for association analysis of nuclear and extended pedigrees. *S.M. Leal, Z. He, D. Zhang, B. Li, G.T. Wang.* Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX.

Whole genome and exome sequence data can now be cost effectively generated for families in order to perform rare variant (RV) association analysis. An advantage of analyzing families is that causal variants that aggregate in families usually have larger effect sizes than those found in the general population, thus increasing the power of family-based association studies. Population-based RV association studies can also suffer from the inflated false-positive rates due to population admixture/substructure which family-based analysis can avoid. We developed RV extensions of the Generalized Disequilibrium Test (GDT; Chen et al 2009) to analyze nuclear and extended families. The GDT utilizes the genotype differences of all discordant relative pairs to assess association within a family, and the RV extension of GDT (RV-GDT) is combination of the single variant GDT statistic over a genomic region of interest. The GDT has increased power by efficiently incorporating information beyond first-degree relatives and can be applied when there is missing genotype, e.g. parental data is missing. Using simulated genetic data, we demonstrate that the RV-GDT method has well-controlled type I error rates, even when applied to admixed populations. We also extended the Pedigree Disequilibrium Test (PDT; Martin et al. 2000) to analyzed RVs but the RV-GDT has higher power. The RV-GDT is also more powerful than the RV Affected Sibpair (Epstein et al. 2015) and Family Based Association Test (FBAT, De et al. 2013). The additional advantage of the RV-GDT over these methods is that it can be utilized to analyze extended families and also allows for missing genotype data. Using the RV-GDT publically available user-friendly software we analyzed the exome sequence data from 2,377 nuclear families with autism spectrum disorder (ASD) from Simons Simplex Collection and whole genome sequence data from 81 nuclear and extended Alzheimer's disease (AD) pedigrees. We identified several genes associated with ASD: *ACSBG2*, *C17orf50* and *SGCA* and AD: *TNK1*. Additional results from these analyses as well as extensive simulation studies will be used to demonstrate the power and capabilities of RV-GDT to identify genes involved in the etiology of complex familial diseases.

399F

Coding and non-coding variants in known Alzheimer's candidate genes co-segregate with Late-Onset Alzheimer's disease (LOAD) in the Alzheimer Disease Sequencing Project genome sequence data. *J. Jaworski[1], G. Beecham[1], B. Vardarajan[2,3,4], E. Blue[5], S. Barral[2,3,4], J. Haines[6], W. Bush[6], C.M. Van Duijn[7], E. Martin[1], G. Schellenberg[8], R. Mayeux[2,3,4], E. Wijsman[5], M. Pericak-Vance[1].* 1) John P. Hussman Institute of Human Genomics, Miller School of Medicine, Universi, Miami, FL; 2) The Taub Institute of Research on Alzheimer□s Disease, Columbia University, New York, NY, USA; 3) The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; 4) The Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 5) Department of Medicine, University of Washington, Seattle, WA, USA; 6) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 7) Erasmus MC, Epidemiologie, Rotterdam, The Netherlands; 8)) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

   Background: The Alzheimer's disease Sequencing Project (ADSP) is an initiative to identify genetic variation influencing LOAD risk. To accomplish this we performed whole-genome sequencing (WGS) on 44 non-Hispanic white (NHW) extended families affected by LOAD and investigate whether variation in known candidate genes co-segregated with LOAD in these families. **Methods**: WGS data were generated for 229 subjects from 44 NHW families at the NHGRI sequencing centers, including both affected individuals and unaffected, elderly relatives. Sequence alignment was performed using BWA, followed by consensus genotype calling using multiple pipelines. Our candidate genes include known early-onset AD (EOAD) genes, genes known to cause dementia, and LOAD genes implicated by GWAS or resequencing studies. Variants within candidate gene regions were annotated for function, frequency, segregation with disease, and information from the ADSP case-control and other datasets. **Results:** We identified 41 variants across the families that were within a candidate gene and co-segregated in >75% of affected genotyped with disease, had minor allele frequency <5%, and were putatively functional (CADD score > 15). These variants were in 12 of the known EOAD genes or LOAD genes implicated by GWAS, most notably *APP*, *PICALM*, *PSEN1*, and *GRN*. *MS4A6A* had a rare, putatively functional (CADD=23.9) splice donor variant. Two other non-intronic, non-coding variants were 3' UTR variants seen in *PICALM* (CADD=20.4) and *MEF2C* (CADD=16.5). Within genes known to cause dementias that can mimic AD, we find co-segregating missense variants in *CSF1R*, *PDGFRB*, and *NOTCH3*, and non-coding variants in *CSF1R* and *SLC20A2*. Within genes known to cause dementias distinct from AD, we find co-segregating missense variants in *ATP13A2* and *LMNB1*, and non-coding variants in *GALC*, *TARDBP*, and *TBP*. All of these non-coding variants with RSIDs are annotated as having enhancer or promoter histone marks. **Conclusion**: This study suggests that rare variation in previously identified candidate genes may play a role in familial LOAD risk. This role extends to both early-onset genes and genes previously implicated with common variation, and may indicate additional mechanisms for LOAD risk.

400W

Whole genome sequence analysis of brain MRI measures in the Framingham Study. *C. Sarnowski[1], C.L. Satizabal[2,3], C. DeCarli[4], A.N. Pitsillides[2], A. Beiser[1,2,3], A.L. DeStefano[1,2,3], J. Dupuis[1,2], S. Seshadri[2,3].* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 3) Department of Neurology, Boston University School of Medicine, Boston, MA; 4) Department of Neurology, University of California at Davis, CA.

   Background: Brain MRI measures such as Total Brain Volume (TBV), Hippocampal Volume (HPV) and White Matter Hyperintensities (WMH) are endophenotypes of Alzheimer disease (AD) and vascular injury. Identifying loci that influence these measures may lead to the discovery of new genes and biological mechanisms underlying these diseases. Genome-Wide Association Studies (GWAS) have identified common genetic variants (Minor Allele Frequency ≥ 5%) with modest effect sizes, and most identified variants reside in non-coding regions. **Aim:** We sought to identify rare variants influencing TBV, HPV and WMH in the Framingham Study by performing Whole Genome Sequence (WGS) analyses within the Trans-Omics for Precision Medicine (TOPMed) Program. **Methods:** A total of 905, 894 and 672 individuals from TOPMed were included in the WGS analyses of TBV, HPV and WMH respectively. Mixed effect linear regression models of brain MRI measures were adjusted for sex, age and total intracranial volume and rank-normalized residuals from these models were tested for association with individual SNPs while taking into account familial relatedness. Gene-based tests were conducted for rare variants with two different methods (SKAT or burden test) using 1) a sliding-window approach or 2) a selection of functional exonic SNPs. Finally, WGS results were compared to GWAS using imputed data from the Haplotype Reference Consortium release 1 (~4,000 genotyped individuals with phenotype data). **Results:** WGS analyses revealed at the genome-wide level (P≤5x10[-8]) a new locus in 3q24 for HPV (P_{WGS}=2.0x10[-9], P_{GWAS}=0.003) and suggestive results (P≤10[-6]) in new regions for HPV (5q35, 8q24, 11p15, 19p13), WMH (11p12) and TBV (4q25). Known associations in 12q24 for HPV (P_{WGS}≤ 2.5x10[-3], P_{GWAS}≤ 4.1x10[-8]) and in 17q25 for WMH (P_{WGS}≤ 2.8x10[-4], P_{GWAS}≤ 4.2x10[-6]) were observed. Gene-based tests using a sliding-window approach detected genome-wide associations (P≤2.5x10[-6]) in new loci for HPV (3q24) and TBV (17p13) with the SKAT method and for HPV (4p15, 12q23, 12q24) and TBV (5q23, 9p21) with the burden test. The 3q24 locus associated with HPV includes *ZIC1,* a transcription factor expressed only in human brain with tumor suppressor function that has an important role in brain development and can transactivate *APOE*, whose ε4 allele is the major known risk-factor for late-onset AD. **Conclusion:** This study reveals intriguing new loci determining brain volumes. Replication is underway to confirm these findings.

401T

Improving pedigree-based genotype imputation with simultaneous use of identity-by-descent and linkage disequilibrium information. *R.A. Nafikov[1], E.M. Wijsman[1,2,3].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Biostatistics, University of Washington, Seattle, WA, USA; 3) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Pedigree-based genotype imputation methods rely on identity-by-descent (IBD) information. Genomic regions with IBD originate from the same pedigree founder chromosome, allowing imputation of missing alleles by copying from observed to unobserved alleles that originate from the same inferred founder chromosome. Recent research by our group showed the advantage of also incorporating knowledge about linkage disequilibrium, the main source of information used by population-based imputation methods, to increase the accuracy and completeness of imputed genotypes. Both approaches to imputation start with a haplotype-phasing step, which assigns alleles at different nearby sites to one of the two members of a chromosomal pair in an individual. We are developing an approach that can simultaneously utilize pedigree- and population-based information, with the aim of obtaining improved imputation quality especially in the context of missing founder samples and low-frequency alleles. Missing data is common in family studies, and current models of the genetic architecture of many traits include rare and low-frequency variants. We used our pedigree-based genotype imputation program, GIGI, modified to perform pedigree-based genotype phasing. Each founder chromosome is assigned a unique founder genome label (FGL). We use inheritance vectors based on genome scan markers as represented through the FGLs that are generated with the gl_auto program of the Morgan package. By recording allele assignments to FGL at each marker position, haplotype sequences for each founder chromosome are obtained. In application to a large 5-generation, 189-member pedigree with Alzheimer disease (AD), phasing of 35 rare variants spanning a region-of-interest defined by linkage analysis (LA) identified six potential haplotypes in 10 affected subjects with strong effects on the LA. Of these, four represented >89% of the potential haplotypes. Of course not all FGLs at every marker position get a forced allele assignment, in which case population-based imputation into founder haplotypes is performed with a population-based method such as Beagle, and propagated via the FGLs into family members with missing alleles. We use the phased genotype data from the 1000 Genomes Project as a reference panel. Performance of our joint imputation algorithm is in the process of being evaluated for simulated and real data from AD pedigrees and will be presented.

402F

Whole-genome sequencing in familial late-onset Alzheimer's Disease identifies *TTC3* **and** *FSIP2* **as putative risk loci.** *G.W. Beecham[1], J. Jaworski[1], W. Bush[2], B. Vardarajan[3], E. Blue[4], S. Barral[3], J.L. Haines[2], C.M. van Duijn[5], E.R. Martin[1], G.D. Schellenberg[6], R.P. Mayeux[3], E. Wijsman[4], M.A. Pericak-Vance[1] for the Alzheimer Disease Sequencing Project.* 1) Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst.for Human Genomics, University of Miami, Miami, FL; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 3) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA; 4) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 5) Department of Epidemiology, Erasmus MC, Rotterdam, Netherlands; 6) Perelman School of Medicine University of Pennsylvania, Philadelphia, PA, USA.

Background: The Alzheimer's Disease Sequencing Project (ADSP) is an initiative to identify genetic variation influencing late-onset Alzheimer's disease (LOAD) risk. As part of the ADSP, we performed linkage analysis in 44 non-Hispanic white (NHW) extended families multiply affected by LOAD, followed by whole-genome sequencing (WGS), segregation analysis, and annotation and filtering of variants. **Methods**: WGS data were generated for 229 subjects from 44 NHW families. Standard bioinformatics protocols were applied, with multiple genotype callers used to develop consensus. Family-specific LOD scores were calculated using array data, and parametric multipoint linkage analyses implemented in the MERLIN software. Single nucleotide variants (SNVs) were filtered to families and genomic regions with family-specific LOD scores greater than 1, then annotated for function, frequency, segregation with disease, and information from additional datasets, such as the ADSP case-control dataset. SNVs were annotated with enhancer and expression-QTL data to identify variants likely to influence expression. **Results:** We identified 33 rare (MAF<0.01) SNVs that were under family-specific LOD regions, segregated with disease, were absent from all unaffecteds in the dataset, and were putatively functional (CADD score > 10). These included a missense SNV in TTC3 (CADD=32) that was absent in reference datasets, and confirmed using orthogonal genotyping technology. The FSIP2 gene harbored a missense SNV that segregated with disease among sequenced affecteds, was rare, and likely functional (CADD=25), and showed nominal association in an ADSP case-control gene analysis (p-value=0.043). Additional SNVs in SLC15A5, COG4, CAPZA3, and CONYL1 showed strong functional evidence, in addition to segregation and low MAF. Analysis of intergenic variation identified 17 SNVs that segregated with disease, were rare, and putatively functional (CADD score > 10). Analysis of enhancer data identified multiple enhancer SNVs that segregate with disease and may influence gene expression, including four in a known chr14 linkage region, and two in a known chr1 linkage region; additional enhancer variants were identified in the family-specific regions. **Conclusion**: These results suggest both intra and intergenic SNVs influence LOAD risk in multiplex families, and show the power of segregation-based family designs in WGS studies of complex diseases like AD, and suggest TTC3 and FSIP2 as AD risk genes.

403W

Genome-wide linkage analysis for Parkinson's disease using high-density SNPs in the Faroese cohort. *L.L. Farrell[1], T. Candido[1], S. Bech[2], I. Guella[1], D. Evans[1], J.O. Aasly[3], M.S. Petersen[2], M.J. Farrer[1].* 1) Centre for Applied Neurogenetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Department of Occupational Medicine and Public Health, The Faroese Hospital System, Sigmundargota 5, FO-100 Torshavn, Faroe Islands; 3) Department of Neurology, St Olav's Hospital, Trondheim, Norway.

In the Faroe Islands, the prevalence of Parkinson's disease (PD) is double that of neighboring countries. Disease is assumed to be multifactorial albeit with unknown etiology. Overall, the relative risk for PD is 2.3 for siblings and 1.4 for first cousins. The Faroese cohort is a family-based study of PD that includes several multi-incident pedigrees with detailed genealogy, clinical information and environmental data. Known mutations underlying Parkinsonism have been excluded. In this cohort prior segregation analysis suggests PD best fits an additive or dominant genetic model (Petersen et al., 2015). A subset of 16 families, a total of 303 individuals (124 with DNA, 54 affected) with an average family size of 18.9 (range 10 to 24) were selected for further genetic analyses. Most of the pedigrees are genealogically extensive but prior generations have sparse clinical data and samples. Hence, comprehensive linkage simulations were performed to maximize information content using non-parametric and model-based methods (setting disease frequency to 1/1000, assuming biallelic SNP allele frequencies (0.5) and various penetrance (0.5-0.9), phenocopy (0.01, 0.05) and age-associated liability parameters). Illumina MEGA SNP array genotyping of 1.8M SNPs was prioritized for families with multi-incident PD. After excluding SNPs with significant deviation from HWE ($p<0.01$), MAF<0.05 and LD ($r^2> 0.8$), genome-wide nonparametric and model-based linkage analyses was performed with MERLIN (Abecasis et al.2002), using high-density SNP sets with ~100 kb spacing throughout the genome (24,679 SNPs). Marker allele frequencies were based on all individuals genotyped within the cohort. To date, under non-parametric models evidence of suggestive linkage (LOD >2) was observed at the 137cM to 139cM region on chromosome 7q33.5-q34.5 and 86cM to 88cM region on chromosome 10q23.1-q.23.2 with a maximum LOD score of 2.19 and 2.16, respectively, which was obtained after controlling for marker-marker linkage disequilibrium (LD) with $r^2>0.2$. Model-based linkage analyses are ongoing. For 31 affected individuals, whole exome sequencing has also been performed, primarily to help improve haplotype informatively in loci shared among family members with PD. Overall, results from pedigree/power simulations, linkage and haplotype analyses, informed by exome variants within shared intervals will be presented.

404T

Mosaic loss of chromosome Y (LOY) in peripheral blood is associated with age, smoking, shorter survival and increased risk of cancer and Alzheimer's disease (AD) in men. *L. Forsberg[1], J.-C. Lambert[2], C. Rasi[1], V. Giedraitis[3], H. Davies[1], B. Grenier-Boley[2], C. Lindgren[4], D. Campion[5], C. Dufouil[6], F. Pasquier[7], P. Amouyel[7], L. Lannfelt[3], M. Ingelsson[3], L. Kilander[3], L. Lind[8], J. Dumanski[1], The European Alzheimer's Disease Initiative Investigators.* 1) Dept. Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala, Sweden; 2) INSERM, U1167, Institut Pasteur de Lille, Université de Lille, Lille, France; 3) Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK and Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA; 5) CNR-MAJ, INSERM, U1079, Rouen University Hospital, Rouen, France; 6) INSERM, U708, Victor Segalen University, Bordeaux, France; 7) CNR-MAJ, Université de Lille and Centre Hospitalier Régional Universitaire de Lille, Lille, France; 8) Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Introduction: In the entire world men live on average 4 years shorter lives compared to women. In the developed countries this difference is even larger and the average life expectancy is more than 6 years shorter in men compared with women. This sex-difference has been known for centuries but the underlying factor(s) have been elusive. Our discoveries show that a male specific genetic risk factor, i.e. LOY (mosaic loss of chromosome Y in blood cells), can help explain this observation. Our previous analysis show that LOY is associated with age, smoking, all-cause mortality and non-hematological tumors (*Nat. Genet.* 2014 PMID:24777449, *Science* 2015 PMID:25477213). We can now demonstrate that in addition to this, LOY in blood is also associated with increased risk for Alzheimer's disease (AD) (*In press AJHG*). Materials and Methods: LOY in blood cells was estimated using SNP-array data from >3200 men from one AD case-control study and two prospective cohorts. Strict QC was applied for included data and whole genome sequencing as well as ddPCR was used for experimental validations. 100% concordance in LOY-scoring was achieved between the different platforms. A set of statistical techniques were used to evaluate association between LOY in blood cells and AD diagnosis. Results: LOY was detected in ~17% of participants (median age=73, range=37-96). We found that men with AD diagnosis had a higher degree of LOY mosaicism in the case-control study (adjusted odds ratio=2.80, AD events=606, p=0.0184). Furthermore, analysis of the two prospective cohorts showed that men with LOY in blood at sampling had an increased risk for incident AD during follow-up time (HR=6.80, 95% CI=2.16-21.43, AD events=140, p=0.0011) (*In press AJHG*). Conclusions: Our results suggest that LOY in blood cells is associated with increased risk for both AD and cancer, suggesting a role of LOY in blood cells on disease processes in other tissues, possibly via defective immunosurveillance functions of immune cells without the Y chromosome. Hence, as a male-specific genetic risk factor, LOY might help explain why males on average live shorter than females. Future and on-going studies will put efforts into understanding what happens in cells with LOY and will thus focus on functional aspects at the cellular as well as the organismal levels.

405F

First study of *C2491T FV* **mutation with ischaemic stroke risk in Morocco.** *B. Diakite[1], K. Hamzi[2], W. Hmimech[2], S. Nadifi[2], GMRAVC.* 1) Genetic and Molecular Pathology, FMOS, USTTB, University of Bamako, 1805; 2) Department of Genetics and Molecular Pathology Laboratory (LGPM), Faculty of Medicine and Pharmacy Hassan II University, 9154 Casablanca, Morocco.

   Ischaemic stroke is is a multifactorial disease. It is known that, individual inherited genetic difference in genes involved in the metabolism of lipid or in the process of inflammation and thrombosis play an important role in the development of stroke. Several previous studies have revealed that the distribution of *FV Leiden* mutation remains low or almost absent in some populations (African population) compared to others. The present study evaluated the association of *C2491T FV* mutation with the risk of ischaemic stroke. Genotyping was performed by High Resonance Melting and PCR-RFLP methods. Carriers of *T* mutated allele was associated with a high risk of ischaemic stroke (OR: 3.77, 95%CI = 2.70-5.25; P <0.0001). But this risk was 8.95 fold when the subject carrying of *TT* genotype (P <0.0001) and 4.08 fold with *CT* genotype. According to the risk factors of ischaemic stroke, a positive correlation was observed only with *C2491T FV* mutation and hypertension (P < 0.001). Thus, we can suggest that *C2491T FV* mutation could be a genetic risk factor for Ischaemic Stroke in Moroccan population.

406W

Phenome-wide association study of 28 variants associated with Parkinson's disease in a web-based cohort. *K. Heilbron, B. Alipanahi, P. Cannon.* 23andMe, Mountain View, CA.

   The phenome-wide association study (PheWAS) is a powerful method for detecting phenotypes that are significantly associated with a genetic variant (or other attribute) of interest. A recent genome-wide association study (GWAS) meta-analysis found 28 genetic variants that were significantly associated with Parkinson's disease (PD). By performing PheWAS on these variants, we set out to find traits that share a common genetic etiology with PD and thereby gain insight into the underlying biology of the disease. We performed a PheWAS on the 26 non-HLA SNPs associated with PD. We tested these SNPs for associations with 1,562 phenotypes derived from survey questions that are available to all 23andMe customers who consented to research. We found ten self-reported phenotypes that were significantly associated with at least two of the 26 SNPs (Bonferroni-corrected $P < 0.05$). For four of these phenotypes, all significant associations had the same direction of effect. Family history of PD and status within one's community exhibited positive associations, while heart or metabolic disease and whether one prefers to talk rather than listen during a conversation exhibited negative associations. Follow-up studies are required to replicate these findings and determine whether any of these candidate phenotypes play a causal role in PD.

407T

Identifying genetic modifiers of age at diagnosis in Parkinson disease – A 23andMe study. *K-D.H. Nguyen[1], N.A. Furlotte[2], L. Yu[2], S. John[1], A. Day-Williams[1].* 1) Computational Biology & Genomics, Biogen, Cambridge, MA; 2) 23andMe, Inc., Mountain View, CA.

Background: Age-at-diagnosis (AAD) varies greatly in both familial and sporadic Parkinson Disease (PD); however, the genetic underpinning (if it exists) has not been determined. Mutations in LRRK2 and GBA have been previously implicated in younger AAD in PD patients. Conversely, smoking has been shown to be correlated with later AAD amongst patients. **Objectives:** We want to identify genetic associations affecting AAD in *LRRK2* PD, *GBA* PD and sporadic PD. Additionally, we want to investigate the interaction between genetic contributions and smoking status. **Methodology:** We performed genome-wide association analyses studies (GWASes) of PD AAD for 3 subgroups of adult PD research participants (age ≥30) from the 23andMe cohort: (1) PD participants with known *LRRK2* mutations (n=153), (2) PD participants with known *GBA* mutations (n=151) and (3) sporadic PD participants with neither *LRRK2* nor *GBA* mutations (n=6,835). To investigate the epidemiological association of AAD with smoking we performed the analyses with and without smoking as a covariate to identify both dependent and independent genetic effects. **Results:** The association landscape changed remarkably upon removing the effect of smoking in identifying smoking-independent signals. By removing the effect of smoking from PD AAD regression analysis, we were able to remove spurious (likely false positive) associations and enhance signals for new loci. Despite limited sample size in the LRRK2 and GBA populations, we identified 1 genome-wide significant signal and ~10 suggestive signals in these 2 populations alone. **Conclusions:** This study comprises one of the largest collections of *LRRK2* and *GBA* carriers (N~150-200) to reveal the genetic effects of PD AAD. In addition, we are performing Mendelian randomization to assess the causality of smoking to PD AAD.

408F

Identıfıcatıon of rare variants in recessıve Hereditary Spastic Paraplegia patients from Turkey. *B. Ozes[1], M. Gonzalez[2], S. Zuchner[2], A. Gunduz[3], Y. Parman[4], E. Battaloglu[1].* 1) Bogazici University, Molecular Biology and Genetics Department, Istanbul, Turkey; 2) University of Miami Miller School of Medicine, Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, Miami, USA; 3) İstanbul University, Cerrahpaşa Medical School, Department of Neurology, Istanbul, Turkey; 4) Istanbul University, Istanbul Medical School, Istanbul, Turkey.

Hereditary Spastic Paraplegia (HSP) is a group of clinically and genetically heterogeneous neurodegenerative disorders. Lower limb spasticity and progressive weakness are manifested in pure HSP patients. In complicated HSP patients, other neurological and nonneurological symptoms can be observed. HSP can be inherited in autosomal dominant, autosomal recessive, or X-linked manner. Forty-nine loci and 43 genes are associated with autosomal recessive form of HSP (ARHSP). Whole exome sequencing data of six ARHSP patients from Turkey were analysed to determine the effect of genetic background on phenoytpic heterogeneity among patients. Filtering was performed to determine rare variants via GEM.app[1]. All rare (MAF<0,01), non-synonymous variants in neurodenegeration/neuropathy related genes were selected. We further filtered the variants that were predicted as 'effective' by *in silico* tools. Forty-two rare variants in six individuals were identified and 35 of those variants were located in the genes related to various neurodenegerative/neuropathic diseases such as HSP, CMT and Usher syndrome. Three variations were in homozygous state underlying the effect of rare variants in ARHSP. When whole genome sequencing data of 15 healthy Turkish individuls from Turkish Genome Project[2] were analyzed with the same criteria, 68 rare variant were identified in total. Since the average of rare variant in patients (7 variant per individual) was higher compared to unaffected individuals (4,5 variant per individual), rare variations observed in these patients are probably effective. These rare variants observed with the variant segregating with the disease require further and detailed analyses because they might explain inter- and intra-familial heterogeneity in phenotype. They might also contribute to fulfill the overall picture of mechanism underlying neurodegeneration. This data might also be important to question the approach to "end the genetic analysis after determining the disease-causing variant" because in this approach other variants that might contribute to disease formation are underestimated. 1. Genomes Management Application (GEMapp), University of Miami Miller School of Medicine (https://genomics.med.miami.edu/gem-app/) [June, 2014]. 2. Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S., et al. 2014. Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. BMC Genomics.15:963.

409W

Genetic variants influencing myeloperoxidase levels increase risk of stroke. *C. Phuah[1], T. Dave[2], R. Malik[3], M.R. Raffeld[4], A.M. Ayres[5], B.D. Mitchell[2], B.B. Worrall[6,7], S.J. Kittner[8,9], M. Dichgans[3], N. Rost[5], C.D. Langefeld[10], D. Woo[11], A. Biffi[12,13], J. Rosand[4,5,14,15], C.D. Anderson[4,5,14,15].* 1) Department of Neurology, Washington University School of Medicine in St Louis, St Louis, MO; 2) Department of Medicine and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 3) Institute for Stroke and Dementia Research (ISD), Ludwig-Maximilians-University of Munich, Munich, Germany; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 5) J. Philip Kistler Stroke Research Center, Department of Neurology, Massachusetts General Hospital, Boston, MA; 6) Department of Neurology, University of Virginia, Charlottesville, VA; 7) Department of Public Health Science, University of Virginia, Charlottesville, VA; 8) Department of Neurology, University of Maryland School of Medicine, Baltimore, MD; 9) Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD; 10) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem,NC; 11) University of Cincinnati College of Medicine, Cincinnati, OH; 12) Division of Behavioral Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA; 13) Division of Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA; 14) Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA; 15) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

  Objective: Primary intracerebral hemorrhage (ICH) and lacunar stroke are acute manifestations of arteriolar injury. Myeloperoxidase (MPO) plays a central role in the initiation and progression of vascular inflammation. We hypothesized that genetic determinants of MPO levels influence risk of ICH and lacunar stroke.   **Methods:** We used a discovery cohort of 1409 ICH cases and 1624 controls from 3 studies, an extension cohort of 12,577 IS cases and 25,643 controls from NINDS-SiGN, and a validation cohort of 10,307 IS cases and 29,326 controls from METASTROKE. An unweighted MPO-increasing genetic risk score (GRS) was constructed from fifteen single nucleotide polymorphisms (SNPs) arising from prior genome-wide studies of circulating MPO levels ($p < 5 \times 10^{-8}$). We used multivariable regression models for association between the MPO-GRS and ICH and IS subtypes, and fixed-effects meta-analyses to pool estimates across studies. We utilized Cox regression models in a prospective cohort of 174 ICH survivors for association with ICH recurrence. **Results:** Genetic determinants of elevated circulating MPO levels were associated with both ICH risk (odds ratio [OR] 1.07, *p*=0.04) and recurrent ICH risk (hazards ratio [HR] 1.45, *p*=0.006). Analysis of IS subtypes demonstrated association with MPO-GRS in only the lacunar subtype (OR 1.05, *p*=0.0012). **Interpretation:** Genetic variants that increase circulating MPO levels increase risk of ICH and lacunar stroke. Because genetic variants are not influenced by environmental exposures, these results provide new evidence for a causal rather than bystander role for MPO in the progression of cerebrovascular disease. Furthermore, this also suggests that chronic inflammation may be a potential modifiable stroke mechanism, and immune-targeted therapies could be useful for treatment and prevention of cerebrovascular disease.

410T

Identifying high genetic risk groups for personalized medicine in Alzheimer's disease. *S.J. van der Lee[1], F.J. Wolters[1], A. Hofman[1,2], M.A. Ikram[1], N. Amin[1], C.M. van Duijn[1].* 1) Erasmus MC, Rotterdam, Rotterdam, Netherlands; 2) Harvard T.H. Chan School of Public Health, Boston, MA, USA.

  Purpose: The gene encoding the apolipoprotein E protein (APOE) is a major determinant for Alzheimer's disease (AD). While the APOE*4 variant associate with an increased risk of AD, the APOE*2 variant associates with a decreased risk. In the past decade genome wide association and sequencing studies identified 24 other genetic variants reaching genome wide significance. These include rare variants with effects similar to APOE but predominantly common variants with small effects. Until now, it is unclear to what extend these variants jointly modify the risk of AD.   **Methods:** We calculated the lifetime cumulative incidence of AD in the Rotterdam study. This prospective cohort study included 12255 subjects (67.48 ± 8.4 years at inclusion) free of AD at age 60. After 12.9 ± 6.3 years (maximum 24 years) of follow-up, 1262 participants developed AD. All participants were characterized for common and rare variants using direct genotyping and imputation. Most recent published effect sizes of new variants were used to calculate an AD specific genetic risk score (AD-GRS) capturing the joint additive effect of the novel variants. Kaplan-Meier-analyses were used to calculate cumulative risk lifetime by age of 85 years, overall and stratified by APOE genotype.   **Results:** We found that 58.6% of the APOE*44 carriers developed AD prior by age 85 years while the risk of APOE*22/23 carriers by age 85 is as low as 6.2%. Overall, there was a significant effect of the AD-GRS on the risk of AD (low risk tertile to the high risk tertile p-value = 1.0×10^{-12}). For those in the low tertile of the AD-GRS the risk of AD was 8.7% by age 85 while the cumulative for those in the high tertile was 17.7%. The cumulative incidence of the low tertile reached 17.7% by age 88.7, translating into a 3.7 year difference in AD-free survival. Although the AD-GRS differentiates the risk of AD across all APOE genotypes, there was evidence for interaction with APOE genotypes (p = 0.01). APOE*44 carriers in the low tertile of the AD-GRS on average developed AD eight years later than those in the high AD-GRS tertile. Of the APOE*44 carriers in the high risk tertile, 82.9% had developed AD by age 85 years given they survived to this age. **Conclusions:** Combining the effects of multiple low and moderate risk variants can identify those at high life-time risk of AD. The results can be used for clinical counseling and patient selection for clinical intervention trials.

411F

Refining the role of *de novo* **protein truncating variants in neurodevelopmental disorders using population reference samples.** *J.A. Kosmicki[1,2,3], K.E. Samocha[1,2,3], D.P. Howrigan[2,3], D.G. MacArthur[2,3], D.P. Wall[4], E.B. Robinson[2,3], M.J. Daly[1,2,3].* 1) Harvard Medical School, Boston, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Stanford University, Palo Alto, CA.

One of the fundamental challenges of human disease genetics is differentiating risk-conferring variants from the overwhelming amount of neutral variation in the genome. *De novo* variants comprise a unique component of the genetic architecture of human disease since, not having yet passed through a single generation, any heterozygous variants with complete or near-complete elimination of reproductive fitness must reside almost exclusively in this category. As such, recent studies found that *de novo* protein truncating variants (PTVs) play a significant role in neurodevelopmental disorders. Despite prior evidence of mutational recurrence (i.e., the same mutation occurring *de novo* multiple times), most studies implicitly assumed each *de novo* variant was novel, in line with Kimura's infinite sites model, and thereafter analyzed *de novo* variants genome-wide without respect to their allele frequency in the population. We examined these assumptions using data from 9477 families with ASD, intellectual disability (ID), or developmental delay (DD), and found that overall ~1/3 of *de novo* variants are independently observed as standing variation in the Exome Aggregation Consortium's (ExAC) cohort of 60,706 adults without severe neurodevelopmental disease. *De novo* PTVs absent from ExAC are more strongly associated with neurodevelopmental risk than those present in ExAC, evidenced by enhanced associations in ASD (rate ratio [RR]=1.8, $P<10^{-12}$) and ID/DD (RR=2.6, $P<10^{-21}$). In contrast, *de novo* PTVs present in ExAC do not exhibit enrichment in ASD (RR=0.67, $P=0.98$) or ID/DD (RR=0.69, $P=0.94$). We further use a loss-of-function intolerance metric derived from ExAC to identify a small subset of genes that contain the entire signal of associated *de novo* PTVs in ASD (RR=3.2, $P<10^{-16}$) and ID/DD (RR=6.7, $P<10^{-38}$). These same genes carry a moderate excess of PTVs inherited by individuals with ASD from their unaffected parents (RR=1.2, $P=0.01$), and a larger excess in 404 ASD cases and 3654 controls (OR=2.7, $P<10^{-19}$), with no association seen outside these genes. Furthermore, of the top 11 genes with ≥3 *de novo* PTVs in ASD, all but 4 (*DSCAM*, *ANK2*, *CHD8*, *ASH1L*) are more significantly enriched in ID/DD (OR=9.3, $P<10^{-6}$), indicating the majority of genes previously implicated in ASD using *de novo* variants are associated with ID/DD, not ASD. Collectively, these findings illustrate the importance of population-based cohorts as reference samples for even *de novo* variation.

412W

From one family to replication in five data sets: Two loci associated with age-at-onset of familial and sporadic Alzheimer disease. *E. Blue[1], E. Wijsman[1,2,3], T. Bird[4,5,6], C-E. Yu[7], T. Thornton[2].* 1) Div Med Gen, Univ. of Washington, Seattle, WA; 2) Dept Biostat, Univ. of Washington, Seattle, WA; 3) Dept Genome Sci, Univ of Washington, Seattle, WA; 4) Dept Medicine, Univ of Washington, Seattle, WA; 5) Dept Neurology, Univ of Washington, Seattle, WA; 6) Dept Psych Behav Sci, Univ of Washington, Seattle, WA; 7) Div Geront Geriatr Med, Univ of Washington, Seattle, WA.

We have fine-mapped a linkage region in a single family, evaluated sequence variants within the region, and replicated association at 2 loci in 5 data sets. These data sets include subjects with early- (EO) and late-onset (LO) Alzheimer disease (AD), with and without family history, and European or Hispanic ancestry. We began with strong linkage signals on chromosomes (chr) 1 and 17 for age-at-onset (AAO) modifiers in a German from Russia (GFR) family with EOAD caused by the *PSEN2* N141I variant. We obtained genome-wide SNP data on 32 members of the family and combined a measured genotype (MG) approach with identity-by-state (IBS) and identity-by-descent estimates to reduce to one tenth the size of the linkage regions in that family. We sequenced the exomes of 6 members of this family, and the genomes of 4 of these, maximizing variant detection and phasing information for the linked haplotypes. We applied bioinformatic and IBS filters as well as the MG approach to derive a short list of candidate variants. We genotyped these 54 variants in 331 subjects of GFR ancestry ascertained from both LO and *PSEN2* N141I families, including the discovery family. We tested for association between the residuals from a Cox proportional hazards analysis with AAO of AD adjusted for both *APOE* and *PSEN2*, incorporating known pedigree information. We found 13 variants that were both nominally significant in the discovery family and in the larger GFR sample. Using a similar model, we tested for association between AAO of AD and these loci (+/-25kb) in exome sequence variants observed in 5,567 cases and 4,961 controls with non-Hispanic European (EUR) ancestry from the Alzheimer's Disease Sequencing Project. We found highly significant ($p<10^{-9}$) association between variants surrounding two GFR loci: a promoter of *NCSTN* (chr1q23.2) and a promoter of *KDM6B* (chr17p13.1). We performed similar analyses of SNP array-focused data in a total of 4,203 EUR and 5,174 Caribbean Hispanic (CH) subjects from 5 different LOAD data sets encompassing both familial and unrelated subjects. We found nominally significant evidence for association between AAO of AD and SNPs surrounding the *NCSTN* locus with SNPs in each of one EUR and one CH sample, and surrounding the *KDM6B* locus in each of 2 CH and 2 EUR data sets. We therefore show that loci associated with phenotypic variation in a Mendelian EO form of AD have similar effects in cases with sporadic or familial LOAD as well as across populations.

413T

Identifying novel genes whose tissue-specific expression level causally influence complex human traits. *E. Porcu[1,2], A. Reymond[1], Z. Kutalik[2,3].* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Institute of Social and Preventive Medicine, CHUV and University of Lausanne, Lausanne, Switzerland.

In the last decade Genome-Wide association studies (GWAS) identified thousands of variants associated with hundreds of complex traits but in many cases the underlying biological reason for such association is unknown. Given that many GWAS loci fall far from coding regions of genes and the importance of regulatory variations in shaping complex phenotypes, we propose to use gene expression quantitative trait loci (eQTLs) to interpret the role of variants in a GWAS region and the complex trait. Recently, a study (Zhu et al. Nature Genetics 48, 481–487 (2016)) applied Mendelian Randomization (MR) approach using a single instrumental variable to search for the most functionally significant genes at the loci associated in GWAS for complex phenotypes by integrating association summary statistics and gene expression eQTL data. Here, we propose an advanced MR approach that uses multiple SNPs jointly as instruments and multiple gene expression traits simultaneously in multiple tissues. Our method is not only more powerful than the previously one, but verifies MR assumptions more rigorously. We applied the method to schizophrenia using summary data from the latest GWAS meta-analyses and various eQTLs datasets and identified more than 200 genes causally implicated in schizophrenia in various tissues. Whereas we replicated 6 out of the 17 signals reported by Zhu and colleagues, for the remaining 11 SNPs we cannot ensure that the MR assumption of absence of pleiotropy holds. Among our results we find examples of genes with a) overestimated effect in single-gene analysis: e.g. *SPCS1*, p=1.2x10^{-34}; b) significant association only in the multiple-gene analysis: e.g. *EFNA1*, p=7.3x10^{-8} ; c) tissue-specific causal effects: e.g. *INPP5B* associated with schizophrenia only using brain eQTLs, p=8.9x10^{-20} ; and d) not significant association in previous GWAS analysis: e.g. *L3HYPDH*, p=1.6x10^{-16}. These results suggest that a multi-SNP and multiple-gene approach verifies MR assumptions more rigorously. It allows unraveling new insights in the tissue-specific transcriptomic regulation of complex traits. Such mechanistic insight should facilitate tissue-specific pathway enrichment analysis. These observations confirm the importance of integrating data from various tissues when trying to interpret GWAS results using gene expression as an intermediate phenotype.

414F

Exploring the genetic architecture of ADHD by assessing the contribution of non-additive genetic effects and the role of rare coding variation. *D.S. Palmer[1,2], A. Bloemendal[1,2], D. Demontis[3,4,5], D.M. Hougaard[3,4,5,6], A.D. Børglum[3,4,5], B.M. Neale[1,2,7], iPSYCH-Broad Consortium.* 1) Broad Institute, Boston, MA, USA; 2) ATGU, Massachusetts General Hospital, Boston, MA, USA; 3) iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; 4) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 5) iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; 6) Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 7) Harvard Medical School, Boston, MA, USA.

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterized by hyperactivity, impulsivity and inattentiveness. Current twin study estimates place heritability at ~0.76 [1], while SNP-based approaches estimate the additive contribution of SNP heritability at ~0.28 [2]. In order to more thoroughly characterize the genetic architecture of ADHD, we examine two additional sources of variance: non-additive genetic effects of common variants and rare coding variation from exome sequencing. SNP-based approaches such as GREML [3] and LD score regression [4] assume a polygenic model, where a large number of small effects combine additively toward a continuous phenotype or case/control status. Here, we extend this 'polygenic model' to allow for dominance effects upon the observed phenotype. Based on the variance decomposition in Fisher's original biometrical model [5], we augment the polygenic model with terms that act orthogonally to the additive genetic effects. We develop an extension of the method of LD score regression to estimate the additional heritability explained, and apply it to genotype data from iPSYCH of 12,000 cases and 20,000 controls. We report on the increase in heritability explained by accounting for dominance effects in our analysis. To further explore the contribution to genetic architecture, we also examine exome sequence data from 3,500 ADHD cases and 6,000 controls. In this study, whole exome data is obtained from stored bloodspots taken from the Danish Neonatal Screening Biobank at Statens Serum Institut, Denmark. Using these data, we find that cases of ADHD show an enriched rate of ultra-rare disruptive and damaging mutations (OR=1.43, p=7e-6) but these mutations explain less than 1% of the variance of ADHD.[1] Faraone et al, 2005. Molecular genetics of attention-deficit/hyperactivity disorder. Biological Psychiatry.[2] Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature Genetics.[3] Yang et al, 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics.[4] Bulik-Sullivan et al, 2015. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics.[5] Fisher, 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. Phil. Trans. R. Soc. Edinburgh. .

415W

De novo mutations account for a small fraction of the epidemiologic association between autism and paternal age. *E.B. Robinson[1,2,3], D.J. Weiner[1,2,3], E.M. Wigdor[1,2,3], S. Ripke[1,2,3,4], J.A. Kosmicki[1,2,3,5], K.E. Samocha[1,2,3,6], A. Bloemendal[1,2,3], M.J. Daly[1,2,3].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA; 3) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Boston, MA; 4) Department of Psychiatry and Psychotherapy, Charite, Campus Mitte, Berlin, Germany; 5) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 6) Program in Genetics and Genomics, Biological and Biomedical Sciences, Harvard Medical School, Boston, MA.

Advanced paternal age is consistently linked to risk for autism spectrum disorders (ASDs). For example, one epidemiologic meta-analysis suggested that men above age 50 are approximately 2.5 times as likely as men under 30 to have a child diagnosed with ASD (Hultman et al. 2011). It is commonly assumed that this increase in risk is driven by *de novo* mutations, which also increase with paternal age. However, a recent test of this assumption used simulated data to estimate that *de novo* mutations were not likely to account for more than 20% of the observed risk relationship between paternal age and ASD (Gratten et al. 2016).   Using **empirical** exome sequence and genotype data from over 2500 families from the Simons Simplex Collections (SSC), we aimed to directly estimate the amount of paternal age risk that could be attributable to *de novo* mutations. We also aimed to directly test the hypothesis that men who have children later in life could have greater common, polygenic risk for ASD, an alternate genetic route to advanced paternal age risk.   In the SSC, nonsynonymous *de novo* mutations in the exome are 20% more common in affected probands than in their unaffected siblings (odds ratio (OR)~1.2). Consistent with previous estimates (Kong et al. 2012), we found that nonsynonymous de novo mutations in controls increase at a rate of approximately 3% per year of paternal age. Modeling this observed increase against the empirical OR, we find that less than 10% of the epidemiologic association between paternal age and ASD risk can be explained by *de novo* mutations in the exome.   Most genetic risk for ASDs reflects common, polygenic variation (Gaugler et al. 2014). We estimated the SSC parents' polygenic risk for ASD, as well as for educational attainment and schizophrenia (traits genetically correlated with ASD; Bulik-Sullivan et al. 2015). Increasing paternal age was significantly associated with higher polygenic risk for schizophrenia (p=0.004) and educational attainment (p=0.001), but not ASDs (p=0.98). These analyses suggest that 1) *de novo* mutations in the exome account for a limited component of the epidemiologic association between paternal age and ASD risk, and 2) some of the association is being driven by complex, polygenic variation, an unchanging property of parents that influences timing of childbearing. These findings have significant implications for public health messaging and genetic studies of ASD.

416T

Genetic influence on age of onset of obsessive-compulsive disorder. *H. Qin[1], K. Wen[1], M. Ritter[1], G. Nestadt[2], Y. Yao[1].* 1) National Institute of Mental Health (NIMH/NIH), Bethesda, MD; 2) Johns Hopkins University School of Medicine, Baltimore, MD.

Obsessive-compulsive disorder (OCD) consists of obsession and compulsion domains, characterized by frequent upsetting thoughts and rituals/repeat behaviors that to relieve the anxiety caused by an obsession. The National Comorbidity Survey Replication (NCS-R) estimated in 2005 that 1.0% of the US adult population had been affected by OCD in the past 12 months. Individuals with OCD begin to develop symptoms during childhood or adolescence. Genetic predictors for the young onset OCD are solely needed. In addition, genetic studies of young onset OCDs may lead to understanding of the genetic underpinnings of OCD development. Our OCD Collaborative Genetics Association Study project (OCGAS) recruited families with at least two OCD-affected siblings from outpatient and inpatient clinics. Subjects were diagnosed as "affected" if they had met DSM-IV OCD diagnosis criteria at any point in their lives. Diagnostic assessments were conducted using the OCD Collaborative Genetics Study Assessment Package. Probands were included if age of onset ≤18. We hypothesized that young onset OCDs might be more genetically driven. To take the advantage of the unique characteristics of our samples, we conducted a genetic analysis to explore the potential relationship between age of onset and genetic variants in a sample of OCD patients (n = 1,711). We categorized patients into two groups using the age of onset of 8 as threshold to determine the early onset (≤8 years old) and late onset (>8 years old). We used the PLINK software to analyze genotypic and phenotypic data and tested for potential association. As a result, we found a potential correlation between the SNP rs4656781 and the age of onset of OCD (early vs. late onset), with p-value of 6.58×10^{-7}. The SNP is located in SPON1 gene. A recent study showed that haplotypes on SPON1 were associated with plasma concentrations of Methadone R- and S-enantiomers in heroin-dependent patients. Also, SPON1 is known as a cell adhesion protein contributing to the growth and guidance of axons in both the spinal cord and the peripheral nervous system. However, we need note the limitations of our study, including relatively small sample size, and the self-reported age of onset. To conclude, albeit the fact that our study did not identify a genome-wide level significant SNP, the preliminary results presented gave a suggestive association between genetic variations and age of onset. This potential finding warrant further investigations. .

417F

A semi-supervised learning approach to functional annotations for coding and noncoding variants. *Z. Liu[1], J. Zou[2], I. Ionita-Laza[3].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 2) Microsoft Research New England and MIT, One Memorial Drive, Cambridge, MA 02142; 3) Department of Biostatistics, Columbia Mailman School of Public Health, NYC, NY.

   Predicting and interpreting the pathogenicity of mutations is a fundamental challenge in human genetics. Great efforts have been devoted to generate functional annotation of variation in the human genome sequence through either high-throughput experiments such as the ENCODE project or computational predictions such as genomic conservation. Such annotations can help identify putatively causal genetic variants obtained from genetic association studies. There has been intense research efforts to develop supervised and unsupervised methods that integrates functional annotations predict the effects of genetic variants. Supervised methods use variants that are known to be pathogenic or benign to train prediction algorithms. The major challenge with this approach that only a small proportion of genetic variants can be labeled reliably with its effects on humans. Methods such as CADD uses evolutionary conservation as a surrogate for true pathogenicity. Recently we have proposed two unsupervised approaches named Eigen and Eigen-PC, that do not make use of any labeled data, and tend to outperform CADD. However, its prediction accuracy can be improved if the information contained in the small proportion of high quality labeled data is properly used. Here, we propose a novel semi-supervised co-training approach to derive an annotation score that, unlike any existing methods, can integrate both the labeled and unlabeled data to increase predictive accuracy. This framework allows us to fully capture the complementary benefits of previous methods. We show that the resulting annotation score has better discriminatory power than Eigen and Eigen-PC on noncoding variants from GWAS and eQTL studies, as well as in coding mutations in Mendelian neuropsychiatric diseases.

418W

Whole genome analysis of the genetic relationship between attention deficit disorder and obsessive compulsive disorder. *M. Ritter[1], G. Nestadt[2], W. Guo[1], H. Qin[1], Y. Yao Shugart[1].* 1) Unit on Statistical Genomics, National Institute of Mental Health, NIH, Bethesda, MD; 2) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD.

   Attention deficit disorder (ADHD) and obsessive compulsive disorder (OCD) are both highly heritable, neurodevelopmental disorders that onset in childhood. The comorbidity of ADHD and OCD has often been observed in pediatric populations, ranging from 10-50%. In this study, we aimed to examine the genetic relationship between ADHD and OCD. In particular, this was done by calculating polygenic risk scores. To take advantage of the publically available data shared by the Broad Institute, we downloaded the summary statistics from the PGC-GWAS analysis and used that dataset to predict the susceptibility of OCD. The OCD dataset includes 2,998 individuals from nuclear families. The ADHD dataset contains 2,960 cases, as well as parental and independent controls. Our meta –analysis did not detect any genome-wide significant signals. However, while conducting protein-protein interaction analyses, several proteins were predicted to play a role in a complex network that may jointly contribute to ADHD and OCD susceptibility. We would like to recognize that the limitations of using polygenic risk scores and meta analyses to predict psychiatric diseases require further work and a larger sample size. This will be done by combining other relevant datasets into the above mentioned ones.

419T

Identification of novel genetic variants of *DSM-5* Nicotine Use disorder: Exome array analysis in National Epidemiology Survey on Alcohol and Related Conditions-III. *J. Jung, H. Zhang, B. Grant.* Laboratory of Epidemiology and Biometry, NIH/NIAAA, Rockville, MD.

DSM-5 nicotine use disorder (NUD) is a complex addictive disorder that affects about 32% of European Americans (EA) and 24% of African Americans (AA). Moderate to severe DSM-5 NUD and DSM-4 nicotine dependence (ND) have a concordance rate of 93% and DSM-4 ND has heritability of 50%-72% based on twin studies. We analyzed data from the National Epidemiologic Survey on Alcohol and Related Condition-III (NESARC-III) to perform an exome array study of nicotine use disorder with moderate to severe symptoms (NUD >=4). EA and AA in NESARC-III samples were genotyped by Affymetrix Axiom Exome Chip 319K and 100K additional customized SNPs. EA samples of 3,470 cases and 4,078 supernormal controls assessed DSM-5 with no identified psychiatric disorders, and AA samples of 733 cases and 1,766 supernormal controls were analyzed. A single SNP based analysis with common variants (minor allele frequency (MAF)>=0.01) was performed to test associations with moderate to severe NUD using an additive model, and a gene (regional) based analysis with rare variants (MAF<0.05) was performed using a Sequence Kernel Association Test (SKAT) after controlling for sex, age, family income, marital status, education, and two population stratification scores. In a gene based analysis using SKAT-OPT method, we found 21 genes with p-value <2.5e-6 of a gene level threshold in EA and 7 of them (*BCAP29, FER1L6, LRRFIP2, MAP4K3, OMD, SPOCK2, SPTBN5*) were detected in AA with p-value <0.005. Based on a single SNP analysis, we identified 12 novel SNPs in EA with p-value <2e-7: *MAP4K3, WDR17, BAZ2B, FOLR2, EIF4E2, ZAK, LRRFIP2, DEPDC5, KIF25, OR5K4, SPTBN5, SLC22A15* and 7 of 12 identified genes in EA were significant in AA with p-value <0.007; *MAP4K3* (rs140914094)*, FOLR2* (rs150788760)*, ZAK*(rs149819262)*, LRRFIP2* (rs144038552)*, DEPDC5* (rs199749859)*, OR5K4* (rs78102601) *, SPTBN5* (rs200664511). Both the gene based approach and the SNP based approach detected associations between *LRRFIP2, MAP4K3,* and *SPTBN5* with NUD in both EA and AA. In addition, we confirmed the previous findings on association of *SPOCK2, MAP4K3* with nicotine-related phenotype such as smoking behavior or lung cancer. Our preliminary data identified novel associations that promises to improve understanding of the etiology of NUD.

420F

Vertex-wise heritability of human brain cortical thickness and surface area using a twin and non-twin siblings design. *S. Patel[1,2], M. Park[3,4], G.A. Devenyi[3], R. Patel[3], J. Knight[1,2,5], M.M. Chakravarty[3,6].* 1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 3) Cerebral Imaging Centre, Douglas Mental Health University Institute, McGill University, Verdun, QC, Canada; 4) Schulich School of Medicine and Dentistry, Western University, London, ON, Canada; 5) Lancaster Medical School, Faculty of Health and Medicine, Lancaster University, Lancaster, UK; 6) Department of Psychiatry, McGill University, Montreal, QC, Canada.

Introduction Variation in complex phenotypes such as neuroanatomical measurements can be accounted for by three factors: genetics (A), shared (C) and unique environment (E). The genetic component is defined as heritability and can be quantified using twin studies. The volume of the whole brain is highly heritable with a heritability score of 90% [Baare et al., 2001]. Investigating the heritability of vertex based cortical thickness (CTh) and surface area (SA) allows us to explore the genetics influence across brain regions to make inference and interpretation of studies relating behavior and neuropsychiatric disorders to the same structural phenotypes. **Methods Database:** High-resolution 3 Tesla, T1 weighted magnetic resonance images (MRI; 0.7mm isotropic) from the Human Connectome Project (HCP) [Van Essen et al., 2013] were used in this study (542 healthy subjects: 126 monozygotic twins, 144 dizygotic twins and 272 siblings; age range of 22-36 years old). **Image processing**: CIVET 1.1.12 pipeline [Collins et al., 1994, Lyttelton et al, 2007] was used to measure CTh and SA at each vertex (40,962 in each hemisphere). Each voxel is classified into either white matter (WM), gray matter (GM) or cerebrospinal fluid (CSF). CTh is calculated by the distance in millimeters between WM surface (interface between GM and WM) and GM surface (interface between GM and CSF) at each vertex. The middle cortical surface which is at the geometric center between the inner and outer cortical surface is used to measure SA. **Heritability calculations:** Broad-sense heritability of vertex wise CTh and SA was estimated using structural equation modeling implemented with OpenMx package [Neale et al., 2015] within R. A univariate model controlling for sex and age was used to examine the genetics, shared and unique environment. **Results** Preliminary results demonstrate the overall heritability of SA was higher than CTh. The posterior cingulate gyrus had a mean SA heritability score of 69% (SD± 0.07). The highest mean heritability score of CTh was within the paracentral lobule of 49% (SD± 0.11). **Conclusion** A continuous heritability spatial brain map is created from vertex based CTh and SA. The preliminary results suggest that SA and CTh can be partly accounted by genetic variation. Currently we are working with a larger HCP dataset (n=970) to identify potential quantitative phenotypes to be used in neuropsychiatric disorders based on heritability scores of SA and CTh.

421W

Integrative analysis of eQTL and GWAS detects novel genes having pleiotropic effects on schizophrenia and gene regulations. *Q. Wang[1,2,3], J. Gelernter[2,3,4,5], H. Zhao[1,4,6,7].* 1) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; 2) Department of Psychiatry, Yale University School of Medicine, West Haven, CT, USA; 3) VA CT Healthcare Center, West Haven, CT, USA; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; 5) Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA; 6) Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA; 7) VA Cooperative Studies Program Coordinating Center, West Haven, CT, USA.

Although genome-wide association studies (GWAS) have identified many SNPs associated with psychiatric disorders, some have been difficult to replicate and interpret functionally. This is due in part to polygenicity; therefore, combining signals at the gene level should aid in interpretation and would have the further potential to reveal underlying mechanisms. In search of disease-associated genes, a number of studies aimed to identify differential expression between cases and controls have yielded helpful insights. However, they have frequently suffered from confounding environmental factors; thus the directionality of observed differential expression as consequences of, or causal, genetic factors cannot be interpreted. Therefore, there is need for statistical methods that can identify genes whose expression levels share genetic factors with psychiatric disorders. We developed a novel method that integrates GWAS summary statistics and expression quantitative trait locus (eQTL) summary statistics to identify genes with pleiotropic effects. Observed GWAS signals in genes are combined based on their eQTLs, and linkage disequilibrium (LD) information extracted from 1000 Genomes Project data was used in testing the level of statistical significance. Simulations under multiple settings indicate well-controlled type I error rate and good statistical power for this method. We applied our method to study the genetics of schizophrenia, using GWAS summary statistics from the Psychiatric Genomics Consortium (PGC), *cis*-eQTLs from the Genotype-Tissue Expression Project (GTEx), and *cis-/trans*-eQTLs from the seeQTL database. A list of novel genes that are not significant in VEGAS gene based test were identified by our method, including 108 genes using GTEx eQTLs, 14 genes using seeQTL cis-eQTLs and 9 genes using seeQTL *trans*-eQTLs. For example, a mitochondrial gene, *NDUFC1,* was implicated with combined statistical significance level p<1e-6 using seeQTL *trans*-eQTLs, and has been reported to be differentially expressed in schizophrenia cases vs. controls. We also found that genes prioritized in our method are enriched in the set of differentially expressed genes in a tissue specific manner, using schizophrenia gene expression datasets from brain (GSE17612) and blood (GSE27383). Our results indicate the importance of both *trans*-eQTLs and *cis*-eQTLs in capturing pleiotropic associations, and might lead to further insights in the pathophysiological mechanisms of schizophrenia.

422T

Mediation methods applied to post-traumatic stress disorder to identify genomic effects. *S. Gaynor[1], G. Guffanti[3], T. Jovanovic[2], L. Almli[2], A. Lori[2], A. Wingo[2], E. Binder[2], K. Ressler[2,3].* 1) Harvard University, Boston, MA; 2) Emory University, Atlanta, GA; 3) McLean Hospital, Belmont, MA.

Mediation methods have been developed to characterize causal relationships that decompose a total effect into natural direct and indirect effects. Recent studies have suggested that genomic analyses may be improved by jointly analyzing SNP and gene expression data to analyze phenotypes of complex diseases. We are able to perform such a joint analysis via the mediation framework in order to: (1) elucidate the complete relationship between genomics and diseases, (2) detect biologically important effects from genetic disease-causing mechanisms and (3) better assess contributions by weak genomic effects. Genetics of psychological disorders have been particularly challenged by the weakness of sparse effects across the genome. We apply the mediation approach to the Grady Trauma Project (GTP), a cross-sectional study of post-traumatic stress disorder (PTSD). PTSD is an anxiety disorder that can occur after an individual experiences or witnesses a traumatic event. Individuals respond differently to traumatic stress and it has thus been suggested to have a genetic component. GTP has been collecting genetic, genomic and clinical data on minority individuals in Atlanta, GA that have been highly traumatized. We will leverage the GTP data to identify biological mechanisms and mutations via mediation analysis in order to better biologically characterize PTSD.

423F

Genome-wide polygenic atlas of the phenome in emerging adulthood. *A. Moscati[1], A.R. Docherty[1], D.M. Dick[2], J.E. Savage[1], J. Salvatore[2], M.E. Cooke[2], F. Aliev[2], A.A. Moore[1], A.C. Edwards[1], D.E. Adkins[3], B.P. Riley[1], B.T. Webb[1], S.A. Bacanu[1], K.S. Kendler[1].* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA; 2) Virginia Commonwealth University Department of Psychology, Richmond, VA; 3) Virginia Commonwealth University Center for Biomarker Research and Personalized Medicine, Richmond, VA.

Goal: Identifying genetic relationships between complex traits in emerging adulthood can provide useful etiological insights into risk for psychopathology and other adverse outcomes. This study examined genomic and clinical assessment data from a large sample of young adults (The VCU Student Survey, a cohort sequential assessment of behavior and substance use in college undergraduates, N = 5,951) to construct an atlas of polygenic risk that indexed all clinical, health factor, and outcome data. Methods Used: This study used genome-wide association studies of 34 diverse psychiatric phenotypes and health factors as discovery samples to calculate genome-wide polygenic scores (GPS) and used them to predict 55 phenotypes in the sample of undergraduate young adults. A special emphasis was placed on *a priori* testing of previously published phenotypic and genetic relationships. All analyses were tested separately in each ancestry group in the sample (African ancestry, native populations of the Americas, East Asian ancestry, European ancestry and South Asian ancestry) and corrected for multiple testing within group. GPSs were also correlated with one another to examine polygenic relationships between profile scores in the young adult sample Summary of Results: The analyses resulted in over 1,800 associations between GPS and phenotype, with over 80 reaching robust significance. The majority of *a priori* hypotheses were replicated. A number of notable findings emerged beyond the expected within-trait prediction (GPS for height and body mass index predicted phenotypic height and BMI, respectively). The GPS for schizophrenia predicted depressive symptoms, anxiety symptoms, and nicotine use, as well as experiences of interpersonal trauma and family history of mental health problems. Furthermore, significant positive associations were observed between schizophrenia, bipolar disorder, and major depressive disorder GPS. Conversely, the subjective well-being GPS predicted fewer depressive symptoms, fewer anxiety symptoms, less family history of mental health problems, as well as higher social support and relationship satisfaction. Many of these associations were consistent across all ancestry groups. These analyses highlight the power of a careful polygenic modeling framework in younger samples, and provide potential avenues for prediction of risk and resilience in emerging adulthood.

424W

Investigating the inflammatory hypothesis of depression using Mendelian randomization in the Avon Longitudinal Study of Parents and Children. *H.M. Sallis[1,2], L. Paternoster[1], J. Evans[2], G. Davey-Smith[1].* 1) MRC IEU, School of Social and Community Medicine, University of Bristol, Bristol, Bristol, United Kingdom; 2) Centre for Academic Mental Health, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

To date, our understanding of the aetiology of depression and its mechanisms is relatively limited. However, there is currently much interest in the 'inflammatory hypothesis' of depression, which postulates that depressive symptoms could stem from an increase in levels of inflammatory markers such as pro-inflammatory cytokines. Many groups have reported increased levels of inflammatory biomarkers, including tumor necrosis factor (TNF)-α, c-reactive protein (CRP) and interleukin (IL)-6, among patients with a depressive disorder or increased depressive symptoms. However, the majority of these are observational associations from cross-sectional data. It is thus unclear whether these levels of circulating inflammatory markers are a potential cause, or effect, of the mood disorder itself, or simply a chance finding. Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), we have investigated this hypothesis by looking at the association between depression and a number of exposures with known inflammatory responses and strong genetic instruments, using a combination of Mendelian randomization (MR) approaches. These exposures were interleukin (IL)-6 (a pro-inflammatory cytokine), vitamin D, which is thought to have anti-inflammatory properties, and the chronic, inflammatory skin condition, eczema. For each exposure, we looked at the instrument-outcome association and a 2 stage least squares regression. For the IL-6 analyses, we were also able to incorporate data from 2 independent European cohorts to perform a 2-sample MR. Initial results from our MR analyses suggest some evidence of an association between increased IL-6 levels and depressive symptoms, and between decreased vitamin D and depressive symptoms, while the link between eczema and depression is less clear. Viewed together, these results may provide some support for the inflammatory hypothesis of depression. 1 Maier SF, Watkins LR. *Psychol Rev* 1998; **105**: 83–107. 2 Smith RS. *Med Hypotheses* 1991; **35**: 298–306. 3 Howren MB, Lamkin DM, Suls J. *Psychosom Med* 2009; **71**: 171–86. 4 Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK *et al. Biol Psychiatry* 2010; **67**: 446–57. 5 Liu Y, Ho RC-M, Mak A. *J Affect Disord* 2012; **139**: 230–9. 6 Valkanova V, Ebmeier KP, Allan CL. *J Affect Disord* 2013; **150**: 736–44.

425T

Shared genetic effects between clinical ADHD and smoking, alcohol and breastfeeding in mothers from the general population. *E. Stergiakouli[1,2], J. Martin[3], M. Hamshere[3], B. St Pourcain[4], N. Timpson[1], A. Thapar[3], G. Davey Smith[3].* 1) MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom; 2) School of Oral and Dental Sciences, University of Bristol, Bristol, United Kingdom; 3) Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK; 4) Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Introduction: Smoking and alcohol consumption during pregnancy have been suggested as possible risk factors for ADHD. Inferring causality has not been possible because the mother provides both the prenatal environment and genetic risk factors for ADHD. We investigated if there are shared genetic effects between ADHD and possible risk factors for ADHD using polygenic risk score analysis in mothers from the general population. Materials and Methods: ADHD polygenic risk scores were calculated for 8340 mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC) based on the results of a genome-wide ADHD case-control study (Stergiakouli et al. 2012). We tested polygenic scores for association with smoking status and alcohol consumption before pregnancy and during the first trimester and with breastfeeding status at 2 months postnatally. Results: Higher genetic risk for ADHD was associated with higher odds of smoking before and during the first trimester of pregnancy and not breastfeeding at 2 months after birth (see Table). However, there was no evidence of association with alcohol consumption. Adjusting for the ADHD polygenic score of the child did not change the association. Conclusions: Our results show for the first time that there are shared genetic effects between ADHD and life style choices in adults from the general population that do not reach diagnostic criteria for the disorder. The mother can not only transmit genetic risk for ADHD to her offspring but also expose the child to risk factors through her life style choices that are in turn influenced by her genetic risk for ADHD.

Outcome	N	OR (95% CIs)	p value
Smoking before pregnancy	7530	1.05 (1.01 to 1.1)	0.03
Smoking at first trimester	7543	1.08 (1.03 to 1.15)	0.002
Alcohol consumption before pregnancy	7530	0.99 (0.91 to 1.08)	0.9
Alcohol consumption at first trimester	7543	0.98 (0.94 to 1.03)	0.5
Breastfeeding	6604	1.06 (1.01 to 1.11)	0.03

426F

Dynamic trajectories of psychiatric hospitalizations in the Danish population for subjects with 22q11.2 deletions and duplications. *W.K. Thompson[1], L. Olsen[1], S. Weinsheimer[1], M. Bertalan[1], W. Mazin[1], J. Grauholm[2], M. Hollegard[2], D. Hougaard[2], C.B. Pedersen[3], P.B. Mortensen[3], T. Werge[1], T. Sparso[1], iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research.* 1) Institute of Biological Psychiatry, Roskilde, Region Sjæland, Denmark; 2) Section of Neonatal Genetics and the Danish Neonatal Screening Biobank, Copenhagen, Denmark; 3) National Centre for Register-based Research, Aarhus, Denmark.

Purpose: Genetic rearrangements in the 22q11.2 region are mediated by LCR. Frequency estimates of the 3Mb deletion has so far been based on hospital recruitments and commonly referred to as 1:4000 newborns. A unique source for population-based studies is provided by the iPSYCH initiative (www.ipsych.au.dk). Dried bloodspots from all newborns have been stored since 1982 and today DNA has been extracted from 76,109 individuals (N=25,703 are random samples). Here, we present the first accurate population-based prevalence of both 22q11.2 deletions and duplications. Using the National Health Registry, we provide population-based incidence rate ratio (IRR) estimates for neurodevelopmental disorders in individuals with the 22q11.2 rearrangement. We used incidence density sampling to calculate incidence rate ratio (IRR) estimates for the psychiatric diagnostic categories (ADHD, autism, schizophrenia, and mental retardation). We also perform dynamic survival analyses showing the pattern of diagnoses as a function of age and 22q11.2 rearrangement type. These estimates are weighted using appropriate case-cohort weights and are therefore interpretable as an unbiased estimated of the effect of 22q11.2 rearrangements on psychiatric hospitalizations in all of Denmark. Results: Among the 25,703 population-based samples we observed a frequency of 0.027% (1:3672) for 22q11.2 deletions and 0.0622% (1:1606) for the reciprocal duplications. In our samples of ADHD (N=16,715), autism spectrum (N=14,333), schizophrenia (N=2,623) and mental retardation (N=4,097), we found a significant increased IRR for individuals carrying the 22q11.2 deletion (ADHD: IRR=5.00 (p=0.0026), autism: IRR=6.00 (p=0.00309), mental retardation: IRR=14.29 (0<0.00001). Carriers of the reciprocal 22q11.2 duplication showed similar increased risk for neurodevelopmental disorders (ADHD: IRR=3.75 (p=0.00053), autism: IRR=2.86 (p=0.018), schizophrenia: IRR=4.29 (p=0.035), and mental retardation: IRR=4.10 (p=0.00398), however, not as profound as for deletion carriers. Conclusion: True population-based frequency of the 22q11.2 deletion is marginally higher than the widely accepted 1:4000. Both the 22q11.2 deletion/duplication confer risk of neurodevelopmental disorders at the population-level.

427W

Gene-based meta-analysis of genome wide association studies uncovers novel genes associated with alcohol use disorder. *H. Zhang, J. Jung, B. Grant.* NIH/NIAAA, Rockville, MD.

Alcohol use disorder (AUD) is widespread around the world. Research shows that genetic factors account for about half of the risk for alcoholism (heritability = 0.49). Previous genome-wide association studies (GWAS) typically focus on single variant analysis, which may not have enough power to detect the majority of truly susceptibility genes. We performed gene based meta-analysis using publicly available data genotyped at the Illumina 1M bead-chip from the Study of Addiction: Genetics and Environment (SAGE; dbGap study Accession: phs000092), which included 1,235 alcohol dependent cases and 1,433 unrelated, alcohol-exposed, nondependent controls in the European ancestry (EA) cohort and 662 alcohol dependent cases and 499 controls in the African American (AA) cohort. The VEGAS (VErsatile Gene-based Association Study) program was used to conduct gene-based association analysis in each cohort separately before gene-based meta-analysis was performed using Fisher's method. Among 38 genes that yielded $P < 10^{-3}$ in the EA discovery cohort, six novel genes (GATAD1, ING3, OR52R1, CTD-2270F17.1, ASB13 and PIWIL4) passed a marginally significant threshold ($P < 0.10$) in the AA replication cohort and generated significant outcomes (combined $P < 10^{-3}$) in meta-analysis. Notably, variants at the GATAD1 locus on 7q21 were associated with alcohol use disorder (discovery $P = 2.8 \times 10^{-4}$, replication $P = 0.022$, combined $P = 7.8 \times 10^{-5}$). Our gene-based meta-analysis of GWAS indicated several novel susceptibility genes are involved in alcohol use disorder. Our results also suggest that gene-based GWAS models may be powerful in identifying and prioritizing candidate loci for association studies.

428T

Evaluating polygenic risk for narcolepsy and essential hypersomnia, including SNP and CNV. *M. Yamasaki[1], T. Miyagawa[1,2], H. Toyoda[1], S.S. Khor[1], X. Liu[3], H. Kuwabara[4], Y. Kano[4], T. Shimada[5], T. Sugiyama[6], H. Nishida[7], N. Sugaya[8], M. Tochigi[9], T. Otowa[10], Y. Okazaki[11], H. Kaiya[12], Y. Kawamura[13], A. Miyashita[14], R. Kuwano[14], K. Kasai[10], H. Tanii[15], T. Sasaki[16], Y. Honda[17], M. Honda[2,17], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Sleep Control Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 3) Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, Kanagawa, Japan; 4) Department of Child Psychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 5) Division for Counseling and Support, The University of Tokyo, Tokyo, Japan; 6) Department of Child and Adolescent Psychiatry, Hamamatsu University School of Medicine, Shizuoka, Japan; 7) Asunaro Hospital for Child and Adolescent Psychiatiry, Mie, Japan; 8) Department of Epidemiology and Public Health, Yokohama City University Graduate school of Medicine, Kanagawa, Japan; 9) Department of Neuropsychiatry, Teikyo University Hospital, Tokyo, Japan; 10) Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 11) Tokyo Metropolitan Matsuzawa Hospital, Tokyo, Japan; 12) Panic Disorder Research Center, Warakukai Med. Corp., Tokyo, Japan; 13) Department of Psychiatry, Shonan Kamakura General Hospital, Kanagawa, Japan; 14) Department of Molecular Genetics, Bioresource Science Branch, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; 15) Department of Psychiatry, Institute of Medical Life Science, Graduate School of Medicine, Mie University, Japan; 16) Department of Physical and Health Education, Graduate School of Education, The University of Tokyo, Tokyo, Japan; 17) Japan Somnology Center, Neuropsychiatric Research Institute, Tokyo, Japan.

In humans, narcolepsy is a sleep disorder that is characterized by sleepiness, cataplexy, and rapid eye movement (REM) sleep abnormalities. Essential hypersomnia (EHS) is another type of sleep disorder that is characterized by excessive daytime sleepiness without cataplexy. A human leukocyte antigen (HLA) class II allele, *HLA-DQB1*06:02*, is a major genetic factor for narcolepsy. Almost all narcoleptic patients are carriers of this *HLA* allele, while 30% to 50% of EHS patients and 12% of all healthy individuals in Japan carry this allele. The pathogenesis of narcolepsy and EHS are thought to be partially shared. In order to evaluate the contribution of common SNP to narcolepsy onset and to assess the common genetic background of narcolepsy and EHS, we conducted a polygenic analysis that included 393 narcoleptic patients, 38 EHS patients with *HLA-DQB1*06:02*, 119 EHS patients without *HLA-DQB1*06:02,* and 1,582 healthy individuals. We also included 376 individuals with panic disorder and 213 individuals with autism to confirm whether the results were biased. Polygenic risks in narcolepsy were estimated to explain 58.1% ($P_{\text{HLA-DQB1*06:02}} = 2.30 \times 10^{-48}$, $P_{\text{whole genome without HLA-DQB1*06:02}} = 6.73 \times 10^{-2}$) including *HLA-DQB1*06:02* effects and 1.3% ($P_{\text{whole genome without HLA-DQB1*06:02}} = 2.43 \times 10^{-2}$) excluding *HLA-DQB1*06:02* effects. The results also indicated that small-effect SNPs contributed to the development of narcolepsy. Reported susceptibility SNPs for narcolepsy in the Japanese population, *CPT1B*, *TRA@*, and *P2RY11*, were found to explain 0.8% of narcolepsy onset ($P_{\text{whole genome without HLA-DQB1*06:02}} = 9.74 \times 10^{-2}$). EHS patients with *HLA-DQB1*06:02* were estimated to have higher shared genetic background to narcoleptic patients than EHS patients without *HLA-DQB1*06:02* even when the effects of *HLA-DQB1*06:02* were excluded (EHS with *HLA-DQB1*06:02*: 40.4%, $P_{\text{HLA-DQB1*06:02}} = 7.02 \times 10^{-14}$, $P_{\text{whole genome without HLA-DQB1*06:02}} = 1.34 \times 10^{-1}$, EHS without *HLA-DQB1*06:02*: 0.4%, $P_{\text{whole genome without HLA-DQB1*06:02}} = 3.06 \times 10^{-1}$). Meanwhile, the polygenic risks for narcolepsy could not explain the onset of panic disorder and autism, suggesting that our results were reasonable. In addition to SNP-based polygenic analysis, we evaluated polygenic risks including not only SNP but also CNV, so that we can get more integrated genetic background.

429F

Pleiotropic effects of CSF levels of Alzheimer's disease proteins. *O. Vsevolozhskaya[1], D. Zaykin[2], D. Fardo[1].* 1) Biostatistics, University of Kentucky, Lexington, KY; 2) Biostatistics & Computational Biology Branch, National Institute of Environmental Health Sciences, Cary, NC.

Cerebrospinal fluid (CSF) analytes harbor potential as diagnostic biomarkers for Alzheimer's Disease (AD). Quantitative measures of CSF proteins comprise a set of often highly correlated endophenotypes that have previously shown promise in genetic analyses (Cruchaga et al., 2013; Kauwe et al., 2014). Pleiotropic impact of genetic variations on this set may provide additional insights into AD pathology at its earliest stages. To determine which specific endophenotypes are pleiotropic, one can employ methods based on the reverse regression of genotype on phenotypes. Recently, we proposed a method based functional linear models (Vsevolozhskaya et al, 2016) that utilizes reverse regression and simultaneously evaluates all variants within a genetic region for an association with multiple correlated phenotypes. Here we apply our novel methodology to explore pleiotropic effects of CSF analtyes using Alzheimer's Disease Neuroimaging Initiative (ADNI) data.

430W

GWASs of ability to carry a musical tune and mathematical educational attainment. *R. Bell, D. Hinds, C. Tian.* 23andMe Inc., Mountain View, CA.

We describe the results of large-scale GWAS of two interesting cognitive traits in the 23andMe cohort — the self-reported ability to carry a musical tune and the highest level of mathematics classes completed. Our musical tune analysis included 130,673 cases and 85,081 controls of European ancestry and revealed twenty-five novel associations that exceed the genome-wide significance threshold. These included a noncoding variant rs6882046 upstream of MEF2C (p=1.3e-23), which is in the MEF2 family of proteins that play a role in myogenesis and muscle differentiation and also play a role in brain development, learning and memory. The musical tune trait is genetically correlated in the 23andMe cohort with other music-related phenotypes such as the ability to clap to a beat (LD score rg=.75), perfect pitch (LD score rg=.88), and the ability to sing back musical notes (LD score rg=.97). It also shares significant associations with several other music-related phenotypes. The math analysis included 210,135 research participants of European ancestry and revealed twenty-nine associations that exceeded the genome-wide significance level. Associations included a loss of function variant rs10555297 in GIGYF2 (p=2.5e-13), which is involved in tyrosine kinase regulation, and rs4500960 in SLC4A10 (p=1.5e-11), a sodium coupled bicarbonate transporter that regulates intracellular pH of neurons. The trait is genetically correlated in the 23andMe cohort with other academic achievement traits such as having completed college, years of education (both LD score rg=.80), and other phenotypes relating to math ability such as working in a technical occupation (LD score rg=.74). It also shares significant associations with many of those phenotypes. We found that the math class and carry a tune phenotypes are not significantly correlated (LD score rg=-.05, p=.08) but they do share several significant associations both with the same effect direction and opposite effect direction. This data suggests there may be both synergy in the underlying mechanisms in some cases versus a tradeoff in other cases. Overall the analysis suggested that the complex phenotypes included in our study are highly polygenic, share common factors with related phenotypes, but are nonetheless quite distinct.

431T

A genome-wide association study of mitochondrial DNA copy number in mothers and children from the Avon Longitudinal Study of Parents and Children (ALSPAC). *A.L. Guyatt[1,2], K. Burrows[1,2], P.A.I. Guthrie[2], T.R. Gaunt[1,2], S. Rodriguez[1,2].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom.

Background: The mitochondrial genome (MtDNA) is expressed at variable copy number between cell populations and individuals. Since rare mutations in nuclear genes lead to a reduction in MtDNA copy number (the 'mitochondrial depletion syndromes'), we hypothesised that smaller, population-level differences in MtDNA copy number might be attributable to common, nuclear DNA polymorphisms. **Methods:** We performed a genome-wide association study (GWAS) in the ALSPAC study. For the main analysis, 2805 mothers (mean age 28 years), 3712 children (mean age 8 years) and 2164 neonates were studied (these participants were unrelated). Since cell populations vary in their mitochondrial content, we undertook additional analyses in which we controlled for white cell proportions, as estimated previously from methylation data by the Houseman method. **Results:** rs57968500 (*PSMD3* intron) on chr17 was associated with a -0.16 (S.E. 0.027, p=3.98e-09) change in SD units of log MtDNA copy number, in mothers. In unrelated neonates, but not children, this association was also present (beta [SE]=-0.13 [0.031], p=3.49e-05, beta [SE]=-0.02 [0.024], p=0.525, respectively). In addition, to assess the effects of cell proportions on the rs57968500 chr17 association (a known GWAS white cell count locus), a larger population of mothers was used. In order to maximise power, it was necessary to include some who were related to the offspring: beta (unadjusted) [SE]=-0.12 [0.019], p=1.26e-09, N=5431; beta (cell-proportion adjusted) [SE]=-0.04 [0.060], p=0.516, N=642. In neonates, an association at rs10424198 (chr19, *ABHD8* intron) was observed (beta [SE]=-0.26 [0.034], p=3.84e-14). *ABHD8* is in head-to-head orientation with the mitochondrial ribosomal protein L34: a SNP in *MRPL34* rs11548548 [synonymous]) was in LD (r2=0.27, D'=0.89) with rs10424198, and neared genome-wide significance (beta [SE]=-0.25 [0.047], p=2.43e-07). However, the rs10424198 association was not replicated in mothers or children. **Conclusion:** We observed associations between MtDNA copy number and a white cell count locus. Although this association diminished after controlling for cell proportions, this analysis had limited power, and replication in larger cohorts is warranted. A chr19 locus was associated with MtDNA copy number in neonates: given its location in proximity to a mitochondrial ribosomal protein implicated in organelle biogenesis, it will be important to see whether this locus replicates in other neonatal populations.

432F

Mitochondrial haplogroup backgrounds modify the phenotypic impact of SNPs in genes relevant to mitochondrial function. *R.T. Levinson[1], D.C. Samuels[1,2].* 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN.

Mitochondrial haplogroups are patterns of genetic variation that are maternally inherited and are specific to continental ancestry. Mitochondrial haplogroups have been associated with neurological and immune function phenotypes, among others. Despite having their own genomes, mitochondria are still dependent on ~1000 proteins encoded in the nucleus for proper function. Using BioVU, a biobank linked to de-identified EHRs, we explored the relationship between nuclear and mitochondrial genetic variation in humans. We determined the mitochondrial haplogroup of 20,064 White individuals randomly split into discovery (n=13391) and replication (n= 6673) sets. Common nonsynonymous SNPs in genes on the MitoCarta2 list of genes relevant to mitochondria and present on the Illumina Human Exome BeadChip were obtained for these individuals. We performed a phenome-wide association study (PheWAS) to evaluate whether the presence of a SNP in individuals of a given mitochondrial haplogroup increased the likelihood of a medically relevant phenotype compared to individuals without the SNP or the haplogroup. ICD.9 codes aggregated into groups of similar phenotypes were used as our outcomes, and all regressions were adjusted for age at last record and gender. Several different nuclear SNPs in combination with the common mitochondrial haplogroup H increased the risk of cancer phenotypes. In contrast, on the background of haplogroup IWX, the SNPs tested were more likely to increase the risk for immune phenotypes. Overall, we noticed that different haplogroup backgrounds had different general areas of significant phenotypes. In conclusion, our study provides evidence that mitochondrial haplogroup backgrounds may impact the phenotypic effect of SNPs in genes relevant to mitochondrial function.

433W

Comparison of mitochondrial DNA copy number estimation techniques from multiple platforms. *R.J. Longchamps[1], R.N. Eggebeen[1], J.A. Lane[2], F.N. Ashar[1], T.M. Bartz[3], N. Pankratz[2], D.E. Arking[1].* 1) Mckusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, MN; 3) Department of Biostatistics, University of Washington, Seattle, WA.

Mitochondrial DNA copy number (mtDNA-CN) declines with age and is associated with frailty, cardiovascular disease, and overall mortality. The gold standard for determining mtDNA-CN is qPCR; however this method is cost prohibitive for large scale association studies. We have recently developed several methods (available in Genvisis) to overcome this issue by leveraging preexisting data in the forms of microarray raw probe intensities and next generation sequencing read counts. Here, we compare mtDNA estimated from qPCR, the Affymetrix 6.0, the Illumina Exome Chip, Whole Exome Sequencing (WES) data and Whole Genome Sequencing (WGS) data to highlight the best technique for accurately quantifying mtDNA-CN. White blood cell (WBC) count is known to be negatively associated with mtDNA-CN. SNPs associated with WBC count were used as proxies to evaluate how well each platform accurately measured mtDNA-CN. Data from 29,273 individuals (22.3% African American [AA]) across four independent cohorts (ARIC, MESA, CHS, and LLFS) were used to evaluate effect size (beta) concordance with known WBC count SNPs (n = 13) to determine which method maximized known association signals. Initial analysis for AA in MESA determined Affymetrix optimally maximized signal from rs2814778, the top Duffy locus SNP, a well-established WBC count locus in African Americans. The effect size estimate for Affymetrix was -0.29 standard deviation units (SE = 0.043) compared to -0.16 and -0.17 (SE = 0.044 and 0.044) for the Exome chip and qPCR metrics, respectively. A similar analysis within the ARIC cohort revealed WES to perform similarly to the Exome chip. In a further comparison of effect size estimates, a GWAS for WBC count was compared to a GWAS for mtDNA-CN within ARIC for all SNPs p < 0.0001. The highest correlation belonged to the Affymetrix array with a correlation coefficient of -0.579 compared to -0.228 for the Exome Chip, -0.262 for WES and -0.128 for WGS. Here we illustrate the ability to use WBC as a proxy for evaluating mtDNA-CN estimates derived from five separate platforms. With the ability to collect information from hundreds of mitochondrial probes, estimates from microarrays appeared to maximize signal relative to qPCR. Interestingly, our metric derived from sequencing read count data does not appear to outperform mtDNA-CN calculated from microarray data. However, our results show that despite the poor signal, mtDNA-CN can still be derived from the Exome Chip, WGS and WES data.

434T

A genome wide association study for exfoliation syndrome in a European-ancestry sample. *R.P. Igo[1], J.N. Cooke Bailey[1], J.H. Kang[2], P. Kraft[3], R. Allingham[4], M. Hauser[5], J. Fingert[6], R. Ritch[7], A. Sit[8], T. Aung[9], C.C. Khor[10], R. Lee[11], M.A. Pericak-Vance[12], W. Scott[12], L.R. Pasquale[13], J.L. Haines[1], J.L. Wiggs[13].* 1) Epidemiology & Biostatistics, Case Western Reserve Univ, Cleveland, OH; 2) Channing Division of Network Medicine, Brigham and Women's Hopital, Harvard Medical School, Boston, MA; 3) Program in Genetic Epidemiology and Statistical Genetics, T.H. Chan Harvard School of Public Health, Boston, MA; 4) Department of Ophthalmology, Duke University Medical Center, Durham, NC; 5) Department of Medicine, Duke University Medical Center, Durham, NC; 6) Department of Ophthalmology, University of Iowa, College of Medicine, Iowa City, IA; 7) Einhorn Clinical Research Center, Department of Ophthalmology, New York Eye and Ear Infirmary of Mt. Sinai, New York, NY; 8) Department of Ophthalmology, Mayo Clinic, Rochester, MN; 9) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore and Eye Academic Clinical Program, Duke-National University of Singapore Graduate Medical School, Singapore; 10) Division of Human Genetics, Genome Institute of Singapore, Singapore and Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 11) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 12) Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 13) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA.

Exfoliation syndrome (XFS), involving the deposition of proteinaceous fibers on ocular tissues, is a major cause of glaucoma, cataract and other eye diseases. The *LOXL1* gene is the only well-characterized risk locus for XFS with strong effect, and while *CACNA1A* has recently been implicated as a causal locus with small effect from a Japanese sample (Aung et al., 2015, *Nat Genet* **47,** 387), there still remain undiscovered genetic determinants of XFS. We conducted a genomewide association study (GWAS) on a sample of 1,124 XFS cases and 4,894 controls of European ancestry, using the Illumina OmniExpress panel (532,011 common variants with minor allele frequency ³ 0.01 after quality control). Association analysis for single common variants was carried out by means of logistic regression, adjusting for sex and six principal components for population structure, using an additive genetic model. Consistent with previous studies, we observed an exceedingly strong association signal at *LOXL1*, with numerous genomewide significant markers with odds ratios near 4.0 for the common alleles (most significant marker: rs2165241, OR = 4.11 for the A allele, 95% CI = (3.65, 4.62), $p = 4.0E-121$). In addition, suggestive association results ($1E-07 < p < 1E-06$) appeared on chromosome 3q (rs7629566, intronic to *CEP19* and *PIGX*) and on chromosome 6p (rs2763122, within pseudogene *HNRNPLP1*). We did not, however, observe strong association in or near previously reported XFS loci *CACNA1A* or *CLU* (best $p > 0.001$ within 100 kilobasepairs of gene). Moreover, our chromosome 3q and 6p findings were not strongly supported by results from three previous EXF GWAS on cohorts of European ancestry. Overall, these results support *LOXL1* as a major risk factor for XFS in this European Caucasian population, and suggest that there are no other common genetic risk factors in this sample with moderate to large effects. Grant support: NIH/NEI R01 EY020928.

435F

Common and rare variants of congenital cataract genes and age-related nuclear cataract: International Cataract Genetics Consortium. *E. Yonova[1], W. Zhao[2], R.P. Igo Jr[3], P. Sundaresan[4], A. Fletcher[5], C.C. Klaver[6], B.E. Klein[7], J.J. Wang[8], S.K. Iyengar[3], C.J. Hammond[1,9], C.Y. Cheng[2,10], International Cataract Genetics Consortium.* 1) Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 2) Singapore Eye Research Institute, Singapore National Eye Center, SG; 3) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, USA; 4) Department of Genetics, Aravind Medical Research Foundation, #1, Anna Nagar, Madurai¬ 625020, Tamilnadu, IN; 5) Department of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, UK; 6) Department of Epidemiology, Erasmus Medical Centre, Rotterdam, NL; 7) Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, USA; 8) Centre for Vision Research, Westmead Institute for Medical Research, University of Sydney, Sydney, AU; 9) Department of Ophthalmology, King's College London, London, UK; 10) Duke-NUS Medical School, National University of Singapore, SG.

   Purpose: A recent genome wide association study (GWAS) in Asian populations identified common variants within *CRYAA* to be associated with age-related nuclear cataract. Mutations in *CRYAA* are the most common cause of congenital cataracts. As part of the International Cataract Genetics Consortium (ICGC), we performed GWAS meta-analysis of 14,151 individuals from European and Asian ancestry and found variants in three additional congenital cataract genes to be associated ($p<1\times10^{-5}$) with age-related nuclear cataract, two at genome-wide significance (GJA3 and SOX2-OT/SOX2). The aim of this study was to explore whether common and rare genetic variants in other congenital cataract genes were also associated with age-related nuclear cataract.   **Methods:** 95 congenital cataract genes (longest transcript +/- 100kb), selected from 3 publicly available databases (OMIM, CatMap and ClinVar) and an in-house database, were tested for single-variant association with nuclear cataract using common variants from the ICGC GWAS meta-analysis. We also performed a burden test for rare variants (SKAT) in those genes using whole-genome sequencing data from 905 unrelated individuals from the TwinsUK cohort. Finally we performed a gene-set enrichment test (DAVID) to explore whether any of the associated congenital cataract genes belonged to the same pathways.   **Results:** Common variants in 47 congenital cataract genes (50%) were associated with age-related nuclear cataract ($p<0.05$). The most strongly associated variants belonged to the following genes: *BFSP1* ($p=3.5\times10^{-5}$), *LIM2* (1.4×10^{-4}), *MIP* (3.4×10^{-4}) and *CHMP4B* (3.8×10^{-4}). *BFSP1* encodes for a filament-like protein that is part of a structure (the beaded element) that is unique to the differentiated lens fibers. *LIM2* encodes for another lens specific structural protein found at lens fiber junctions. *MIP* encodes for an aquaporin that plays a role in cell communication while *CHMP4B* encodes for a part of a complex which functions in the sorting of endocytosed cell-surface receptors. Rare variant (MAF<0.05) burden test analysis showed only two genes to be associated with nuclear cataract: *POMT2* ($p=1.5\times10^{-3}$) and *CRYBA4* ($p=0.02$), although sample size was smaller, limiting power to find associations. Only one pathway – focal adhesion was enriched ($p=0.004$). **Conclusions**: These results suggest that common polymorphisms within genes where coding mutations cause congenital cataract may be important contributors to age-related nuclear cataract.

436W

Marital status, marriage patterns and fertility of deaf people in the Tyva Republic (South Siberia, Russia). *O.L. Posukh[1,2], M.S. Bady-Khoo[1,3], V.Yu. Mikhalskaia[1,2], M.V. Zytsar[1,2], N.A. Barashkov[4,5], G.P. Romanov[5], S.A. Lashin[1,2].* 1) Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation; 2) Novosibirsk State University, Novosibirsk, Russian Federation; 3) Perinatal Center of the Tuva Republic, Kyzyl, Russian Federation; 4) Laboratory of Molecular Genetics, Yakut Scientific Center of Complex Medical Problems, Yakutsk, Sakha Republic, Russian Federation; 5) Laboratory of Molecular Biology, Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Yakutsk, Sakha Republic, Russian Federation.

   High frequency of deafness caused by mutations in the *GJB2* (Cx26) gene might be the consequence of combined effects of high rate of assortative mating among deaf individuals and the relaxed selection against deafness (Nance et al., 2000). Our previous studies revealed that 18.9% of deafness in the Tyva Republic (South Siberia, Russia) were caused by *GJB2* mutations. This study aims to evaluate marital status, marriage patterns and fertility of deaf persons (n=456) at the age of 19 to 86 years (mean 39.6±0.7) living in the Tyva Republic. Most of them have congenital or early onset severe-profound deafness. Deaf individuals indicated sign language as a common communication mode with other deaf persons, whereas in communication with hearing people they use lip-reading, writing, and oral language skills, if any. 329 (72.1%) deaf individuals are married and the data on 200 marriages were available for analysis. We revealed assortative mating rate of 0.65 in total sample of married deaf people with significant differences in urban (0.71) and rural (0.58) groups (p=0.036). Fertility of married deaf individuals is slightly lower compared with their hearing siblings (2.22 vs. 2.40). We also found tendency to decreasing fertility in assortative marriages (2.14±0.08) compared with marriages in which only one of spouses was deaf (2.50±0.16) although the differences were not significant. This study provides insight to social aspects of deafness in the Tyva Republic and our data are relevant for predicting the prevalence of inherited deafness in studied region. Work is supported by the RFBR grant #15-04-04860_a.

437T

Identifying rare and low-frequency variants associated with axial length in the Beaver Dam Eye Study (BDES). *S.M. Bomotti[1], P. Duggal[1], F. Chen[1], B.E.K. Klein[2], K.E. Lee[2], B. Truitt[3], R. Klein[2], S.K. Iyengar[3], A.P. Klein[1,4,5].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 2) Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA; 4) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, USA; 5) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.

Refractive errors, myopia and hyperopia, are the most common causes of visual impairment in the US and the world. Although studies suggest a strong genetic component, only ~3.4% of the genetic variation in refractive errors has been explained to date. The primary determinant of refractive errors is ocular axial length. The goal of this analysis is to identify rare and low-frequency variants (minor allele frequency (MAF) ≤ 5%) associated with axial length. A subset of 1,908 individuals aged 58-94 years from the population-based Beaver Dam Eye Study (BDES) cohort were successfully genotyped using the Illumina exome array. Following standard quality control procedures, 10,397 autosomal genes and 34,981 single variants from 874 individuals were available for analysis. Axial length measurements were obtained from the fourth follow-up visit of the BDES using partial coherence laser interferometry (IOL Master, Carl Zeiss, Jena, Germany). We conducted a gene-based analysis using SKAT-O to increase our power to detect the effects of rare and low-frequency variants on axial length variation. A single variant analysis was then conducted in PLINK under an additive model. All analyses were adjusted for age, sex, height, and education, and were conducted in the right eye followed by the left eye to ensure consistency of results. The Bonferroni-corrected threshold for significance was $P < 4.81 \times 10^{-6}$ for the gene-based analysis and $P < 1.84 \times 10^{-6}$ for the single variant analysis. In the gene-based test, the *RAD17* checkpoint clamp loader component gene (MIM 603139) on chromosome 5q13.2 was significantly associated ($P = 3.44 \times 10^{-6}$) with axial length. The single variant test supported this finding with a suggestively associated missense variant (rs35440980, c.1778G>A [MAF = 0.76%], p.R593K) in *RAD17* in the right ($P = 7.48 \times 10^{-6}$) and left ($P = 1.65 \times 10^{-5}$) eyes. The adhesion G protein-coupled receptor G7 gene (*ADGRG7* [MIM 612307]) on chromosome 3q12.3 was also significantly associated with axial length ($P = 1.56 \times 10^{-6}$). A missense variant (rs61742836, c.187G>A [MAF = 0.68%], p.D63N) in *ADGRG7* was suggestively associated in the right eye ($P = 8.18 \times 10^{-5}$) and significantly associated ($P = 1.26 \times 10^{-6}$) with axial length in the left eye. These results suggest that rare and low-frequency variants within the *RAD17* and *ADGRG7* genes are associated with variation in axial length. These results may help to elucidate the biological mechanism underlying ocular refraction.

438F

Variation in *RNF149* associated with nuclear lens opacity. *S.J. Loomis[1], A.P. Klein[1,2,3], B.E.K. Klein[4], R. Klein[4], K.E. Lee[4], B. Truitt[5], S.K. Iyengar[5], P. Duggal[1].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland; 3) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland; 4) Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; 5) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio.

PURPOSE: Nuclear sclerotic cataract (NSC; MIM 601371) is the most common form of age-related cataract and is one of the leading causes of blindness worldwide. NSC is severe nuclear sclerosis, or opacity of the nucleus (center) of the optic lens. NSC is multifactorial, with evidence for both genetic and environmental risk factors, such as smoking. We sought to better understand the genetic factors associated with increasing nuclear lens opacity through interrogation of rare and low frequency coding variants using exome array data. **METHODS:** We analyzed genotype data from the Illumina Human Exome Array for 1489 participants in the Beaver Dam Eye Study for association with nuclear sclerosis grade. All individuals were of European ancestry and had not undergone cataract surgery. We conducted analysis using highest nuclear sclerosis grade in either eye prior to lens replacement surgery or loss to follow up after regressing off the effects of age and sex as outcome. We then performed single-variant regression analysis for 32,138 variants with MAF≥0.003. We used 11,844 genes with at least two variants with MAF<0.05 (N=63,598 SNPs) to perform a gene-based unified burden and non-burden sequence kernel association test (SKAT-O) using Madsen-Browning weights. We analyzed the data overall and stratified by smoking status (never vs. past or current smokers). **RESULTS:** None of the single variant tests were statistically significant after controlling for multiple tests (p<1.6 x 10^{-6}; top SNP: rs144458991, p=2.83 x 10^{-5}). Gene-based tests showed suggestive evidence of association for the gene *RNF149* (p=8.25 x 10^{-6}) in the overall analysis. This finding reached genome-wide significance (p<4.2 x 10^{-6}) when the analysis was limited to never smokers (N=790, p=2.67 x 10^{-6}), but not when limited to past or current smokers (N=698, p=0.07). *RNF149* codes the ring finger 149 protein, involved in ligase and protein ubiquitination activity. **CONCLUSIONS:** Our study suggests exonic changes in the *RNF149* gene may affect risk of nuclear lens opacity, Further research is necessary to fully understand the genetic architecture of nuclear sclerosis. .

439W

Whole genome sequencing of 4,806 age-related macular degeneration cases and controls. *A. Kwong[1], L.G. Fritsche[1,2], X. Zhan[3], J. Bragg-Gresham[4], K.E. Branham[5], M. Othman[5], L. Gieser[6], R. Ratnapriya[6], D. Stambolian[7], E.Y. Chew[8], A. Swaroop[6], G. Abecasis[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway; 3) outhwestern Medical Center, University of Texas, Dallas, TX, United States; 4) Kidney Epidemiology and Cost Center, Department of Internal Medicine-Nephrology, University of Michigan, Ann Arbor, MI, United States; 5) Department of Ophthalmology and Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, United States; 6) Neurobiology-Neurodegeneration and Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, MD, United States; 7) Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania Medical School, Philadelphia, PA, United States; 8) Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD, United States.

Age-related macular degeneration (AMD) is a leading cause of visual impairment among people age 50 and older. Previous AMD genetics studies have focused on subsets of the genome accessible via genotyping arrays or targeted sequencing. While this approach is affordable and accurate, it limits the number and types of variants that can be discovered. We therefore conducted whole genome sequencing (WGS) to assess the contribution of the full spectrum of genetic variants to AMD pathogenesis.We sequenced, at 6x coverage, 4,806 samples (2,413 cases and 2,393 controls, age- and sex-matched) of European ancestry, from the University of Michigan Kellogg Eye Center, the National Eye Institute, the University of Pennsylvania, and the Michigan Biobank. Our cases consist of subjects with large drusen (intermediate AMD), geographic atrophy (GA), choroidal neovascularization (CNV), and mixed GA/CNV. Our WGS analysis discovered over 45 million variants (42 million SNPs and 3 million indels), a majority of which are very rare (>35 million with allele frequency <1%) and previously unreported (>26 million variants not in dbSNP version 138).We rediscovered significant genomewide association signals (P-value <5x10^{-8}) at four loci (*CFH*, *C2/CFB/SKIV2L*, *ARMS2/HTRA1*, and *C3*). In a follow-up association study restricted to advanced AMD (excluding participants withintermediate AMD), we discovered 6 new genomewide significant signals near the known risk loci: an intergenic deletion near *ST3GAL6-AS1* and an intronic deletion in *lnc-GTPBP8-1* on chr 3, an intergenic deletion near *MARCKS* on chr 6, an intronic deletion in *CAMSAP1* on chr 9, an intergenic deletion near *FLJ26245* on chr 16, and an intergenic deletion near *VPREB1* on chr 22. We were also able to better characterize rare variants: for example, the *CFH* R1210C variant was detected in 23 cases vs. 2 controls (P-value = 3.0x10^{-7}).In addition to the GWAS, our study contributes a large catalog of rare variants from a large number of AMD cases, which can be used for AMD-specific reference panel to impute into much larger genotyped panels, potentially increasing both the number of AMD-associated imputed variants and the imputation accuracy of those variants.In summary, our WGS and association study of AMD replicate 4 previously-discovered loci and identified significant signals outside of known loci, providing a more complete list of potentially causal variants genomewide.

440T

Linkage analysis of common myopia using whole exome genotyping in highly aggregated Ashkenazi families finds novel signals and replicates known loci. *C.L. Simpson[1,2], A.M. Musolf[2], L. Portas[2], F. Murgia[2], Q. Li[2], J.E. Bailey-Wilson[2], D. Stambolian[3].* 1) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, United States; 3) Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States.

Myopia is a common refractive error which affects at least a third of most populations. Both genetic and environmental factors influence myopic development. It has a significant impact on the lives of affected individuals and carries high economic costs associated with treatment, loss of productivity and co-morbidity from vision impairment. Here we report the results of linkage analyses on highly aggregated Ashkenazi Jewish families with common myopia using an exome-targeted genotyping array. Sixty-four extended Ashkenazi Jewish families previously collected from New Jersey. Myopia defined as mean spherical equivalent (MSE) of -1D or worse and both parametric two-point linkage analyses (using MCLink) and collapsed haplotype pattern linkage using SEQlinkage were performed. Significant evidence of linkage was seen on chromosome 16q22.1 (max two-point LOD=7.28), chromosome 1p36.2 (max two-point LOD=4.47) which had previously been linked to ocular refraction but not myopia in these families (MYP14) and chromosome 8q24.22 (max two-point LOD=3.75). Suggestive evidence of linkage was seen on multiple chromosomes including the 7p14 close to the MYP17 locus and 11p13-p15.4 which overlaps the MYP18 locus and includes a 3' UTR variant in PAX6. Collapsed haplotype pattern linkage analysis of these data identified four genomewide significant loci, including the 16q22.1 locus (max HLOD 7.78, α=0.8) centered on the PDPR gene and 1p36.1 (max HLOD 3.47, α=0.57) centered on LINC00339. There was one signal unique to this analysis from the SSPO gene on chromosome 7q36.1 (max HLOD 4.02, α=1), which does not overlap with the known chromosome 7 myopia locus MYP17. Here we report novel loci and replicated a locus previously reported in refractive error but not myopia in these families. Targeted sequencing in these regions to will be necessary to further find causal variants under these linkage peaks.

441F

Modeling misclassified cases in electronic medical records using genotype information. *X. Zhao[1], E. Schmidt[1], L. Fritsche[2], G. Zajac[1], C. Brummett[3], S. Kheterpal[3], G. Abecasis[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

Combining the clinical data in electronic medical records (EMR) and genetic data provides us a chance to accelerate the pace of genomic discovery on thousands of traits. One major challenge in using EMR data is the potential misclassification of cases and controls. Previous studies have shown that misclassification of disease can cause biased effect size estimates and misleading association results. Existing methods for correcting misclassification require either knowing the misclassification rate or identifying a gold standard set of sample. We focus on the setting where neither is available. Here, we propose a method that estimates the misclassification rate by examing genotype information of dozens of disease associated loci. Instead of using gold standard samples, we use the odds ratios of known risk loci from previous large-scale genome-wide association studies (GWAS). Assuming that these odds ratios in our study are similar to those in previous GWAS, we are able to construct the likelihood and obtain the MLE of the misclassification rate. Then we can correct biased association statistics using estimated misclassification rate. In a simulation study, we generated phenotypes and genotypes for 5000 samples at 52 variants. The 52 variants were previously reported to be associated with age-related macular degeneration (AMD) by the International AMD Genomics Consortium. The simulation demonstrated that this method accurately estimated misclassification rates for a variety of different misclassification rates (range: 0.1-0.5) and disease prevalence (range: 0.05-0.5). For example, in the setting with disease prevalence 0.05 and misclassification rate 0.3, the mean square error of our estimated misclassification rate was 0.002. We also show that estimates of the misclassification rate are robust to errors in odds ratios. We applied our technique to the data from the Michigan Genomics Initiative (MGI), a collaborative research at the University of Michigan aiming to analyze the patient EMR jointly with genetic data. Using the odds ratios for 52 associated variants of AMD, we found that 36% of cases were likely to be misclassified when using the ICD-9 based classification scheme implemented in the PheWAS R package. Our method enables the evaluation of new EMR-based case definition schemes and provides accurate estimates of disease odds ratios and other association measures when cases are misclassified.

442W

Socioeconomic status, genetic risk, and how their interactions affect risk of oral clefts. *M.M. Jenkins[1], L.M. Almli[2], C.P. Bay[2], P.A. Romitti[3], L.M. Uribe[3], G.L. Wehby[3].* 1) Centers for Disease Control and Prevention, Atlanta, GA; 2) Carter Consulting, Inc., Atlanta, GA; 3) University of Iowa, Iowa City, IA.

Low socioeconomic status (SES) has been linked to increased risk of oral clefts. However, little is known about how SES interacts with genetic risks for oral clefts. Using data from a population-based, case-control study (National Birth Defects Prevention Study [NBDPS]), we examined whether the association between SES and risk of cleft lip with/without cleft palate (CL/P) was modified by genetic predisposition to CL/P. Our sample included 614 cases with isolated CL/P and 2213 controls without any major birth defect; birth years were between 1997 and 2009. SES was estimated by maternal education (≤high school [HS] vs >HS [reference]) or household income (1st and 2nd tertiles vs 3rd tertile). Genetic risk was estimated by a genetic risk score (GRS) based on a set of top genes and loci associated with CL/P in previous genome-wide association studies; SNPs of interest were assayed using a Fluidigm platform (San Francisco, CA). Logistic regression analyses were used to estimate CL/P risk for cases relative to controls, adjusting for maternal age at conception, NBDPS site, and maternal race/ethnicity. The risk for CL/P was greater for mothers with ≤HS education compared to mothers with >HS education (OR=1.27, 95%CI 1.03-1.58) and for mothers in the lowest income tertile compared to those in the highest tertile (OR=1.35, 95%CI 1.01-1.80). As expected, the GRS was significantly associated with increased risk for CL/P (OR=2.49, 95%CI 2.17-2.85). A significant SES x GRS interaction was observed. Specifically, the risk of CL/P associated with ≤HS education was higher at 1 standard deviation (SD) below the GRS mean (i.e., a low genetic risk profile; OR=2.00, 95%CI 1.43-2.81) than at the mean GRS (OR=1.40, 95%CI 1.11-1.76). Educational status was not associated with CL/P risk at 1 SD above the GRS mean (i.e., a high genetic risk profile; OR=0.98, 95%CI 0.75-1.27). Similarly, comparing the lowest tertile of household income to the highest, the association with CL/P was stronger at 1 SD below the GRS mean (OR=2.13, 95%CI 1.36-3.33) than at the mean GRS (OR=1.47, 95%CI 1.07-2.00). There was no association between household income and risk of CL/P at 1 SD above the GRS mean (OR=1.01, 95%CI 0.71-1.43). Our preliminary results suggest that for individuals with a higher genetic risk profile for CL/P, the influence of lower SES on risk of CL/P is attenuated. Similarly, SES seems to be a more influential risk factor for CL/P among individuals with lower genetic risk scores.

443T

Genetic variations on chromosome 21 and AVSD risk in Down syndrome: Results from chromosome 21 genome sequencing. *X. Blanc[1], F.A. Santoni[1], C. Borel[1], P. Makrythanasis[1], R. Sailani[1], M. Guipponi[1], F. Bezrukov[1], E. Falconnet[1], P. Ribaux[1], U. Radhakrishna[1], F. Bajolle[2], D. Laux[2], J. Okoronkwo[2], M.A. Alcántara-Ortigoza[3], S. Vicari[4], K. Rabionet[5,6,7], L.A. Pérez-Jurado[6,8], R. Touraine[9], S. Kitsiou[10], H. Fryssira[10], G. Merla[11], J.M. Delabar[12], D. Bonnet[2], S.E. Antonarakis[1,13].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneva 1211, Switzerland; 2) M3C-Necker, Pediatric Cardiology, University Paris Descartes, 75015 Paris, France; 3) Department of Human Genetics, National Institute of Pediatrics, Mexico City, Mexico; 4) Department of Neuroscience, Children's Hospital Bambino Gesù, 00165 Rome, Italy; 5) Centre for Genomic Regulation, Biomedical Research Park (PRBB), E-08003 Barcelona, Spain; 6) Hospital del Mar Research Institute (CRG), E-08003 Barcelona, Spain; 7) CIBER in Epidemiology and Public Health (CIBERESP), E-08003 Barcelona, Spain; 8) Department of Experimental and Health Sciences, Universitat Pompeu Fabra, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), E-08002 Barcelona, Spain; 9) CHU de Saint-Etienne, Hôpital Nord, 42055 Saint-Etienne Cedex 2, France; 10) Department of Medical Genetics, University of Athens, 11527 Athens, Greece; 11) Medical Genetics Service, Hospital, IRCCS, 71013 San Giovanni Rotondo, Italy; 12) Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS EAC 4413, Université Paris Diderot, Sorbonne, Paris Cité, 75205 Paris, France; 13) iGE3 Institute of Genetics and Genomics of Geneva, Geneva 1211, Switzerland.

Down syndrome patients are often born with congenital heart defects (CHD). Among those, atrioventricular septal defect (AVSD) is the most frequent and one of the most severe. The incidence of AVSD is 43% in CHD T21 patients (17% in the trisomic population) and almost absent in the general population. The genetic architecture of CHD regarding chromosome 21 variants has been previously studied with SNP arrays and oligonucleotide hybridization. With this approach our group identified two SNPs and three CNVs regions significantly associated with the risk of CHD (Sailani et al., Genome Research, 2013). Here we use targeted chromosome 21 ultradeep sequencing (>120x) using a custom DNA selection design to perform a case-control association study in a trisomic population to identify SNVs, structural variants and CNVs associated with an increased risk of AVSD. The study cohort includes 162 Down syndrome patients with AVSD at birth and 178 Down syndrome patients without detectable cardiac defect as controls. A total of 179931 SNVs has been called through our in-house pipeline implementing BWA, GATK and Annovar. The association study of SNVs and structural variants has been performed with Fisher's exact test and revealed rs1156253 in the MIR99AHG cluster and rs869725 in a regulatory area nearby PDE9A, a gene known to be expressed in heart tissues and upregulated in case of hypertrophy and cardiac failures ($p<10^{-6}$). We performed CNVs detection with CoNIFER and CONTRA. We were able to call in average more than 20 CNVs per sample; the association analysis is ongoing. This study shows that common variants on chromosome 21 contribute to the developmental phenotype of AVSD and might provide insights to the understanding of the biological and genetic mechanisms underlying the development of the disease.

444F

Whole exome association of rare deletions in multiplex oral cleft families. *J. Fu[1], T. Beaty[2], A.F. Scott[3], M. Parker[4], J. Wilson[5], M. Marazita[6], E. Mangold[7], H. Albacha-Hejazi[8], J. Murray[9], A. Bureau[10], J. Carey[2], S. Cristiano[1], I. Ruczinski[1], R. Scharpf[11].* 1) Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 2) Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 3) Center for Inherited Disease Research and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 4) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 5) Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 6) Department of Oral Biology, Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, PA; 7) Institute of Human Genetics, University of Bonn, Bonn, Germany; 8) Dr. Hejazi Clinic, Damascus, Syrian Arab Republic; 9) Department of Pediatrics, School of Medicine, University of Iowa, IA; 10) Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec et Département de Médecine Sociale et Préventive, Université Laval, Québec, Canada; 11) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

By sequencing the exomes of distantly related individuals in multiplex families, rare mutational and structural changes to coding DNA can be characterized and their relationship to disease risk assessed. Recently, several rare single nucleotide variants (SNVs) were associated with an increased risk of non-syndromic oral cleft, highlighting the importance of rare sequence variants in oral clefts and illustrating the strength of family-based study designs. However, the extent to which rare deletions in coding regions of the genome occur and contribute to risk of non-syndromic oral clefts is not well understood. To identify putative structural variants underlying oral cleft risk, we developed a pipeline for the delineation of rare hemizygous deletions in families from whole exome sequencing and statistical inference based on rare variant sharing. Our method utilizes a read-depth approach, with median centering across samples, and includes adjustments for sequencing artifacts like GC-content and mappability. We further employ the 1000 Genomes database as well as a Bayesian mixture model to filter for the rarity of identified deletions. Among 46 multiplex families, we identified 51 regions with one or more rare hemizygous deletions. We found 43 of the 51 regions contained rare deletions occurring in only one family member. Members of the same family shared a rare deletion in only 8 regions, which included a rare deletion on chr13 at 53,078,416 - 53,158,768bp showing nominal statistical significance (p = 0.004). We also devised a scalable global test for enrichment of rare variant sharing. .

445W

Associated anomalies in cases with anotia and microtia. *C. Stoll, Y. Alembik, B. Dott, M.P. Roth.* genetics , Faculté de Médecine, Strasbourg, France, France.

Infants with anotia and microtia (AM) often have other non-AM associated congenital anomalies. The purpose of this investigation was to assess the prevalence and the types of these associated anomalies in a defined population. The associated anomalies in infants with AM were collected in all livebirths, stillbirths and terminations of pregnancy during 29 years in 387,067 consecutive births in the area covered by our population-based registry of congenital malformations. Of the 146 cases with AM registered during this period, representing a prevalence of 3.77 per 10,000, 49.3% had associated anomalies. There were 14 (9.6%) cases with chromosomal abnormalities including 5 trisomies 18, and 18 (12.3%) nonchromosomal recognized dysmorphic conditions including 6 cases with oculo-auriculo-vertebral spectrum. However, numerous other recognized dysmorphic conditions were registered. Forty (27.4%) of the cases had multiple congenital anomalies (MCA). Anomalies especially in the cardiovascular, the musculoskeletal, the urogenital, the central nervous, and the digestive systems, and facial clefts were the most common other anomalies. This study included special strengths: each affected child was examined by a geneticist, all elective terminations were ascertained, and the surveillance for anomalies was continued until 2 years of age. In conclusion the overall prevalence of associated anomalies, which was one in two cases, emphasizes the need for a thorough investigation of cases with AM. A routine screening for other anomalies may be considered in infants and in fetuses with AM.

446T

Genetic determinants of telomere length in a Bangladeshi cohort. *D. Delgado[1], B. Pierce[1,2,3], C. Zhang[1], J. Farina[1], M. Kibriya[1], H. Ahsan[1,2,3,4].* 1) Department of Public Health Sciences, University of Chicago , Chicago , IL. 60637; 2) Department of Human Genetics, University of Chicago, Chicago, IL 60637; 3) Comprehensive Cancer Center, University of Chicago, Chicago, IL 60637; 4) Department of Medicine and Human Genetics, University of Chicago, Chicago, IL 60637.

Introduction: Attrition of telomere length (TL) occurs with each round of cell division and is associated with aging and increased risk for various age-related chronic diseases. TL has high heritability and shows differences by race/ethnicity. We performed a genome-wide association study (GWAS) to identify genetic determinants of TL in a Bangladeshi cohort. Methods: We measured TL for 5,019 individuals participating in one of two prospective cohort studies in Bangladesh (2.416 men and 2,603 females with a mean age of 39.9 years). TL was measured using one of two technologies: a qPCR-based assay (n=2,203) and a Luminex–based method (n=2.816). All subjects were genotyping using the Illumina HumanCytoSNP array. Prior to analysis, we used mixed effects models to adjust our TL measures for plate, position, and batch effects, and we used the standardized residuals from these regressions as our TL phenotypes for GWAS. We used the Genome-wide Complex Trait Analysis (GCTA) software to identify single nucleotide polymorphisms (SNPs) associated with TL, using a mixed effects model to account for the presence of cryptic relatedness among our participants. Results: After standard quality control procedures, a total of >250,000 genotyped and 8.5M imputed SNPs were included in the analysis. We replicated loci reported in prior GWAS studies of TL in European populations, including the *TERC* region (P = 1.7x10-9), the *TERT* region (P = 4.1x10-7) and *RTEL1* region (P = 9.3x10-7). We identified a suggestive novel association with TL in the *NCR2* region (P = 7.4x10-7). The SNP –based heritability estimate based on all genome-wide SNP was 26.7% (P=4.7x10-10). Conclusion: Our data suggest that the *TERC* region harbors the strongest common genetic determinants of TL in this Bangladeshi cohort. Our results also suggest that several of TL-associated regions identified in GWAS of European ancestry show consistent association with TL among Bangladeshi participants. In addition, we also observe suggestive evidence for a novel association signal for TL in the *NCR2* region.

447F

Simultaneous scan of whole-genome genetic variants in family-based association studies via supervised dimension reduction. *D. Zhang, L. Wang, Q. Wang, M. Zhang.* Depart of Statistics, Purdue Univ, West Lafayette, IN.

Population stratification, family structure and cryptic relatedness are known to produce spurious association if not properly corrected. Recent Study showed that mixed model is an effective way to handle these confounding factors. Most of the current mixed model based methods are typically carried out by testing one or few loci at a time, making these algorithms essentially limited in identifying genetic variants of weak effects. We propose an algorithm that extends penalized orthogonal component regression (POCRE) to family-based association studies (fPOCRE), with the merit of accounting for multiple loci at the same time. Our proposed algorithm is a hybrid of linear mixed model that is well known for handling confounding issue and penalized regression that performs well in large p small n setting. We compared fPOCRE with existing methods, and demonstrated that fPOCRE has promising performance over others in terms of both power and false discover rate. We have applied our method to a genome-wide association study in the Framingham Heart Study.

448W

On predicting tissue specific functional effects of genetic variation with applications to complex diseases. *D. Backenroth[1], B. Xu[1], K. Krzysztof[1], E. Khurana[2], J. Buxbaum[3], I. Ionita-Laza[1].* 1) Columbia University, New York, USA; 2) Weill Medical College, Cornell University, New York, USA; 3) Mount Sinai School of Medicine, New York, USA.

Over the past few years, large scale genomics projects such as the ENCODE and Roadmap Epigenomics have produced genome-wide data on a large number of biochemical assays for a diverse set of human cell types and tissues. Such data can play a critical role in identifying putatively causal variants among the abundant natural variation that occurs at a locus of interest. We will present unsupervised approaches to integrate these diverse sets of annotations for specific tissues and cell types into a single predictor of functional importance. We have calculated such integrative functional scores for every possible position in the human genome for 127 tissues and cell types available in Roadmap Epigenomics. We demonstrate the usefulness of such scores in the context of complex disease genetics, using GTEx data on over 44 tissues, and summary statistics from over 20 GWAS. Specifically we show that tissue specific functional variants are enriched among eQTLs in related tissues in Roadmap. Furthermore, we show how these integrated functional scores can be used to identify the most likely tissue for a complex disease using summary statistics from GWAS, and can shed light on pathways implicated in a disease using existing enhancer-gene target maps.

449T

Modeling prediction error improves power of transcriptome-wide association studies. *K. Bhutani[1], A. Sarkar[2], Y. Park[2], M. Kellis[2], N. Schork[1].* 1) Bioinformatics & Systems Biology, UCSD, La Jolla, CA; 2) CSAIL, MIT, Cambridge, MA.

Thousands of loci associated with hundreds of complex diseases have been reported in the NHGRI catalog of genome-wide association studies (GWASs), but most genome-wide significant loci are devoid of protein-coding alterations and likely instead affect transcriptional regulation. Several studies have directly investigated the role of transcriptional regulation on complex diseases by jointly considering genotype, expression, and phenotype. However, such studies require all data to be measured in all samples, which is still prohibitive at the scale of GWAS. Recent large-scale efforts such as the Gene-Tissue Expression Project (GTEx) have produced reference profiles of transcription, enabling transcriptome-wide association studies (TWASs). TWAS tests for association between gene expression and phenotype by predicting gene expression in GWAS cohorts (where expression is not measured) using models of transcriptional regulation trained on reference transcriptomes. However, current methods for TWAS only use point estimates of imputed expression and ignore uncertainty in the prediction. Here, we develop a method to explicitly model error in imputed expression and propagate this error through TWAS. We demonstrate that imputed expression has high uncertainty, possibly due to genetic factors not included in typical models, environmental factors which vary between reference transcriptome cohorts, and technical biases in training the predictive models. We show through simulation and application to real data that our method improves power to detect genes associated with phenotype.

450F

Exploring the Latino asthma paradox: Significant interactions between genetic ancestry, psycho-social stressors, and environmental exposures on asthma susceptibility in Puerto Rican and Mexican children. *M.G. Contreras[1], M.J. White[2], D. Hu[2], S. Huntsman[2], C. Eng[2], N. Thakur[2], S. Thyne[2], A. Davis[3], K. Meade[3], H.J. Farber[4], P.C. Avila[5], R. Kumar[6], E. Brigino-Buenaventura[6], M.A. LeNoir[7], D. Serebrisky[8], W. Rodriguez-Cintron[9], J.R. Rodriguez-Santana[10], L.N. Borrell[11], E. Burchard[2].* 1) San Francisco State University 1600 Holloway Ave, San Francisco, CA 94132; 2) University of California, San Francisco Departments of Medicine and Bioengineering and Therapeutic Sciences. 1550 4th St. Rm 584D San Francisco, CA 94158; 3) Children's Hospital and Research Center Oakland, Oakland, CA; 4) Department of Pediatrics, Section of Pulmonology, Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 5) Children's Memorial Hospital; Feinberg School of Medicine, Northwestern University, Chicago, IL; 6) Department of Allergy and Immunology, Kaiser Permanente–Vallejo Medical Center, Vallejo, California; 7) Bay Area Pediatrics, Oakland. CA; 8) Pediatric Pulmonary Division, Jacobi Medical Center, Bronx, NY; 9) Veterans Caribbean Health Care System, San Juan, Puerto Rico; 10) Centro de Neumologia Pediatrica, San Juan, Puerto Rico; 11) Department of Health Sciences, Graduate Program in Public Health, Lehman College, City University of New York, Bronx, NY.

Asthma, an inflammatory condition of the airways, is the most common chronic disease among children. In the U.S. Puerto Ricans have the highest asthma prevalence and mortality rates compared to all other racial/ethnic groups; particularly as compared to Mexican Americans who have the lowest asthma prevalence, this phenomenon is known as the Latino Asthma Paradox. Asthma is a complex disease likely affected by genetic and environmental interactions. We used a non-parametric method, VISEN, to identify genetic x environmental interactions, including genetic ancestry, associated with asthma susceptibility in Puerto Rican and Mexican children with and without asthma. Our study population included 1520 participants (Mexican: 525 cases, controls 224; Puerto Rican 651 cases, controls 120) from the Genes, Environment, and Admixture in Latino Americans (GALA II) study. Our environmental variables included evidence of discrimination, sex, BMI, age, genetic ancestry, and prenatal smoking. We identified several interactions associated with asthma in Puerto Rican and Mexican American children. One of the strongest interactions in Mexicans included BMI * Discrimination * Age; interestingly, in Puerto Rican children we discovered a model that included BMI * Discrimination * Native American Ancestry. Our results indicate that interactions play a significant role in asthma susceptibility in these two Latino populations; and the nature of those interactions may be population specific. These results emphasize the need to recognize differences between populations that self-define as Latino.

451W

Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. *A. Cortes[1,2], D. Vukcevic[3,4,5], A. Motyer[3,4], L. Jostins[1], C. Dendrou[2], L. Fugger[2], S. Leslie[3,4,5], G. McVean[1].* 1) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 2) Nuffield Department of Clinical Neurosciences Weatherall Institute of Molecular Medicine John Radcliffe Hospital University of Oxford, Oxford, UK, OX3 9DU; 3) School of Mathematics and Statistics, University of Melbourne, Parkville, Vic 3010, Australia; 4) School of BioSciences, University of Melbourne, Parkville, Vic 3010, Australia; 5) Data Science, Murdoch Childrens Research Institute, Parkville, Vic 3052, Australia.

The past decade has seen an accelerated rate of discovery of the genetic determinants of complex traits. These discoveries have shed light on the molecular pathways involved in the aetiology of common diseases and have led to new therapeutics. Moreover, there is a growing interest in using genetic associations for assessing the efficacy and safety of therapeutic candidates, and this could be enabled with a better understanding of the spectrum of clinically phenotypes associated to genetic variants. So far, this has been limited by insufficient phenotyping in GWAS cohorts and improper methodological approaches. Cross-sectional cohorts, such as the UK Biobank, provide a resource that aims to overcome this limitation. The UK Biobank has recruited half a million individuals, derived genome-wide genotype data, and collected phenotypic data through questionnaires and linking of national health registries. Such resources require novel analytical strategies to jointly analyse genetic variation in individuals with a wide range of phenotypes collected. These methods must be robust to complexities in the data, with expected correlated effect sizes, abundant missing data and varied sample sizes across different phenotypes. Here we propose a Bayesian analysis framework that uses the hierarchical structure of diagnosis classifications to jointly analyse genetic variants against phenotypes in the UK Biobank Hospitalisation Episode Statistics and the self-reported diagnosis datasets. We validate our method using simulations and the analysis of allelic variation in the MHC. Simulations demonstrated an increase in statistical power when jointly analysing all phenotypes compared to individually analysing each clinical phenotype. The gain in power was quantified as the area under the ROC curve at different allele frequencies. For example, for a genetic variant with population allele frequencies of 1% an improvement from 0.757 to 0.902 was observed. Analysis of MHC allelic variation showed we could identify known associations to common immune-mediated diseases with routine health-care data. Increased evidence in the tree analysis was also notable, with support for HLA-B*27 association to ankylosing spondylitis at log10(BF)=29.96, whilst the evidence with the single phenotype analysis was log10(BF)=20.34. Furthermore, we observe allele associations with secondary outcomes common in disease prognosis, such as HLA-B*27 predisposing to glaucoma, a common complication in AS.

452T

Meta-imputation: A flexible approach to combine genotype data imputed against multiple reference panels. *S. Das, G. Abecasis.* Department of Biostatistics, University of Michigan, Ann Arbor.

Genotype imputation has proven to be an invaluable tool for human genetic studies - increasing power of gene mapping analyses, facilitating combination of results across studies (meta-analyses), and accelerating fine-mapping efforts. Imputation involves an in-silico method in which a reference panel is used to infer genetic variants that were not genotyped in the study sample. In the last few years, reference panels have proliferated - starting with the Hap-Map Consortium and the 1000 Genomes Project, and now including multiple disease specific sequencing efforts. Combining these reference panels for imputation leads to increased imputation accuracy, especially for rare variants. However, constructing a combined (joint) reference panel from these datasets can be quite difficult, and may often be impossible due to data-sharing restrictions. This is especially true when some panels, such as the Haplotype Reference Consortium, are not publicly available but can only be accessed through web imputation services. In this study, we explore the idea of meta-imputation, an approach that combines results from imputation against different reference panels to obtain a consensus imputed genotype. We construct this imputed genotype as a weighted average of individual imputed dosages. The weighing scheme is based on new individual-specific imputation diagnostic metrics, estimated by minimac3, reflecting the combined effects of chip density, probability of template switches, historical recombination events around the site being meta-imputed, and the genetic distance between study samples and the reference panel. We show that meta-imputation performs well compared to more cumbersome joint imputation in terms of imputation accuracy (measured by Pearson R^2). The prime benefit of our our tool (MetaMinimac) is to allow users to combine their imputed data from multiple reference panels in-silico without requiring direct access to individual haplotypes from all potential reference panels. An added bonus is that this method allows imputation analyses to be run in parallel and thus reduces net compute time. In summary, we describe a method that makes it easier for genetic studies to take advantage of the many currently available reference panels and improve imputation accuracy in genetic studies while abiding by data sharing restrictions.

453F

Statistical tests for Hardy-Weinberg equilibrium at X-chromosomal genetic markers. *J. Graffelman[1], B.S. Weir[2].* 1) Department of Statistics and Operations Research, Polytechnic University of Catalonia, Avinguda Diagonal 647, 08028 Barcelona Barcelona, Spain; 2) Department of Biostatistics, University of Washington, Box 359461, Seattle, WA 98195-9461, USA.

In large-scale genotyping studies genetic markers are usually tested for Hardy-Weinberg equilibrium (HWE). Disequilibrium is often related to genotyping error, and tests for HWE form a useful device for detecting genotyping error. It has become computationally feasible to test millions of autosomal genetic markers for HWE using an exact test. For markers on the X chromosome, the same autosomal exact test is typically applied, but using only the females in the database. Hitherto, the hemizygous males were considered to be uninformative with regard to HWE. It is well known that an autosomal marker will achieve HWE in one generation of random mating. However, for a marker on the X chromosome the situation is different. If there is a difference in initial allele frequency between males and females, then it can take several generations before equilibrium is reached. Equilibrium refers to the stability of allele and genotype frequencies over the generations, and this condition will only be attained if male and female allele frequencies are equal. In a recent paper in Heredity, we have therefore argued that males are informative, because a difference in allele frequency between males and females may indicate that equilibrium does not hold. For the X-chromosome, it seems more sensible to jointly test Hardy-Weinberg proportions in females and equality of allele frequencies in males and females in a single omnibus test. We have developed such omnibus tests, specifically tailored to biallelic markers on the X-chromosome, using a classical frequentist approach. We have designed X-chromosomal versions of the chi-square test, the exact test, the likelihood ratio test and the permutation test, and updated our R software (the HardyWeinberg package) implementing these tests. Statistical inference with a test including males can differ considerably from a test that uses females only. In this contribution we address ongoing extensions of our work on HWE for X-chromosomal markers with multiple alleles and present our latest methodological advances. We apply our newly developed methodology to complete X-chromosomes with data from different populations of the 1000 genomes project. We create tracks for HWE test results, and suggest these to be included in genome browsers. We show examples of such tracks where we uncover areas of Hardy-Weinberg disequilibrium on the X-chromosome.

454W

Improved imputation accuracy of rare and low-frequency variants using population-specific imputation reference panel. *R. Magi[1], M. Kals[1,2], K. Parn[1], M. Mitt[1,3], T. Esko[1,4], P. Palta[1,5].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia; 3) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 4) Broad Institute, Cambridge, MA, US; 5) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

Genetic imputation has shown to be a cost efficient way to improve the power and resolution of GWAS studies. Current large reference panels work well for common variants, but the imputation quality of rare and low-frequency (RLF) variants is rather limited. For better imputation of RLF variants, two approaches could be employed: (i) increase the size of the reference panel; or (ii) use a population-specific reference panel. In the current study we evaluate and compare imputation quality and accuracy among several different panels into individuals from the Estonian population: (a) 1000 Genomes (HRC); (b) Haplotype Reference Consortium (HRC); (c) Estonian-specific whole-genome sequence; and (d) merged panels. As a population-based reference, we used the Estonian Genome Center, University of Tartu (EGCUT) high-coverage (30x) whole-genome sequenced dataset – 2,244 geographically diverse samples from Estonia, selected randomly by county of birth. All reference panels were tested with imputation into 6,400 Estonian samples genotyped on Illumina HumanCoreExome chip. These samples were pre-phased (SHA-PEIT2 software) and imputed (IMPUTE2 software) into each reference panel and two combinations of the EGCUT and 1000G panels (EGCUT+1000G and 1000G+EGCUT). As 500 of these individuals also had whole-exome sequencing (WES) data available, these overlapping samples were used to determine the imputation accuracy. The EGCUT reference panel has less variants than the multi-ethnic 1000G and HRC reference panels, but more than 75% of those are confidently imputed (IMPUTE2 INFO>0.8). The average imputation quality is fairly similar across all reference panels for common variants (MAF>5%), with INFO ranging between 0.93-0.96 and is the most discrepant for the rare variants (MAF<0.5%), where the average imputation quality differs the most between EGCUT (average INFO 0.75) and 1000G panels (average INFO 0.40). When comparing imputed data with available WES data from the same individuals, we were able to test the imputation quality of all references. We observed that the imputed genotypes matched the real genotypes from WES data better in the whole frequency spectrum when using the Estonian population-specific panel alone or merged with 1000G reference. In conclusion, we have shown that population specific imputation reference panels do improve the imputation quality across entire allele frequency spectrum with the greatest improvement in rare variants.

455T

Canonical correlation analysis separates genetic and non-genetic confounders. *R. Schweiger[1], E. Rahmani[1], L. Shenhav[1], E. Halperin[1,2,3].* 1) Computer Science, Tel Aviv University, Tel Aviv, Israel; 2) Molecular Microbiology and Biotechnology Department, Tel Aviv University, Tel Aviv, Israel; 3) International Computer Science Institute, 1947 Center Street, Berkeley, CA 94704, USA.

Estimation of statistical confounders is a necessity in genetic and epigenetic studies in order to control for false discoveries in association testing. However, different confouders, both genetic and non-genetic, may combine in the same dataset and could be difficult to estimate separately. For example, Genome-wide association studies (GWAS) routinely apply principal component analysis (PCA) to infer population structure within a sample to correct for confounding due to ancestry. However, the top PCs do not always reflect population structure: they may reflect family relatedness, long-range LD (for example, due to inversion polymorphisms), or completely non-genetic confounders, such as assay artifacts. Another example is epigenome-wide association studies (EWAS), where different methylation profiles of distinct cell types may lead to false discoveries. Dimensionality reduction techniques, such as PCA and variants thereof, are often used for the correction of cell type heterogeneity, since the first several principal components are correlated with cell type composition. However, EWASs reported thousands of CpG methylation sites to be associated with genetics and ancestry, and applying PCA on a group of genome-wide CpGs correlated with genetics can be used in order to efficiently capture population structure. We show that applying cross decomposition techniques, such as canonical correlation analysis (CCA), on datasets of different measurements of the same individuals separates genetic and non-genetic confounders. These techniques allow for more accurate estimation of various types of confounders, which can then be used as covariates in statistical analyses or studied separately.

456F

A two-stage hidden Markov model design for biomarker detection, with application to microbiome research. *Y. Zhou[1], X. Wang[2].* 1) North Carolina State University, Raleigh, NC; 2) IMEDACS, LLC.

High-throughput genomic and transcriptomic sequencing technologies promise to yield great advances in biomedicine, but the enormous amount of data produced raises statistical challenges. Typical data in a microbiome study consist of the operational taxonomic unit (OTU) counts that have the characteristic of excess zeros. There are few methods which can control false discovery rate as well as increase the power to detect significant associations. For many high-throughput technologies, the cost still limits their application. Two-stage designs are attractive, in which a set of interesting features or biomarkers is identified in a first stage, and then followed up in a second stage. However, the statistical principles of two-stage FDR control with hidden Markov model (HMM)s have not been established. In this paper, we study an efficient HMM-FDR based two-stage design, using a simple integrated analysis procedure across the stages. Numeric studies show its excellent performance when compared to available methods. A power analysis method is also proposed. We use examples from microbiome data to illustrate the methods.

457W

Large scale brain eQTL meta-analysis from multiple RNA-sequencing cohorts. *S.K. Sieberts[1], M. Allen[2], K. Dang[1], J. Calley[3], P.J. Ebert[3], T. Perumal[1], L.M. Mangravite[1], the AMP-AD Consortium eQTL Working Group and the CommonMind Consortium (CMC).* 1) Sage Bionetworks, Seattle, WA; 2) Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL; 3) Lilly Research Labs, Eli Lilly and Company, Indianapolis, IN.

To date, hundreds of risk loci have been associated with neurodevelopmental, neuropsychiatric and neurodegenerative diseases. When risks are mediated through gene expression, eQTL can implicate the specific genes involved through colocalizing algorithms or global tests of case-control differences in the predicted gene expression. In each case, eQTL from the relevant tissue are necessary to understand the specific expression patterns contributing to disease, which may not be captured in readily available tissue such as blood. Recently, several large initiatives, including the CommonMind Consortium (CMC), the Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD) and GTEx, have made high-quality data available from the RNA-sequencing and genotyping of post-mortem brain collections representing neurodegenerative and neuropsychiatric cases, and undiseased individuals. Here we generate the best-powered brain eQTL resource to date which spans multiple brain regions and diseases. In total, we have collected 6 cohorts, comprised of tissue from approximately 2400 individuals and representing 4 brain regions, the largest collection of which is from dorsolateral prefrontal cortex comprised of 1800 samples. A common analysis pipeline was applied to each cohort. Samples were imputed to the Haplotype Reference Consortium panel. For each of the AMP-AD cohorts, RNA-seq data was realigned and quantitated using parameters appropriate to specifics in library prep strandedness and paired-end sequencing. Following quality control, quantified expression from each of the 6 cohorts was then normalized via *voom*, adjusting for available known clinical and technical covariates, and hidden confounders, prior to applying a linear model to detect eQTL, adjusting for inferred genetic structure and diagnosis. Meta-analysis across cohorts was then performed. Preliminary analysis of the Caucasian samples from the CMC MSSM-Penn-Pitt cohort (n=467) imputed to 1000 Genomes v1 identified >2.1 million proximal (distance ≤ 1 Mb) eQTL with MAF ≥ 0.05 and FDR ≤ 5%, which is an order of magnitude more than reported in previous expression or exon array cohorts, or in the publically available GTEx PFC (v6) eQTL. Despite this, the replication of GTEx eQTL in CMC is extremely high ($\pi_1 = 0.98$) confirming these experiments are capturing the same information, despite substantially different processing and analysis pipelines, and are appropriate for meta-analysis.

458T

Interactive effect between ATPase-related genes and early-life tobacco smoke exposure on bronchial hyper-responsiveness detected in asthma-ascertained families. *M-H. Dizier[1,2], P. Margaritte-Jeannin[1,2], L. Pain[3], C. Sarnowski[1,2], M. Brossard[1,2], H. Mohamdi[1,2], N. Lavielle[1,2], J. Just[4], M. Lathrop[5], E. Bouzigon[1,2#], C. Laprise[3#], F. Demenais[1,2*], R. Nadif[6,7*], #* contributed equally.* 1) Inserm U946, Paris, France; 2) Univ Paris Diderot, Paris, France; 3) Université du Québec à Chicoutimi, Québec, Canada; 4) Centre de l'Asthme et des Allergies – Inserm, UMR-S 1136, Equipe EPAR, Paris, France; 5) Mc Gill University, Montréal, Canada; 6) Inserm, UMR-S 1168, Villejuif, France; 7) Univ Versailles Saint-Quentin en Yvelines, France.

In a previous positional cloning study of the 17p11 region, we identified genetic variants interacting with exposure to tobacco smoke (ETS) in early life for bronchial hyper-responsiveness (BHR) in the French Epidemiological study on the Genetics and Environment of Asthma (EGEA) (Dizier *et al.*, 2007). These variants were located in Dynein, Axonemal, Heavy Chain 9 gene, *DNAH9* (Dizier *et al.*, 2016), a promising candidate gene, which has a key role in motile cilia function and is associated to primary ciliary dyskinesia, a disease associated with bronchiectasis and abnormalities of pulmonary function. The aim of the present study was to identify other genetic variants interacting with ETS for BHR by investigating genes involved in the "ATPase binding" and the "ATPase activity" pathways. These two pathways were selected because they both include *DNAH9*, are known to be target of cigarette smoke and are involved in the movement of respiratory cilia. Analyses were first conducted in a discovery sample of 388 asthmatic EGEA families, using family-based association test (FBAT). For all SNPs located within 277 genes belonging to the two pathways, we applied FBAT homogeneity test between exposed *vs.* unexposed siblings to detect SNP x ETS interactive effects on BHR. Replication was performed in Saguenay-Lac-Saint-Jean (SLSJ) asthma collection (Laprise 2014), another asthma-ascertained sample of 253 French-Canadian families. In EGEA families, 25 SNPs showed interaction signals ($P \leq 5 \times 10^{-3}$) with ETS among BHR siblings. One SNP on 4p13 region reached the threshold ($P = 2 \times 10^{-5}$) for a significant interaction with ETS when correcting for multiple testing. This result did not quite reach the nominal level for replication in SLSJ families ($P = 0.13$) but there was improvement of evidence for interaction in the meta-analysis of the two samples ($P = 10^{-5}$). Another SNP on 9q31 showed a stronger signal of interaction in the SLSJ study ($P = 0.003$), and a suggestive interaction by meta-analysis ($P = 6 \times 10^{-5}$). Further replication studies are now necessary to confirm these results. This study highlights that incorporating biological knowledge in the selection of genes may identify novel genes potentially interacting with ETS in BHR siblings.

459F

A novel statistical method for genetic pleiotropic analysis of multiple phenotypes. *N. Lin[1], Y. Zhu[2], R. Fan[3], J. Zhao[2], M. Xiong[1]*. 1) University of Texas School of Public Health, Houston, TX; 2) Tulane University, New Orleans, LA; 3) NIH, Washington DC.

   Next-generation sequencing technologies and modern biosensing produce dozens of millions of SNPs and large number of phenotypes. A key issue for learning the intricate structures of genotype-phenotype is how to effectively extract a few informative internal representation and features from extremely high dimensional genotype and phenotype data. Deep learning uses statistical tools to extract information and reduce the size of "Big Data". Both genome and phenotype signals are compositional hierarchies, in which low level features are combined into higher-level features. Deep hierarchical representations of the genome and phenotype data transform the internal representation at one level into a representation at a higher and more abstract level, extract better informative features and capture invariant features that are only sensitive to the phenotype variability. Motivated by deep learning, we develop a novel quadratically regularized functional canonical analysis (QFCCA) for genetic pleiotropic analysis and apply it to UK-10K dataset with 46 traits and 2,240,049 SNPs in 33,746 genes typed in 765 individuals. The number of genes that were significantly associated with 46 traits after Bonferroni corrections(P-value less than 1.5E-6) identified using QFCCA, FCCA, GAMuT, functional MANOVA and USAT were 80, 50, 0, 50, 0, respectively. The 46 phenotypic variables can be classified as 13 sub- groups according to their biological functions. Then by applying the QFCCA statistics to these sub-group phenotype separately, we can identify 8 genes that are significantly associated with 3 sub-group phenotypes and 43 genes were significantly associated with 2 sub-group phenotypes. A total number of identified genes that were significantly associated with at least one of 46 traits using QFCCA, FCCA, functional MANOVA, USAT and GAMuT were 299, 262, 262, 0 and 43, respectively. The results show that the QFCCA substantially outperforms all existing statistics for genetic pleiotropic analysis.

460W

A comparison of methods for genome-wide association mapping in the PAGE study. *R. Tao[1], J. Haessler[2], D.Y. Lin[1], T. Thornton[3], S. Buyske[4,8], Y. Patel[5], C.R. Gignoux[6], E. Stahl[7], T. Matise[8], G.L. Wojcik[6], K.E. North[9], S. Bien[2], C.D. Bustamante[10], L. Hindorf[11], R.J.F. Loos[7], E.E. Kenny[7], C. Carlson[2], C. Kooperberg[2], D.O. Stram[5]*. 1) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 2) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Department of Statistics, Rutgers University, Piscataway, NJ; 5) Keck School of Medicine, University of Southern California, Los Angeles, CA; 6) Department of Genetics, Stanford University, Palo Alto, CA; 7) Department of Genetics and Genome Sciences, Ichan School of Medicine at Mount Sinai, New York, NY; 8) Department of Genetics, Rutgers University, Piscataway, NJ; 9) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 10) Stanford Center for Computational, Human, and Evolutionary Genomics, Palo Alto, CA; 11) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD.

   Purpose: Recently 53,424 participants in the diverse PAGE study were genotyped for 1.7 million SNPs on the Illumina MEGA array. Genome-wide association studies of the PAGE data require statistical methods to control for hidden population stratification, admixture, and also for a high degree of both known and unknown relatedness. Failure to adjust for population structure and relatedness will lead to high false positive rates in genetic association testing. Two adjustment methods one based on linear mixed models (LMM) and another on generalized estimating equations (GEE) were compared in order to provide guidance to investigators concerning the relative performance of these methods in analysis of the large number of phenotypes available for the consortium. **Methods:** The LMM used is the Genesis program implemented in Bioconductor package for R, while the GEE method considered is the SUGEN program of Lin et al. Markers of frequency of 0.001 or greater in the PAGE genotype data were used in the association tests. All association analyses included principal components and (when appropriate) age and sex, as adjustment variables. **Results:** After examining nearly twenty different phenotypes with up to 49,000 individuals in a given analysis we found relatively few instances where the output from SUGEN and Genesis differed markedly between one another. Overall the genomic control parameters were slightly smaller when using the Genesis program than the SUGEN program, indicating somewhat better control for relatedness using the former. However p-values, effect sizes, and test statistics were generally very highly correlated between the two methods with little reordering of the top associations. The greatest differences in pooled versus stratified analyses were seen for phenotypes showing heterogeneous variances by strata (i.e. by racial/ethnic group). In general the Genesis program was much slower than the SUGEN program for the pooled analyses, for stratified analyses, where there were smaller numbers of subjects per analysis, the run times were much more similar. **Conclusions:** Compared to most GWAS datasets the PAGE samples genotyped on the MEGA-array exhibit an unusual degree of structure especially relatedness. The LMM-based Genesis program theoretically allows for a much more complex relatedness structure than does the GEE-based SUGEN program. However, in these data the results of the two approaches are quite compatible.

461T

Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Z. Zhu[1], F. Zhang[1], H. Hu[2], A. Bakshi[1], M. Robinson[1], J. Powell[1,3], G. Montgomery[4], M. Goddard[5,6], N. Wray[1], P. Visscher[1,7], J. Yang[1,7].* 1) Queensland Brian Institute, University of Queensland, Brisbane, QLD, Australia; 2) State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, China; 3) Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia; 4) Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 5) Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia; 6) Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia; 7) University of Queensland Diamantina Institute, Translation Research Institute, Brisbane, Queensland, Australia.

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human complex traits. However, the genes or functional DNA elements through which these variants exert their effects on the traits are often unknown. We propose a method (called SMR) that integrates summary-level data from GWAS with data from expression quantitative trait locus (eQTL) studies to identify genes whose expression levels are associated with a complex trait because of pleiotropy. We apply the method to five human complex traits using GWAS data on up to 339,224 individuals and eQTL data on 5,311 individuals, and we prioritize 126 genes (for example, *TRAF1* and *ANKRD55* for rheumatoid arthritis and *SNX19* and *NMRAL1* for schizophrenia), of which 25 genes are new candidates; 77 genes are not the nearest annotated gene to the top associated GWAS SNP. These genes provide important leads to design future functional studies to understand the mechanism whereby DNA variation leads to complex trait variation.

462F

Improving power for rare variant tests by integrating external controls. *S. Lee[1,2], S. Kim[1], C. Fuchsberger[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Due to the drop in sequencing cost, the number of sequenced genomes is increasing rapidly. To improve power of rare variant tests, these sequenced samples could be used as external control samples in addition to control samples from the study itself. However, when using external controls, possible batch effects due to the use of different sequencing platforms or genotype calling pipelines can dramatically increase type I error rates. To address this, we propose novel summary statistics-based single and gene- or region-based rare-variant tests that allow the integration of external controls while controlling for type I error rates. Our approach is based on the insight that batch effects on a given variant can be assessed by comparing allele frequencies between internal and external control samples. Specifically, to mitigate possible batch effects for the single variant test, we propose a shrinkage estimator of a log odds ratio; we then extend the shrinkage estimator to gene- or region-based tests by aggregating single variant log odds ratios. Through extensive simulation experiments, we show that the proposed method can control type I error rates at the exome-wide significant level (alpha=2.5x10^{-6}), and have substantially increased power over an approach using the internal controls only. By using European ESP samples as external controls in the analysis of age-related macular degeneration (AMD) targeted sequencing data, the proposed method can replicate rare-variant associations between AMD and *CFH* (*p*-values=8.27×10^{-6}) and *C3* (*p*-values=1.57×10^{-5}) genes. These genes are not significant when the internal control samples are exclusively used (*p*-values > 0.01). The proposed method is implemented in iECAT R-package.

463W

Adjustment for confounders in random forests analysis. *Y. Zhao, D.F. You, F. Chen, H. Yu.* Department of Biostatistics, Nanjing Medical University, Nanjing, China.

   AIM: Random forest (RF) has been increasingly applied in the analysis of high dimensional genetic data. However, if the confounding effects has not been appropriately removed, RF may produce biased results.**METHODS:** We propose to adjust the confounding effects by a residual-based procedure. Suppose that there are 1 response (Y), p confounders and k predictors. We firstly fit a RF model by using Y as the response and the p confounders as the predictors. We then generated the adjusted response as the difference between Y and the prediction of the RF model. Similar procedures were performed to generate the adjusted predictors. The adjusted response and predictors are then used as the outcome and covariates in the RF analysis.**RESULTS:** Extensive simulations were performed to evaluate the proposed procedure. The proposed procedure has advanced performance when there are one or two confounders. The proposed adjustment can also be used when there are interactions between confounders. We also applied the proposed method in a real GWAS dataset.**CONCLUSIONS:** The proposed adjustment procedure can be used in the RF analysis on high dimensional data when there are potential confounders.

464T

Are population-specific panels of exomes useful to identify disease variants: Insights from the French Exome Project. *E. Genin[1], C. Dina[2], T. Ludwig[1], O. Quenez[3], S. Letort[1], P. Lindenbaum[2], J. Giemza[2], B. Grenier-Boley[4], C. Charbonnier[3], C. Bellenguez[4], D. Bacq[5], D. Lechner[5], A. Boland[5], V. Meyer[5], S. Chatel[2], K. Rouault[1], G. Nicolas[3], C. Férec[1], H. Le Marec[2], L. Letenneur[6], J-F. Dartigues[6], J-C. Lambert[4], D. Campion[3], R. Redon[2], J-F. Deleuze[5].* 1) U1078 Genetics, functional Genomics, Biotechnology, INSERM, CHRU, UBO Brest, France; 2) Inserm UMR-1087 / CNRS UMR 6291, l'institut du thorax, Univ. Nantes, Nantes, France; 3) Inserm UMR-1079, Faculté de Médecine, Univ. Rouen, Rouen, France; 4) Inserm UMR-1167, Institut Pasteur, Lille, France; 5) Centre National de Génotypage, CEA, Evry, France; 6) Inserm UMR-1219, Univ. Bordeaux, Bordeaux, France.

   Novel high throughput sequencing technologies enable the characterization of all the genetic variations in the coding sequence (exome) of individuals and the discovery of novel variants involved in monogenic and complex diseases. The study of these two kinds of diseases require different methods and raises specific problems however, in both situations, the major challenge is to understand the functional role of the identified variants to filter out the neutral ones. For this purpose, investigators often rely on the observed frequencies of variants in public databases. In this work, we show that these public databases are not well representatives of the different ancestries in Europe and this is true in particular for France. Indeed, comparing the exomes of 573 individuals sampled in 6 different regions of France to the frequency data available in ExAC, we identified several variants with significant allele frequency differences. We also found that compared to the different public panels, the "French exomes" allow a more efficient filtering of variants leading to a substantial reduction in the number of candidate variants retained for validation. These allele frequency differences are also detected within France between the different geographic regions, pointing out, for example, to some loss-of-function variants that could have been under positive selection. Moreover, at this fine geographical scale, we were able to detect differences in the rare variant burden in several genes that could have consequences for the setting-up of rare variant association studies. Population-specific exome panels such as the one we have developed through the French Exome (FREX) Project are clearly useful to avoid false positives findings due to population structure and to allow a better understanding of genetic diversity at fine geographic scales.

465F

Functional regression method for whole genome epistatic eQTL analysis with sequencing data. *K. Xu[1], L. Jin[1], M. Xiong[2,1].* 1) School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, Shanghai, China; 2) Department of biostatistics, University of Texas School of Public Health, Houston, TX.

Epistasis plays an essential rule in understanding the regulation mechanisms and is an essential component of the genetic architecture of the gene expressions. However, interaction analysis of gene expressions remains fundamentally unexplored due to great computational challenges and data availability. Sparse multivariate regression models (MLMs) are widely used methods for expression epistasis analysis. The current MLMs were designed for microarray gene expression measurements and common variants. They consider a SNP at a time and take one number as a summarized gene expression level. Due to variation in splicing, transcription start sites, polyadenylation sites, post-transcriptional RNA editing across the entire gene, and transcription rates of the cells, RNA-seq measurements generate large expression variability and collectively create the observed position level read count curves. A single number for measuring gene expression which is widely used for microarray measured gene expression analysis is highly unlikely to sufficiently account for large expression variation across the gene. Simultaneously analyzing epistatic architecture using the whole genome RNA-seq and next-generation sequencing (NGS) data poses enormous challenges for methodologies and computations. To meet these challenges, we proposed a functional regression model (FRM) for epistasis analysis of gene expression, which takes a gene as a basic unit and view the RNA-Seq of each gene as a continuous curve. By large-scale simulations, we demonstrate that the proposed method can achieve the correct type 1 error and has higher power to detect the interactions between genes than the existing methods. The proposed methods are applied to the RNA-Seq and next-generation sequencing data from the 1000 Genome Project. The numbers of pairs of significantly interacting genes after Bonferroni correction identified using FRM, RPKM and DESeq were 16,2361, 260 and 51, respectively, from the 350 European samples. The QQ plots for FRM had excellent behavior. In summary, the FRM is a good choice for the whole genome epistatic eQTL analysis. .

466W

X-inclusion: Integrating X chromosome in whole genome association studies of variance heterogeneity. *W.Q. Deng[1], S. Lei[1,2].* 1) Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

Genetic variants associated with heterogeneity of quantitative trait variance are considered promising candidates to follow up. Not surprisingly, this approach to discover potentially interacting SNPs has so far been restricted to autosomes as X-chromosomes are often excluded from genome-wide analyses. In this short report, we qualitatively examined three grouping strategies to test variance heterogeneity of X-chromosome SNPs using Levene's test. We showed theoretically, in the context of variance heterogeneity testing, the three-genotype grouping assuming either the presence or absence of X-inactivation could potentially cause spurious variance signals as a result of confounders. For example, an inherent confounder for X-chromosomal SNPs is sex. Since sex necessarily correlates with the the number of minor alleles of X-chromosomal SNPs, and the distribution of certain quantitative traits due to physiological differences between female and male. Simulation studies and empirical analyses using data from Multi-Ethnic Study of Atherosclerosis are provided to show that 1) the extent the confounding depends on the sex-stratified distributions of the quantitative trait, 2) the five-group genotype/allele combined scheme is the most robust strategy for X-chromosome analysis of phenotypic variability, 3) the pattern of variance heterogeneity signals from X-chromosome SNPs also supports the complex genetic architecture of polygenic traits.

467T

Effect of bias and misclassification on gene-environment studies conducted in observational cohort settings: A simulation study. *A.A. Seyerle[1], C.M. Sitlani[2], K.E. North[1], C.R. Lee[3], E.A. Whitsel[1,4], T. Stürmer[1,5], C.L. Avery[1].* 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 2) Department of Medicine, University of Washington, Seattle, WA; 3) Division of Pharmacotherapy and Experimental Therapeutics, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC; 4) Department of Medicine, University of North Carolina, Chapel Hill, NC; 5) Center for Pharmacotherapy, University of North Carolina, Chapel Hill, NC.

Gene-environment studies represent an increasingly popular avenue to characterize complex traits. Yet few investigations have evaluated the degree to which gene-environment interaction estimates are influenced by sources of bias that affect estimates of the association between the environmental exposure and the outcome. Using pharmacogenomics as our gene-environment model, we therefore simulated three designs (cross-sectional, repeat cross-sectional, incident exposure), two referent groups (whole cohort, alternate exposure), and two scenarios (extreme or modest environmental effects) of N=120,000 participants using 1 million iterations to enable comparisons of 12 studies of the electrocardiographically measured QT interval (QT). For each study, we simulated a causal SNP with minor allele frequency of 25% and an index exposure frequency of 17%, using an $\alpha=5\times10^{-8}$. When selection bias was introduced to the extreme environmental effect scenarios, modest bias in the environment-SNP interaction was observed for all designs and referent groups with the exception of the two incident exposure designs (relative bias range: 2-18%); statistical power also was modestly decreased for these studies (e.g. relative power decrease of 1-37% to detect a simulated 2 ms environment-SNP interaction). However, minimal bias in the environment-SNP interaction effect was observed when selection bias was introduced to the modest effect scenario. When plausible degrees of exposure misclassification (sensitivity=97%, specificity=79%) were introduced, substantial bias towards the null in the environment-SNP interactions was observed across all 12 study settings (relative bias range: 25-51%); statistical power also was substantially decreased (e.g. relative power decrease of 50-99% to detect a simulated 2 ms environment-SNP interaction). In the presence of exposure misclassification, the detection of a simulated 2 ms effect required a repeat cross-sectional design with at least 150,000 participants. Exposure misclassification poses a sizable threat to gene-environment studies, greatly reducing estimated effect sizes, statistical power, and therefore perceived clinical and public health impact. Efforts to reduce the effects of exposure misclassification in gene-environment studies are therefore warranted.

468F

Heritability and ancestry influence the genomic architecture of height in South African KhoeSan populations. *C. Gignoux[1], M. Lin[2], J.M. Granka[1], A.R. Martin[1], J. Myrick[2], E. Atkinson[2], C.J. Werely[3], D. Gurdasani[4], C. Pomilla[4], T. Carstensen[4], B. Scelza[5], M. Moller[3], M. Sandhu[4], C.D. Bustamante[1], E.G. Hoal[3], M.W. Feldman[1], B.M. Henn[2].* 1) Stanford University, Stanford, CA; 2) Stony Brook University, Stony Brook, CA; 3) SA MRC Centre for Tuberculosis Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; 4) Wellcome Trust Sanger Institute, Cambridge, UK; 5) Department of Anthropology, University of California, Los Angeles, CA.

Height is a standard quantitative trait for genetic associations and has been extensively studied in European populations, however like most complex traits the architecture is less well known in other populations. However even with study sizes in the hundreds of thousands, GWAS hits explain only a limited amount of phenotypic variance relative to the high heritability of the trait. Very little is known about the genetics of height elsewhere, particularly in indigenous populations. To understand shared and unique components of height in a different population we collected DNA from over 400 individuals in extended families from two KhoeSan communities from the Northern Cape in South Africa, the ≠Khomani San and Nama. These groups continue to practice elements of traditional hunter-gatherer or pastoralist subsistence and therefore have the potential for different environmental influences. Pedigree-based heritability estimates of height in the KhoeSan is quite high (h2=0.95), although average height is lower than in other neighboring populations. Estimating heritability via relatedness matrices from SNP array and exome data are similar, likely due to inclusion of close relatives. After imputation we used linear mixed models to control both for population stratification and relatedness in our GWAS. We identified several signals associated with height variation in *VWA8* and *KSR2*, previously associated with BMI and energy regulation. We confirmed the associations at these regions through targeted resequencing in the same cohorts, boosting common variant signals at associated loci (p~5.19[-8] in *VWA8*, p~5.33[-9] in *KSR2*). We were able to replicate our top hits in the GIANT Consortium, primarily using a gene-based test (p=0.01), suggesting a shared influence on height in both KhoeSan and European populations. Interestingly, while both populations identify as KhoeSan, they historically have had different cultural practices and patterns of admixture with neighboring groups, resulting in a significantly different local ancestry patterns. This affects cross-population replicability and places additional importance on understanding fine-scale population structure in these communities. Taken together, these results show that genetically diverse cohorts can be effectively leveraged for characterization of genetic architecture and large effect variants, and contribute to the trans-ethnic basis of height.

469W

Highly differentiated *HLA-DPA1* **variant rs3077 may account for global disparity of hepatitis B virus infection.** *P. An, C.A. Winkler.* Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

Hepatitis B Virus (HBV) infection is a pandemic global disease. However, HBV prevalence varies greatly in the world populations, with high prevalence in East Asia and Africa and low prevalence in Europe. Recent GWAS and candidate studies have reported that SNP rs3077 in the *HLA-DPA* gene is consistently associated with HBV infection and HBV clearance. The risk C allele of rs3077, located in the 3'UTR, was associated with lower *HLA-DPA1* mRNA expression in normal human liver. To evaluate whether *HLA-DPA1* rs3077 contributes to the global disparity of HBV prevalence, we compared the global distribution of *HLA-DPA1* allele frequencies from HGDP and HapMap data and HBV prevalence in the world populations. We calculated Fixation index (Fst) values reflecting population differentiation of rs3077 among populations. Based on our previous finding (J Infect Dis. 2011;203:943-7), the risk ancestral C allele of rs3077 was associated with higher acquisition risk of HBV infection (Odds ratio [OR] 1.61) and failure to clear HBV infection (OR = 0.42). We calculated population attributable risks of 36% for HBV infection and 56% for failure to clear HBV, respectively, for the risk rs3077 C allele in Han Chinese. The rs3077 C allele frequency was highest in Africa Yoruba (74%) and Han Chinese (63%) and lowest in Europeans (12%). The pairwise Fst values of 0.25 between Asians and Europeans and 0.42 between Africans and Europeans indicate that rs3077 is highly differentiated among populations. In a simple linear regression analysis of 20 populations with both rs3077 allele frequencies and HBsAg (HBV surface antigen) prevalence rates available, the rs3077 C allele frequency was in strong positive correlation with HBsAg prevalence. The correlation between rs3077 and HBV carrier rates was highly correlated (correlation coefficient 0.81, $P < 0.001$). Although a causal relationship remains to be established, the *HLA-DPA1* rs3077, would explain 66% ($r^2=0.66$) of the variance in global HBV prevalence. The highly-stratified distribution of rs3077 across human populations and correlation with HBV prevalence suggest that *HLA-DPA1* rs3077 may account in part for disparate HBV chronic infection rates in human populations. (Funded by the National Cancer Institute under Contract HHSN261200800001E).

470T

Genetic variation and smoking behavior among diverse populations: Multi-Ethnic Genotyping Array (MEGA) analysis in the PAGE Study. *I. Cheng[1], Y. Li[1], S.L. Park[2], K. Young[3], S. Bien[4], C. Schurmann[5], T. Matise[6], L. Hindorff[7], R. James[8], S. Buyske[6], S.P. David[9], N. Zubair[4], C.A. Haiman[2], L.R. Wilkens[10], R. Loos[5], U. Peters[4], C. Kooperberg[4], K. North[3], D.O. Stram[2], L. Le Marchand[10] for the PAGE Study.* 1) Cancer Prevention Institute of California, Fremont, CA; 2) University of Southern California, Los Angeles, CA; 3) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 4) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Rutgers University, New Brunswick, NJ; 7) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD; 8) Clinical and Health Services Research, NIMHD, NIH, Bethesda, MD; 9) Stanford University, Palo Alto, CA; 10) University of Hawaii Cancer Center, Honolulu, HI.

Background: Globally, cigarette smoking accounts for 6.1 million deaths and 143.5 disability adjusted life years and is the primary risk factor for many cancers as well as other common chronic diseases such as cardiovascular disease. Numerous common variants have been associated with various smoking traits. However, many of these variants have been studied in predominately European populations. The objective of this analysis was to identify novel genetic risk variants for smoking behavior, using the Illumina Multiethnic Genotyping Array (MEGA) that was designed specifically to improve variant discovery and fine-mapping for metabolic, anthropometric, and lifestyle loci in non-European populations. Methods: We included 14,785 smokers from the Population Architecture using Genetics and Epidemiology (PAGE) Study that were genotyped using MEGA with information on the number of cigarettes smoked per day. We examined over 1.1 million SNPs with a minor allele frequency (MAF)>0.01. Race/ethnicity-specific ordinary linear regression results for African Americans, Asian Americans, Native Hawaiians, and Native Americans and generalized estimation equation results for Hispanics were meta-analyzed. Results: We confirmed associations for known smoking loci at chromosome 15q25.1 and chromosome 19q13.2. Potential novel associations for the number of cigarettes smoked per day were observed at chromosome 8p23.2 with rs7003625 in *CSMD1* (average MAF= 0.31, beta= 0.040, standard error= 0.008, P= 4.1x10^{-7}) and chromosome 20q11.2 with rs182236729, a missense variant in *MYH7B* (average MAF =0.01, beta= -0.641, standard error= 0.129, P=6.9x10^{-7}). There was no evidence of heterogeneity in associations by race/ethnicity. Conclusions: These preliminary findings suggest novel associations for smoking behavior among diverse racial/ethnic populations. Future work will further examine rare variants, fine-map known smoking risk loci, and incorporate data on imputed genotypes. The discovery of novel risk variants and the genetic characterization of known smoking loci in non-European populations will increase our understanding of the shared and population-specific genetic susceptibility of smoking. Such findings may ultimately guide tailored smoking cessation and screening efforts for tobacco-related diseases.   .

471F

Exposure to polychlorinated biphenyls (PCBs) and uric acid transporter polymorphisms in American Indians: The Strong Heart Family Study. *G. Chittoor[1], K. Haack[2], A. Sjodin[3], S. Laston[4], L.G. Best[5], J.W. MacCluer[2], J.G. Umans[6], S.A. Cole[2], A. Navas-Acien[7,8,9,10], V.S. Voruganti[1].* 1) Department of Nutrition, and Nutrition Research Institute, University of North Carolina at Chapel Hill, NC; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 3) Centers for Disease Control and Prevention, Atlanta, GA; 4) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX; 5) Missouri Breaks Industries Research Inc., Timber Lake, SD; 6) Georgetown-Howard Universities Center for Clinical and Translational Science, Washington, DC; 7) Department of Environmental Health Sciences, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 8) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 9) Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 10) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

   Background: Polychlorinated biphenyls (PCBs) are among several persistent organic pollutants (POPs) recognized to be toxic to the environment and humans by the United Nations environment program at their Stockholm Convention. Like other POPs, PCBs are highly resistant to biodegradation and tend to accumulate in living organisms, particularly in lipid rich liver and adipose tissues. Studies have shown a positive relationship between exposure to PCBs and metabolic disease risk factors such as hyperuricemia, hypertension, insulin resistance, inflammation, and oxidative stress. Previously we have shown that 25% of American Indians participating in the Strong Heart Family Study (SHFS) had hyperuricemia or elevated serum uric acid (SUA) concentrations and that SUA is strongly influenced by uric acid transporter gene variants. Therefore, our aim in this study was to assess genetic influences on serum PCBs and their relationship with SUA in SHFS.   **Methods**: Serum levels of 24 PCBs were ascertained in a cohort of 169 SHFS participants using GC/HRMS. DNA samples were genotyped using Illumina Vera Code Golden Gate genotyping assay kit for 78 SNPs that were selected for their strong association with SUA concentrations in previous studies. A variance components-based quantitative genetic analysis was performed using SOLAR. The association of serum PCBs with SUA-associated SNPs was estimated using a measured genotype analysis accounting for family relationships.   **Results:** Mean values of serum PCB concentrations ranged from 0.94 to 183.33 pg/g. Serum PCBs showed a positive but non-statistically significant trend of association with SUA concentration. Of the 24 PCBs, 7 analytes (PCB66: 2,3',4,4'-tetraCB; PCB99: 2,2',4,4',5-pentaCB; PCB105: 2,3,3',4,4'-pentaCB; PCB118: 2,3',4,4',5-pentaCB; PCB183: 2,2',3,4,4',5',6-heptaCB; PCB206: 2,2',3,3',4,4',5,5',6-nonaCB; PCB209: decaCB) exhibited moderate-to-high heritabilities (p < 0.03). Significant associations were observed between uric acid transporter SNPs and PCBs, (p < 0.006, MAF between 12 and 42%, effect sizes up to 7%) rs2078267 (*SLC22A11*) with PCB105, rs2242206 (*SLC16A9*) with PCB118 and PCB66, rs1183201 (*SLC17A1*) and rs942379 (*SLC17A3*) with PCB206 and PCB209, and rs2231142 (*ABCG2*) with PCB183. **Conclusions:** Several of the serum PCB measures were heritable and associated with uric acid transporter gene variants suggesting a pleiotropic relationship between serum PCBs and SUA concentrations in SHFS Participants. .

472W

A variance-components approach to identifying causal haplotypes in quantitative-trait linkage. *J.E. Hicks[1,2], M.A. Province[1,2].* 1) Division of Biostatistics, Washington University School of Medicine in Saint Louis, St Louis, MO; 2) Division of Statistical Genomics, Washington University School of Medicine in Saint Louis, St Louis, MO.

   Variance-components (VC) linkage analysis has been an important tool for identification of quantitative trait loci in humans. However, the VC model has some drawbacks. In conventional linkage analysis, pedigrees are assumed to be independent. This assumption may not always hold, as research shows that cryptic relatedness between 'unrelated' individuals is common. Additionally, effect sizes generated only give the amount of variance attributable to a locus, and not which variants are most influential to that variance.   Recent exponential population growth has resulted in an excess of rare variants across the genome. Population genetics theory predicts that these rare variants reside on long haplotypes in the population. Long shared haplotypes between individuals are identifiable from high-density genotype data both within and across pedigrees. Incorporation of haplotype information may increase power to identify rare variants influencing common traits.   We propose a method based on a linear mixed-effects model to find haplotypes responsible for a linkage signal. Phased haplotype data contains much of the information provided by identity-by-descent (IBD) covariance matrices. Instead of an IBD covariance matrix for a locus, we include a random effect indicating an individual's haplotype state alongside a polygenic effect. Statistical significance can be determined by likelihood ratio test. Duing the model fit, best linear unbiased predictions (BLUPs) of the effect of each haplotype are generated.   To test the power of this method we generated simulated data. One hundred replicates were generated for 50 pedigrees with genotypes based on the Affymetrix 6.0 chromosome 12 marker set. P henotypes were generated based on genotypes to have a standard normal distribution and heritability of 50%. Each pedigree had a 50% chance of carrying a 5,000 SNP haplotype containing a rare variant influencing the trait. We fit a model with 8 SNPs per haplotype in a 1Mb region around the rare variant.   This haplotype VC model showed 100% power at α=0.05 to detect a disease haplotype with effect +2σ . Powers for smaller effect sizes were 95%, 70%, and 23% for sizes +1.5σ, +1σ, and +0.5σ, respectively. In null data, the method showed a false positive rate of 3%.   The ability to identify BLUPs for each haplotype makes this a useful method to better understand the results of VC linkage. Additionally, BLUPs can allow more accurate selection of individuals to sequence for further research.

473T

Association analyses of myopia in multiplex African-American families using FBAT and a rare variant FBAT with exome chip data. *C.D. Middlebrooks[1], C.L. Simpson[2], A.M. Musolf[1], L. Portas[3], F. Murgia[4], D. Stambolian[5], J.E. Bailey-Wilson[1]. 1) National Human Genome Research Institute, Baltimore, MD; 2) University of Tennessee Health Science Center, Memphis, TN; 3) Università degli Studi di Verona, Cagliari Area, Italy; 4) Institute of Population Genetics - National Research council of Italy, Cagliari Area, Italy; 5) Scheie Eye Institute - Penn Presbyterian Medical Center, Philadelphia, PA.*

Myopia is an eye condition in which the light entering the eye does not focus on the retina and instead focuses on a point in front of the retina and results in distant objects appearing out of focus. Within recent years, the incidence and prevalence of myopia have increased, reaching epidemic proportions in several Asian countries. Myopia currently affects about 1 in 4 Americans. Research efforts have sought to identify both genetic and environmental causes of this phenomenon. Populations of African descent have lower rates of myopia than Asian populations, but multiplex families from the African-American (AA) population may represent an opportunity to identify unique susceptibility and protective alleles. We have performed an association study examining common and rare variants using Exome Chip genotyping (Illumina HumanExome v1.1 array plus 24,263 custom SNPs) within a family - study framework in 106 African-American families from the Philadelphia area who have multiple individuals affected with myopia (The Penn Family Study). Individuals in the families were defined as myopic if their average refractive error was <= -1 Diopter (D) and were considered unaffected if their average refractive error was > 0.0 D; others were coded as having an unknown phenotype. More stringent rules were used to code children as unaffected since children's eyes are constantly changing during childhood and adolescence. After quality control there were 242,901 SNPs (3300 AIMs, 5300 common SNPs across the genome and 240,000 coding SNPs) available for analysis. Single-variant analyses using FBAT resulted in a suggestive-level association signal in African Americans within the psoriasis susceptibility 1 candidate 3 gene (rs887468 in PSORS1C3) (P=1.2E-05) on chromosome 6. This association signal is intriguing but the analysis of each genetic variant individually does not adequately make use of the many rare variants available in this dataset. To increase power, we are performing gene-based tests using a rare variant analysis approach within the Family Based Association Test (FBAT) software suite. Using this analysis, we expect to increase our power to identify an association by testing for an association at the gene-level which combines the effects of multiple variants within a gene.

474F

The impact of population structure on rare variant association tests. *E. Persyn[1], L. Bellanger[2], R. Redon[1], C. Dina[1]. 1) l'institut du thorax – UMR Inserm 1087 / UMR CNRS 6291, Nantes, France; 2) Laboratoire de Mathématiques Jean Leray – Université de Nantes / UMR CNRS 6629, Nantes, France.*

Rare variant association studies are now enabled with the advances of next-generation sequencing technologies. They aim to identify new genes of susceptibility for complex diseases. Due to the power limitation of single marker tests to detect association signals, a strategy is to test groups of rare variants, commonly defined as genes. Numerous methods have been proposed during the last decade to test heterogeneous groups of rare variants. Population structure is known to be a confounding factor in genome-wide association studies for common variants. However the impact of the genetic stratification was less studied for rare variant association tests. Some authors as Mathieson and McVean (2012) and O'Connor et al. (2013) have shown that population stratification inflated the results of rare variant association tests but this field requires more investigation. Evaluating the effect of this stratification is important as ad hoc controls is a cost-effective option and often adopted by researchers. We aim to better understand the impact of population structure on different rare variant association tests. We conducted genetic simulations based on a coalescent model implemented in the COSI program (Schaffner et al., 2005). We considered in the demographical model several populations whose geographical positions are defined by mutual migration rates. We simulated different scenarios. The varied parameters are the migration rates, the time to the split between populations and the number of controls sampled from each population. Finally we applied ten rare variant association tests with different strategies in order to identify those who are less sensitive to stratification. We observe that some statistical association tests as SKAT (sequence kernel association test) (Wu et al., 2011), SKAT-O (Lee et al., 2012) and KBAC test (kernel-based adaptive cluster) (Liu and Leal, 2010) are more sensitive to population structure than other statistical such as CAST (cohort allelic sum test) (Morgenthaler and Thilly, 2007) and the weighted sum test (Madsen and Browning, 2009). Appropriate correction methods should be considered for various rare variant association strategies.

475W

Local genetic correlation provides insights into shared genetic basis of 30 complex traits. *H. Shi[1], B. Pasaniuc[2,3].* 1) Bioinformatics, University of California Los Angeles, Los Angeles, CA; 2) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, CA; 3) 3Dept of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, CA.

Many complex traits and diseases share a correlation at a phenotypic level. Such correlations can be attributed to shared environmental or similar genetic architectures. Quantifying the correlation in phenotypes that is due to genetic variations is of great interest in understanding the causal relationship between complex traits. Standard approaches to estimate genetic correlation require individual-level genotype data, and are hindered by the availability of such data due to privacy concerns. Cross-trait LD Score regression (Bulik-Sullivan et al., Nat. Genet. 2015) provide the convenience to estimate genome-wide genetic correlation from GWAS summary statistics, but its random-effect assumption renders it less suitable to estimate genetic correlation at a single locus, where few SNPs may be causal. Here, we propose an estimator of genetic correlation between pairs of traits at a single locus in the genome from summary GWAS data while taking LD and overlapping GWAS samples into account. Importantly, our estimator does not make any assumption on the number of causal variants or their effect sizes by treating them as fixed-effects in our model. Through extensive simulations, we show that our estimator is unbiased given in-sample LD, and approximately unbiased given external reference LD. We apply our estimator to GWAS summary data of 30 complex traits spanning 2.4 million phenotype measurements. We observe substantial local genetic correlations among HDL, BMI, and years of education and many other pairs of traits. To gain insights into the causality link among traits, we decompose the genetic correlation across genomic loci, and identify loci that contribute significant amount of genetic covariance.

476T

To ERV is Human: A phenotype-wide scan linking previously unrecognized heterogeneous human endogenous retrovirus insertions to complex diseases. *A. Wallace[1], B. Rhead[2], X. Shao[2], C. Metayer[1], J. Wiemels[3], S. Francis[3,4].* 1) Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley, Berkeley, CA; 3) Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 4) Community Health Sciences, University of Nevada, Reno, Reno, NV.

Approximately 8% of the human genome is comprised of ancient retroviral insertions (ERVs), which integrated into germ cells and have been preserved through natural selection. While the function of ERVs as regulators of gene expression and immune modulation is well established, it was assumed until recently that the locations of ERV insertions in the genome were fixed across individuals and so their potential contribution to germ-line genetic risk of complex diseases was largely overlooked. However, studies have shown that the most recent ERV integration, HERV-K, is active in a range of complex conditions from cancer to neurologic diseases. After demonstrating the heterogeneous nature of HERV-K insertion sites genome-wide, our group has associated these heterogeneous insertions with SNPs previously established through GWAS of complex phenotypes using a hypothesis-free approach. Due to the functional nature of ERVs, they serve as excellent candidates for causal variants contributing to the hidden heritability of complex diseases. Previously we used an in-house pipeline, HERVnGoSeq, on the 1000 Genomes Phase 3 whole genome sequencing data (n=2557) to 1) identify novel HERV-K insertion sites not present in the human reference and 2) characterize the heterogeneity of known and novel HERV-K insertion sites within and between populations world-wide. Here, we first tested the association of 1252 potential insertions with 24 million SNPs genome-wide among the 1000 Genomes Phase 3 European population (n=489). Insertion sites that were mappable using SNPs were defined as sites with a single, genome-wide significant peak and were confirmed by visual inspection of Manhattan plots. Among mappable insertion sites, associated SNPs positioned within 10Kb of the SNP with the lowest p-value and with p-value $< 1 \times 10^{-2}$ will next be extracted and the UCSC Genome Browser PubsMarkerSnp table and the NHGRI-EBI GWAS Catalog will be used to identify previously published associations of these SNPs with complex traits. Of the nominated SNPs, we expect to identify associations with phenotypes in previous publications. For phenotypes identified via this method, ERVs may play an etiologic role, warranting further study. Furthermore, this tool can be used in future GWAS to identify ERV variants as risk factors for disease.

477F

Generalized polygenic risk prediction for non-medical traits in a direct-to-consumer setting. *M.L. Multhaup, R.P. Smith, D. Hinds, N.A. Furlotte.* 23andMe, Mountain View, CA.

   The use of genetic risk prediction to classify individuals into disease risk categories based on their genetics has remained a challenging problem despite dramatic growth in the number of disease-associated variants identified by genome-wide association studies (GWAS). For a small number of conditions, GWAS have identified single, large-effect variants that effectively discriminate high-risk individuals. For many common complex traits, however, GWAS have found many small effect variants, and it is less clear how to use these variants to quantify risk. Recent work has shown that polygenic genetic risk scores (GRS) can be used to identify high-risk individuals. However, most work has been conducted on a limited set of well-studied phenotypes within clinically ascertained cohorts. The generalizability of these methods has yet to be determined. Furthermore, few studies have examined how GRS results could be communicated to at-risk individuals. These issues are important for the future of personalized medicine, which hinges on the ability to accurately quantify and communicate risk to clinicians and the public. In this study, we developed a generalized system for building polygenic risk prediction models and used it to systematically evaluate the utility of GRS using the cohorts at 23andMe. 23andMe has one of the largest databases of genetic and phenotypic information with over 1.2 million customers, 80% of whom have consented to participate in research and contributed more than 345 million points of phenotypic data. Using these data, we created polygenic risk models for dozens of non-medical traits and validated model performance using multiple metrics and separate testing cohorts. In addition, we have implemented a system for translating model output into easily-understandable consumer reports. We present accuracy metrics for over a dozen GRS models created with cohorts of 50-140,000 individuals and Area-Under-Receiver-Operator-Curves (AUROCs) ranging from 0.55 to 0.737. We discuss how AUROCs, despite being a common one-number summary metric, do not capture relevant model features necessary for evaluating the accuracy of information returned to recipients. To our knowledge, this study is one of the first to evaluate the applicability of GRS methods over multiple traits using information derived from a large database of self-reported phenotypes. Our results provide a blueprint for developing complex genetic risk estimates in a direct-to-consumer setting.

478W

Inheritance modes specific pathogenicity prioritization (ISPP) for human protein coding genes. *J. Hsu[1], J. Kwan[1], Z. Pan[1], M. Garcia-Barcelo[2], P. Sham[1,3], M. Li[1,3].* 1) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, HK; 2) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, HK; 3) Centre for Genomics Science, Li Ka Shing Faculty of Medicine, The University of Hong Kong, HK.

   Motivation: Exome sequencing studies have facilitated the detection of causal genetic variants in yet-unsolved Mendelian diseases. However, the identification of disease causal genes among a list of candidates in an exome sequencing study is still not fully settled, and it is often difficult to prioritize candidate genes for follow-up studies. The inheritance mode provides crucial information for understanding Mendelian diseases, but none of the existing gene prioritization tools fully utilize this information.**Results:** We examined the characteristics of Mendelian diseases genes under different inheritance modes. The results suggest that Mendelian-disease genes with autosomal dominant (AD) inheritance mode are more haploinsufficiency and *de novo* mutation sensitive, whereas those autosomal recessive (AR) genes have significantly more non-synonymous variants and regulatory transcript isoforms. In addition, the X-linked (XL) Mendelian-disease genes have fewer non-synonymous and synonymous variants. As a result, we derived a new scoring system for prioritizing candidate genes for Mendelian diseases according to the inheritance mode. Our scoring system assigned to each annotated protein-coding gene (N=18,859) three pathogenic scores according to the inheritance mode (AD, AR, and XL). This inheritance mode specific framework achieved higher accuracy (AUC=0.84) in XL mode.**Conclusion:** The inheritance mode specific pathogenicity prioritization (ISPP) outperformed other well-known methods including Haploinsufficiency (HI), Recessive (REC), Network centrality (NET), Genic Intolerance (RVIS), Gene Damage Index (GDI) and Gene Constraint (CONS) scores. This systematic study suggests that genes manifesting disease inheritance modes tend to have unique characteristics.**Availability:** ISPP is included in KGGSeq v1.0 (http://grass.cgs.hku.hk/limx/kggseq/), and source code is available from (https://github.com/jacobhsu35/ISPP.git).

479T

GWAS of circulating soluble receptor for advanced glycation end products: Evidence from the Long Life Family Study (LLFS), the Atherosclerosis Risk in Communities Study (ARIC), and the Rotterdam Study (RS). *P. An[1], N.M. Maruthur[2], S. Ligthart[3], M. Li[2], A.T. Kraja[1], M.K. Wojczynski[1], B. Thyagarajan[4], J.H. Lee[5], B.C. Astor[6], I. Miljkovic[7], A. Dehghan[3], K. Christensen[8], T.T. Perls[9], A.B. Newman[7], M.A. Province[1].* 1) Washington University, St Louis, MO; 2) Johns Hopkins University, Baltimore, MD; 3) Erasmus University, Rotterdam, The Netherlands; 4) University of Minnesota, Minneapolis, MN; 5) Columbia University, New York City, NY; 6) University of Wisconsin, Madison, WI; 7) University of Pittsburgh, Pittsburgh, PA; 8) University of Southern Denmark, Odense, Denmark; 9) Boston University, Boston, MA.

The importance of glycation has been highlighted in aging processes and glycosylation related diseases. Advanced glycation end-products (AGEs) and their specific receptor play a central role in vascular outcomes. sRAGE, a soluble form of the AGEs receptor, has been recognized with a counter-regulatory mechanism, identifying it as a biomarker of endothelial function or early stage atherosclerosis. To identify loci associated with circulating sRAGE, we assessed SNPs imputed from the 1000 genome project and conducted GWAS using 6,579 participants of European ancestry in the LLFS (nondiabetic subjects, discovery), ARIC (replication), and RS (nondiabetic subjects, replication). In all three studies, levels of sRAGE were natural log-transformed to approximate normality, and adjusted for age, sex, BMI, and GWAS principal components. sRAGE levels were additionally adjusted for field centers (LLFS), current smoking status (LLFS, RS), and educational level, eGFR, prevalent diabetes and CHD (ARIC), and a kinship model was used to correct for random effects of relatedness (LLFS). A linear mixed effects model, assuming additive genetic effects, was used for testing associations. In the discovery phase, a significant SNP (rs190643062, MAF = 0.011, p = 3e-8, near upstream *CCDC71L* and downstream *PIK3CG*) along with five suggestive loci (p < 1E-6; *ST5*, *TRIM49B*, *LOC646813*, *AGBL1*, *SLC32A1*) were identified in the LLFS. While rs190643062 was not replicated (p = 0.36, ARIC; data not available in the RS), the intergenic locus was supported by a significant (p < 2e-4, simpleM test) and independent SNP near *PIK3CG* (rs73186275, p = 3e-5) in the ARIC (but not in the RS, p = 0.01). Further, rs190643062 is in perfect LD with a ~30 kb upstream SNP rs192033857 at 3'-UTR of *CCDC71L* that is conserved in mammals and is located in the promoter/enhancer histone marks of 7-10 tissues where a DNase and a protein bind; the polymorphism modifies binding motifs of *RXRA* and *SIX5* transcription factors. Interestingly, neighboring intergenic SNPs at this locus (e.g., rs342293, rs17477177, rs4730171) are associated with a broad spectrum of cardiovascular traits including platelet counts, lipids, BP, glucose, IR, atherosclerosis, carotid intima media thickness, presence of atherosclerotic plaque and pulmonary function. In conclusion, we identified a novel locus near *CCDC71L* and *PIK3CG* genes that regulates levels of circulating sRAGE. Further independent replications and function queries are warranted.

480F

Genome-wide association study of *Staphylococcus aureus* **infection and clonal complexes in related samples from Starr County, Texas.** *H. Chen[1], H. Hu[2], L.E. Petty[1], D.A. Robinson[3], H.T. Essigmann[4], R.S. Fischer[4], C.L. Hanis[1], C. Huff[2], E.L. Brown[4], J.E. Below[1].* 1) Human Genetics Center, Division of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX, United States of America; 2) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States of America; 3) Department of Microbiology, University of Mississippi Medical Center, Jackson, MS, United States of America; 4) Center for Infectious Disease, Division of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX, United States of America.

Staphylococcus aureus infections represent a severe public health problem in hospitals and communities. The underlying biology of susceptibility to *S. aureus* has been difficult to unravel due to complex interactions between the microorganism, the environment, and the host. Host genetic susceptibilities have been reported to be associated with malaria, HIV and cholera infections and identification of *S. aureus* susceptibility genes would be helpful in the development of novel intervention strategies. The purpose of this study was to identify genes and variants that are correlated with nasal carriage of *S. aureus,* and to investigate associations between genetic susceptibility and different *S. aureus* clonal complexes. 769 subjects from the Mexican-American community of Starr County, Texas were recruited, including 416 individuals from 176 households. Chip genotyping was conducted on Illumina Expanded Multi-Ethnic Genotyping Array and imputation to 1000 Genomes Project Phase 3 reference data was performed using standard procedures to obtain final genotype data. In addition, the carriage status of *S. aureus*, including clonal complex, was determined based on two nasal swabs collected 11-17 days apart. The sample size of the three different carriage states i.e., persistent, intermittent, and non-carrier was 127, 135, and 507, respectively, and the five major clonal complexes identified were CC5, CC8, CC30, CC45, and CC188. Using Efficient Mixed-Model Association eXpedited (EMMAX) we applied linear mixed models with kinship adjustment to identify variants associated with infection of and clonal complex of *S. aureus*. Furthermore, meta-analysis of current study results and a previous publication was conducted using METAL. In the meta-analysis, two suggestive associations were identified in the comparison between intermittent infection and non-carriers (rs115503846 on chromosome 1, p= 6.9×10[-07]; rs3861635 on chromosome 14, p= 5.8×10[-07]). Although we could not replicate the previously reported significant variants, either through direct replication or meta-analysis, several variants and regions were found significantly associated with carriage and specific clonal complexes of *S. aureus*. The significant association between these variants and infection of *S. aureus* confirmed the interaction of genetic susceptibility and *S. aureus* carriage states.

481W

Winner's curse in quantitative genomics studies. *G. Darnell[1], J. Tung[2], C. Brown[3], S. Muhkerjee[2], B. Engelhardt[1].* 1) Princeton University, Princeton, NJ; 2) Duke University, Durham, NC; 3) University of Pennsylvania, Philadelphia, PA.

Winner's Curse is a phenomenon characterized for common value auctions in economics to describe the winner of the auction tending to overpay for the item, and has been used to explain the overestimation of effect sizes and lack of reproducibility of associations that have plagued genomics studies [Zöllner et al., 2007]. The bias in effect size estimation traces back to agricultural QTL mapping studies as the Beavis effect, but has been widely overlooked in the human genetics community [Göring et. al., 2001]. Standard association mapping techniques in modern eQTL studies are mostly concerned with the trade-off between computational tractability and statistical power. However, little focus has been placed on accurate and reproducible effect size estimates, even when false positives are properly controlled and statistical power is high. A corollary to inflated effect sizes is that the causal association is rarely identified, but, instead, a non-functional genetic variant often has the most significant test statistic value. Even conditioning on detection of an association between the truly causal variant and gene expression level of interest, estimation of the strength of association is almost always inflated. We demonstrate with a suite of previously proposed and novel methods that trade off computational and statistical properties, that each method suffers from one or more of four consequences: (1) increasing false positives and decreasing true positives, (2) enrichment of discoveries at low minor allele frequency, (3) false discoveries that are not in high LD with causal variants, and (4) underestimation of locus heterogeneity. In addition, we show lack of reproducibility and lack of correspondence of effect sizes in real data between gene expression microarrays of the same individuals under baseline conditions and drug responses. We propose new directions for robust and accurate effect size estimates in attempt to increase reproducibility in eQTL studies and mitigate Winner's Curse.

482T

Shared genetic predictors of resting heart rate and all-cause mortality. *R.N. Eppinga[1], Y. Hagemeijer[1], S. Burgess[2], D.J. van Veldhuisen[1], P.B. Munroe[3,4], N. Verweij[1], P. van der Harst[1,5,6].* 1) University of Groningen, University Medical Center Groningen, Department of Cardiology, 9700 RB Groningen, the Netherlands; 2) Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, United Kingdom; 3) Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London, Queen Mary University of London, London, EC1M 6BQ, United Kingdom; 4) NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London EC1M 6BQ, United Kingdom; 5) University of Groningen, University Medical Center Groningen, Department of Genetics, 9700 RB Groningen, the Netherlands; 6) Durrer Center for Cardiogenetic Research, ICIN - Netherlands Heart Institute, 3511GC Utrecht, the Netherlands.

Background Resting heart rate is a heritable trait correlated with lifespan. Little is known about the genetic contribution of resting heart rate and its relationship with mortality. **Methods** We performed a genomewide association analysis of 19.9 million genetic variants in 134,251 individuals from UK Biobank to further our understanding of the genetic basis of resting heart rate. We then used the identified genetic variants as an instrument to study the association between resting heart rate, cardiovascular risk factors and all-cause mortality. **Results** Genomewide association analysis identified 76 loci associated with resting heart rate ($P<5×10^{-8}$), 58 of these were novel. An increase in genetically predicted resting heart rate of 5 beats per minute was associated with a 20% increased mortality risk (hazard ratio 1.20, 95% CI of 1.11-1.28, $P=8.20×10^{-7}$) translating to a 2.9 years reduction in life expectancy for males and 2.6 years for females. Genetically predicted resting heart rate was found to be associated with higher body-mass index and diastolic blood pressure, hypertension, diabetes, smoking, myocardial infarction, heart failure, supraventricular tachycardia, beta-blockers, calcium channel-blockers and device implantation (all $P<0.05$). However, we did not identify such factors as mediators on the causal pathway between resting heart rate and all-cause mortality. Candidate gene and pathway analyses provide strong support for a dominant role of cardiac development and structure. **Conclusion** We discovered 58 novel loci for resting heart rate and provide evidence for shared genetic predictors of resting heart rate and all-cause mortality. Large-scale replication efforts are currently underway.

483F

The susceptibility of genome-wide association studies to recent fine-scale population structure. *J.M. Granka, E. Han, P. Carbonetto, K. Rand, J.K. Byrnes, R.E. Curtis, A. Kermany, Y. Wang, K. Noto, E. Elyashiv, H. Guturu, N.M. Myres, E. Hong, C.A. Ball, K.G. Chahine.* Ancestry.com DNA, LLC, San Francisco, CA.

   Large, heterogeneous genetic datasets have identified associations of thousands of single nucleotide polymorphisms (SNPs), sometimes subtly, with hundreds of human diseases and traits. Numerous methods have been developed to account for confounding effects of population structure unavoidable in such datasets; these include linear mixed models (LMMs) and the inclusion of population-level principal components (PCs) in fixed models. However, the susceptibility of GWAS to subtle, fine-scale population structure unidentifiable from PCs or F_{ST} has not been fully studied. Here, we address whether ultra-fine scale population structure, such as from post-colonial U.S. history or recent British history (Leslie *Nature* 2015), could produce an inflation of false positive GWAS associations, even for common variants. To do so, we use simulated phenotype data and genome-wide SNP data from over 1 million AncestryDNA customers who have consented to research. Among these individuals, we identify very recent fine-scale population structure using community detection algorithms across a network of identity-by-descent (Han *ASHG* 2015). Identified sub-populations, such as individuals of the U.S. Appalachian region, correspond to history within the last several hundred years and are driven by both spatial and cultural factors such as religion. We simulate phenotypes under various null and alternative models, as well as under models where latent environmental variables (such as cultural habits) associated with sub-populations affect the phenotype to varying degrees. Under both balanced and imbalanced sampling across sub-populations, we examine results of GWAS obtained from various methods: from linear fixed models using PCs as covariates to LMMs controlling for cryptic relatedness. The susceptibility of current GWAS approaches to fine-scale structure is dependent upon the degree of genetic differentiation among identified sub-populations, as well as the degree of sampling bias. In cases where latent environmental variables affecting the phenotype are related to fine-scale population structure, accounting for such structure using LMMs markedly improves GWAS accuracy. This highlights that even when underlying confounders are unknown, accounting for fine-scale structure can lessen effects of hidden environmental covariates without compromising power. Furthermore, our results suggest that accounting for fine-scale structure is even more critical when studying rare variant associations.

484W

On testing of gene-gene interaction based on case-control data using genotype similarity between individuals. *P. Mondal, I. Mukhopadhyay.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India.

   Background: Some possible explanations for the limited success of genome wide association study (GWAS) are that the current biostatistical analysis paradigm ignores all prior knowledge about disease pathobiology and/or the linear modeling framework of GWAS considers only one marker at a time thus failing to exploit their full genomic context and giving rise to multiple comparison problems. Also gene–gene (G-G) or marker based interactions might be another cause for this. This led to the development of multilocus association that combines information from multiple markers at a time. In spite of some already available multilocus association methods for testing association, no study is done to see its effectiveness in detecting G-G interaction. **Aims and objectives:** We adopt a more holistic approach that explores G-G interaction, combining information from multiple markers of two genes. This new approach is promising in identifying the variants along with relatively potent causal markers in gene interaction models. It has the flexibility to use knowledge from other sources (biological pathways, databases etc) and helps to prioritize genetic variations to be analyzed for G-G interactions to interpret genetic association studies in a biologically meaningful manner. **Material and Methods:** Our method is based on case-control data and the information contained at each marker is captured through a kernel function based on genotypes of two individuals. We use this genotype similarity score or kernel score to develop a novel statistical method avoiding the huge burden of multiple comparisons to a great extent. The newly proposed statistic for testing any such G-G interaction is easy to compute but the calculation of p-value poses another challenge. However, we are able to derive the asymptotic distribution of the statistic under the null hypothesis of no interaction. **Results:** The asymptotic distribution of the test statistic helps in the fast calculation of p-value even in presence of thousands of markers spread over different genes. Extensive simulations studies show that our method is very powerful in detecting any such interaction under various genetic models and is very robust. **Conclusions:** This method will have immense importance to the geneticists because of its robustness, powerful detection of interaction that may occur with or without main effects, ability to fast calculation of p-value with less multiple comparison burdens.

485T

A robust powerful statistical method to integrate genotype and gene expression data identifies novel associated loci. *I. Mukhopadhyay, S. Das, R. Chatterjee.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India.

Genome wide association studies (GWAS) identify many SNPs that are associated with disease traits. However, there is always a possibility that underpowered single marker test misses out SNPs with moderate effect sizes and fails to capture information embedded in the rare variants that are believed to explain the missing heritability of polygenic traits. On the other hand, gene expression contains information about the deregulation of genes when compared between cases and controls. Due to multiple testing or other issues some signals may remain unidentified especially when the sample size is not so large. Moreover, RNA being unstable compared to DNA, high cost is involved in RNA analysis which tend researchers to have a large population of genotyped individuals with only few of them having expression. Thus, to unravel the genetic architecture of common or complex disease, it is intuitive to integrate information from various data sources to decode biologically sound interpretation on heritable traits. No standard statistical procedure is available to deal with the complexities of these datasets, which motivated us to propose a novel method that tests for multi-loci association in the existing scenario. Based on a simple consistent two-stage regression method using Least Squares procedure, we develop a statistical method that essentially integrates genotype data and gene expression data to study genetic association at a genome-wide level using case-control data. For a large number of individuals gene expression data are not available whereas genotype data are available for all individuals. We integrate the information contained in both data sources into a latent variable based model. Our simple yet powerful multi-loci association test integrates two databases that broadcasts more of the deep-seated features comprehensively in a single test, which would otherwise be lost when databases are considered in singularity. We have developed asymptotic distribution of our test statistic for fast calculation of p-value for real data set. Extensive simulation confirms that our method is robust to many genetic models and disease models with different types of penetrance. This method is very powerful in detecting some novel markers that might be associated with the disease phenotype and that may go missing by genotype-based test only. We also get promising result when applied this method at a genome-wide level even with a small gene expression dataset related to psoriasis.

486F

Genetic variants associated with C-Reactive protein levels in US minorities from the Population Architecture using Genomics and Epidemiology (PAGE II) Study. *K.K. Nishimura[1], C. Hodonsky[2], J. Haessler[1], C.S. Carlson[1], I. Cheng[3], R. Do[4], L.A. Hindorff[5], R. James[6], R.C. Kaplan[7], J. Kocarnik[1], L. Le Marchand[8], S.L. Park[8], A.P. Reiner[1], B. Thyagarjan[9], M. Verbanck[4], G.L. Wojcik[10], C.L. Avery[2], C.A. Haiman[11], R.J.F. Loos[4], T.C. Matise[12], K.E. North[2], U. Peters[1], C.L. Kooperberg[1] on behalf of the PAGE Study.* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Cancer Prevention Institute of California, Freemont, CA; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD; 6) Clinical and Health Services Research, NIMHD, NIH, Bethesda, MD; 7) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 8) Cancer Research Center, University of Hawaii, Honolulu, HI; 9) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 10) Stanford University School of Medicine, Stanford, CA; 11) Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 12) Department of Genetics, Rutgers University, Piscataway, NJ.

C-reactive protein (CRP) is a general marker of systemic inflammation. Consistent positive associations between CRP and coronary heart disease (CHD) have spurred research into the genetic, environmental, and pathophysiological factors influencing normal and disease-related CRP levels. Most genome-wide association studies (GWAS) have been conducted in European populations. However, variation in CRP values exists between racial/ethnic groups, and GWAS using admixed populations can increase statistical power to identify novel associations and fine-map known loci. In this study, we present preliminary results from a GWAS of 24,830 African Americans, Hispanics, Asians and Native Americans from the Population Architecture using Genomics and Epidemiology (PAGE II) consortium. Participants were genotyped on the Multiethnic Genotyping Array (MEGA), which was designed to improve variant discovery and fine-mapping for metabolic, anthropometric and lifestyle traits in US minorities. All racial/ethnic groups were pooled and analyzed in the R package GENESIS with linear mixed models. CRP values were natural log transformed, and all regressions were adjusted for age at time of CRP measurement, sex, body mass index, current smoking status, study, and the top 8 principal components of genetic ancestry derived from the pooled study. Analyses stratified by race/ethnicity were also performed, with results meta-analyzed using fixed-effect models in METAL. Overall, our study found 11 genome-wide significant loci ($p<5x10^{-8}$), confirming 9 loci previously associated with CRP, and identifying 2 novel loci: *TREM2* (6p21.1, rs2234256) and *CD300LF* (17q25.1, rs1037170). Pooled and race-specific meta-analyses had similar effect sizes and p-values. The 6p21.1 and 17q25.1 associations were strongest in African Americans and Hispanics, respectively, where the minor allele frequencies were highest. Both genes encode proteins involved in regulating the immune response. Imputation of all MEGA genotypes on additional participants to the 1000 Genomes Project and Haplotype Reference Consortium is currently underway for replication efforts. Future plans include burden association tests to identify rare-variant associations, conditional analyses to uncover independent signals within known and new loci, and gene-environment interactions with a CRP genetic risk score. We attribute our success to the inclusion of diverse populations and expect that our study will yield additional novel findings.

487W

Gene- and pathway-based association tests for multiple traits with GWAS summary statistics. *W. Pan, I. Kwak.* Division of Biostatistics, University of Minnesota, Minneapolis, MN.

To identify novel genetic variants associated with complex traits and to shed new insights on underlying biology, in addition to the most popular single SNP-single trait association analysis, it would be useful to exploit multiple correlated (intermediate) traits at the gene- or pathway-level by mining existing single GWAS or meta-analyzed GWAS data. For this purpose, we present an adaptive gene-based test and a pathway-based test for association analysis of multiple traits with GWAS summary statistics. The proposed tests are adaptive at both the SNP- and trait-levels; that is, they account for possibly varying association patterns (e.g. signal sparsity levels) across the SNPs and traits, thus maintaining high power across a wide range of situations. Furthermore, the proposed methods are general: they can be applied to mixed types of traits, and to Z-statistics or p-values as summary statistics obtained from either a single GWAS or a meta-analysis of multiple GWAS. Our numerical studies with simulated and real data demonstrated the promising performance of the proposed methods. In particular, we applied the proposed gene-based test to the summary statistics for sex stratified anthropometrics data from The Genetic Investigation of ANthropometric Traits (GIANT) consortium, identifying more genes with significant sex-specific associations than other existing methods. The proposed methods are implemented in R package aSPU, freely and publicly available at https://cran.rproject.org/web/packages/aSPU/.

488T

Designing of an efficient genotyping chip for discovery and pan-disease screening in the Veterans Administrations Million Veteran Program (MVP). *S. Pyarajan[1,2], T. Webster[3], Y. Shi[1], M. Li[1,5], A. Huda[4], A. Clark[4], J. Moser[8], S. Muralidhar[8], T. O'Leary[8], J. Concato[6,7], J. Gaziano[1,2] for The MVP Genomic Working Group.* 1) VA Boston Healthcare System, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Affymetrix Inc., Santa Clara, CA; 4) Booz Allen Hamilton, McLean, VA; 5) University of Utah, Salt Lake City, UT; 6) VA Connecticut Healthcare System, West Haven, CT; 7) Yale University School of Medicine, New Haven, CT; 8) VA Office of Research and Development, Washington D.C.

The Million Veteran Program (MVP) is a national program initiated by the Veterans Affairs (VA) to collect consented biosamples for personalized medicine research from a million Veterans. MVP has already collected over 470,000 samples and is expected to reach over a million in the next 4-5 years. Proposed initial analysis includes genotyping all the MVP samples. Lower density genotyping chips have recently emerged as a cost-effective means for genotyping a very large number of samples. A custom lower density genotyping chip was designed on the Affymetrix Axiom platform for genotyping the MVP samples. The MVP 1.0 genotyping array was designed to allow for investigation of markers of clinical significance for a wide variety of human conditions and diseases. We started with the Affymetrix Biobank Chip as the backbone and designed a set of custom markers for inclusion on the chip. The MVP 1.0 array includes 723,305 probes investigating about 668,543 markers. Clinically significant SNPs from over 3400 genes including all of the 57 genes recommended for reporting of incidental findings by the ACMG have been included. Over 800 clinical conditions, with an emphasis on conditions of high prevalence in US Veterans, are covered by the markers on the chip. For some of the clinically relevant markers we have included SNPs with allele frequencies as low as 0.005% for some of the ethnicities. We also enriched for HLA and KIR markers for multiple ethnicities. Our design philosophy also incorporated a low-density genotyping approach with ability to imputing to high density for wide coverage across the genome for multiple ethnicities. Markers for genome wide coverage for not only Caucasians but specifically for African-Americans and other ethnicities were selected by an imputation aware algorithm for incorporation. The imputation coverage (Mean r^2) for marker with MAF range of 0.05-0.5 in the CEU population is expected to be 0.909 while in YRI and ASW population it is expected to be 0.814 and 0.799 respectively. Overall, the MVP 1.0 Array is an efficient low cost genotyping array that can be used not only as a pan-disease screening chip to design future more comprehensive and potentially expensive "omic" assays for MVP samples but also for discovery with efficient imputation. .

489F

Mapping variants to amino-acid changes in three-dimensional protein space improves aggregate association test power and suggests mechanisms of action. *M. von Grotthuss[1], J. Florez[1,2,3], J. Flannick[1,2].* 1) The Broad Institute, Cambridge, MA; 2) Massachusetts General Hospital, Cambridge, MA; 3) Harvard Medical School, Cambridge, MA.

To explore rarer variants than those identified by GWAS studies of complex traits such as Type 2 Diabetes (T2D), aggregate tests have been proposed to increase power. As aggregate analysis is negatively impacted by the inclusion of non-functional variants, the majority of aggregate associations identified to date have tested variants that truncate the protein (PTVs), or are functional in cellular assays. However, PTVs exclude the majority of variants in a gene, cellular assays are seldom available, and bioinformatics algorithms have been shown to be poor predictors of cellular function. We hypothesized that three-dimensional (3-D) protein structure maps might enable more accurate groupings of variants for aggregate association tests. We developed a novel method to (a) map all missense mutations to 3-D structures of proteins; (b) identify patterns of amino-acid changes in 3-D space; and (c) cluster variants according to these patterns for use in aggregate association analysis. We applied our methodology to analyze missense variants in TBC1D1 and TBC1D4 as identified from sequence data of 16,857 multi-ethnic T2D cases/controls; we chose these genes because TBC1D4 was recently identified to carry a Greenlandic variant of large effect on T2D risk and has an established role in insulin signaling, while TBC1D1 has a potentially (but unverified in humans) analogous role to TBC1D4 based on animal models. PTVs in TBC1D1 (OR=2.14, p=0.02), as well as combined PTVs and bioinformatically deleterious variants in TBC1D4 (OR=1.2, p=0.03), demonstrated nominally significant aggregate association with increased T2D risk in the heterozygous state. However, bioinformatically deleterious missense variants in TBC1D1 showed no association (OR=1.02, p=0.88). In contrast, our mapping algorithm identified a cluster of mutations localized on the surface of the TBC1D1 Rab-GAP binding site, which in aggregate demonstrated a nominal association with T2D risk (OR=2.7, p=0.038). Rab-GAP plays a known role in localization of the glucose transporter GLUT4 to the plasma membrane in insulin responsive tissues, with TBC1D1 and TBC1D4 binding necessary for activation of Rab. Collectively, our results thus suggest (a) that disruption of TBC1D1 in humans increases risk for T2D; (b) that impaired Rab-GAP binding is responsible for this increase in risk; and (c) that analyzing variants in 3-D protein space can inform aggregate tests of missense variants when cellular assay data is unavailable.

490W

Principal component based adaptive-weight burden test for quantitative trait associations. *X. Wu.* Dept of Statistics, Virginia Tech Univ, Blacksburg, VA.

High-throughput sequencing has often been used to screen samples from pedigrees or with population structure, yielding genotype data with complex correlations attributed to both familial relation and linkage disequilibrium. Accounting for such genotypic correlations can improve power for genetic association testing. However, due to model restrictions, existing methods cannot make efficient use of the correlation information appropriately, especially when assessing the contribution of multiple genomic loci. Recognizing this limitation, we develop PC-ABT, a novel principal-component-based adaptive-weight burden test for gene-based association. This method uses a retrospective score test to incorporate genotypic correlations, and employs "data-driven" weights to obtain maximized test statistic. In addition, by adjusting the number of principal components that make major contributions to genetic association, PC-ABT is able to control the degree of freedom of the null distribution to improve power. Simulation studies show that PC-ABT is generally more powerful than previously proposed burden tests that allow related individuals, while controlling for the type I error rate. We illustrate the application of PC-ABT by an analysis of the systolic blood pressure data from the Framingham Heart study to detect associated genes with common and rare variants.

491T

Meta-GWAS in cystic fibrosis indicates common variation in regulatory regions of modifier genes contributes to meconium ileus. *B. Xiao[1,14], N. Panjwani[1,14], J. Gong[1], G. He[1], K. Keenan[2], F. Lin[1], D. Soave[1,3], M. Drumm[4,5], G. Cutting[6,7], M. Knowles[8], H. Corvol[9,10], L. Sun[3,11], J.M. Rommens[1,12], L.J. Strug[1,3,11,13].* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 4) Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio, USA; 5) Department of Genetics, Case Western Reserve University, Cleveland, Ohio, USA; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 7) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 8) Cystic Fibrosis Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, North Carolina, USA; 9) Pediatric Pulmonary Department, Hospital Trousseau, Assistance Publique-Hôpitaux de Paris (AP-HP), Institut National de la Santé et la Recherche Médicale (INSERM), U938, Paris, France; 10) Pierre et Marie Curie University–Paris 6, Paris, France; 11) Department of Statistics, University of Toronto, Toronto, ON, Canada; 12) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 13) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 14) These authors contributed equally to this work.

Genome-wide association studies (GWAS) by the International Cystic Fibrosis (CF) Gene Modifier Consortium (ICFGMC) previously identified SNPs near *SLC26A9* and *SLC6A14,* in addition to genes of the apical plasma membrane that reside alongside the causal CF transmembrane conductance regulator, as contributing to intestinal obstruction at birth (meconium ileus, MI). Here we report on a meta-GWAS for MI in an expanded ICFGMC population. CF patients (N=5,845) from the United States, Canada and France were genotyped genome-wide, and imputed to a common set of >12 million SNPs. Random effects meta-analysis determined association evidence at each SNP. Genome-wide significant findings were assessed in the context of available data from the Roadmap Epigenomics and ENCODE projects. SNPs annotated to genes of the apical plasma membrane as defined by AmiGO were tested as a set (n=157 genes; 4,172 SNPs), and a novel nested subset test was implemented to assess whether association was enriched when restricting to nearby *cis*-regulatory elements of the genes (n=157 genes; 212 SNPs). Five loci had SNPs that reached genome-wide significance, $P < 5*10^{-8}$. These included the previously reported *SLC26A9* and *SLC6A14*, and three novel loci containing *PRSS1*, *ABCG5/ABCG8* and *ATP12A*. *ATP12A* was recently reported to contribute to CF lung disease in animal models, and *PRSS1* encodes cationic trypsinogen, a biomarker of CF in newborns. The top ranked SNP, rs9969188, near the transcription start site (TSS) of *PRSS1* is in high linkage disequilibrium (LD) with a promoter variant associated with pancreatitis, and reported luciferase assays argue decreased PRSS1 levels may increase MI risk. Rs12710568, a SNP in high LD with the top ranked *SLC6A14* variant, falls just 5' of the *SLC6A14* TSS. Analysis of chromatin states, DNase hypersensitivity, phylogenetic conservation and transcription factor binding suggest associated SNPs also influence *SLC26A9* expression levels. In addition, independent of the 5 significant loci, our nested subset test suggests the association of apical constituents is enriched when restricted to nearby *cis*-regulatory regions (-5kb to +2kb of the TSS of each gene; p<0.0001).   The positions of the genome-wide significant SNPs, and of the subset of associated variants from the apical plasma membrane constituents suggest their contributions to MI risk may be through expression regulation. These results may inform a broader mechanism for variation in CF disease severity.

492F

Accounting for population stratification in joint analyses of host and pathogen genomes to control type I and type II error rates. *N. Chaturvedi[1,2], O. Naret[1], J. Fellay[1,2].* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

A joint analysis of host and pathogen genomes (genome-to-genome or G2G analysis) can identify variants in the host genome that put a selective pressure on the pathogen genome. G2G can suffer from inflated type I error rate as well as low power due to systematic differences in host and/or pathogen population. In this work, we show through several simulation studies that correcting for both host and pathogen stratification reduces false positive and false negative results. We used the Balding-Nichols model for generating the host SNPs. Using these simulated host SNPs, the pathogen data was generated from a combination of Balding-Nichols and logistic models. This gave us a binary matrix representing amino acid variations in the pathogen genome. To detect associations, we used logistic regression with dichotomized pathogen data as dependent variables and host SNPs as independent variables. To prevent spurious associations due to host population stratification, we included the top five principal components of the genotyping data as covariates. To correct for pathogen stratification, we included either viral genotypes, or alternatively linear, nonlinear or phylogenetic principal components of viral amino acid variation as covariates. In a scenario where the only observed associations were due to host and pathogen stratification, we observed a better control of type I error rate after adjustment for both host and pathogen genetic structure. For example, the p-value of association between a simulated host SNP and pathogen amino acid (both contributing to stratification, but not associated with each other) was 1.5×10^{-21}; it was still significant after correcting for host or pathogen stratification ($p_{host} = 2.9 \times 10^{-10}$; $p_{pathogen} = 5.2 \times 10^{-9}$), but correcting for both made the association non-significant ($p_{both} = 0.06$). When true associations between host and pathogen variants were simulated, we saw an overall improvement in the strength of associations (example: $p_{nocorrection} = 4.1 \times 10^{-6}$, $p_{host} = 2.3 \times 10^{-8}$, $p_{pathogen} = 3.5 \times 10^{-7}$, $p_{both} = 6.2 \times 10^{-10}$) upon adjustment for stratification from both sides. Furthermore, even when the stratification was only from one side (host or pathogen), adding both as covariates in the model did not over-correct or diluted the true signals. Currently we are applying our method to jointly analyze HIV host and viral genome for detecting potential host genetic effects on the diversity of HIV virus.

493W

Statistical methods for rare variant test for multiple phenotypes. *D. Dutta, S. Lee.* Department of Biostatistics, University of Michigan, Ann Arbor, MI.

In genetic association analysis, a joint test of related multiple phenotypes can provide novel insights into the genetic architecture of complex diseases. Although a number of methods have been developed for multiple phenotype tests for common variants, only a few exist for rare variants. Here we present two broad strategies to combine multiple phenotypes, nonparametric PCA-based and parametric regression-based approaches, and incorporate them into gene or region-based rare variant tests. The PCA-based approach performs a variance component test for associations using the principal components of the phenotypes, with no assumption on the effect sizes of the variants, and aggregates association signals across multiple PCs with leveraging on their orthogonality. The regression-model-based approach models the distribution of the effect sizes of the variants to multiple phenotypes through a correlation structure in mixed models and conducts a variance component test for associations. In addition, we combine the results of these different approaches through various aggregating strategies such as the minimum p-value and the weighted sum of test statistics. Furthermore, using Copula, we propose a resampling-based fast computation method to obtain a p-value of the minimum p-value-based combined test. From extensive simulation studies, we show that the proposed tests can improve power over a standard approach of aggregating single-phenotype test p-values, while maintaining type-1 error rate. The relative performance of these tests depends on the number of associated phenotypes and correlation patterns. The combined tests had robust power regardless of the genetic model. From the analysis of a real data example, we also show the superior performance of our methods.

494T

Assessing phenome-wide consequences of gene regulation using 40 human tissue expression prediction models. *H.K. Im[1], A. Barbeira[1], J. Zheng[1], E.S. Tortensen[2], J.M. Torres[1], S.P. Dickinson[1], K.P. Shah[2], H.E. Wheeler[3], T.L. Edwards[2], G.I. Bell[1], D.L. Nicolae[2], N.J. Cox[2].* 1) The University of Chicago, Chicago, IL; 2) Vanderbilt University, Nashville, TN; 3) Loyola University Chicago, Chicago, IL.

GWAS and sequencing studies have yielded thousands of genetic variants robustly associated with complex traits. However, the underlying biology of those associations needs to be further elucidated. To address this issue we have proposed a method called PrediXcan that links these genetic variants with likely causal genes using the correlations between genetically predicted expression levels and phenotypes. Our method tests whether the transcript-level consequences of genetic variation have an effect on a specific phenotype and is similar in some respect with Mendelian randomization-based approaches. Recently, we have developed an extension called MetaXcan, which can infer PrediXcan results using only summary results from GWAS. Among the advantages of our approach are that it a) directly tests a biological mechanism, b) provides gene level results, c) provides directions of effect (which can be used to prioritize drug targets since positively correlated genes have the potential to reduce disease risk when knocked down), d) the multiple testing burden is reduced, and e) provides tissue specific results. To implement our approach we have developed prediction models for gene expression in 40 tissues using the GTEx Consortium and Depression Genes Network data and applied it to 117 phenotypes with publicly available GWAS meta analysis results. These include results from large consortia such as DIAGRAM, MAGIC, PGC, among others. We validate our approach by re-identifying many established genes but in many cases, we find evidence that genes in the vicinity of reported ones are more likely mediators of the phenotype. These results, software, and all prediction models necessary to reproduce them or apply to new datasets are made publicly available on https://github.com/hakyimlab/PrediXcan.

495F

Efficient phenome-wide analyses of large-scale sequencing data. *X. Zhan[1], B. Li[2], G. Abecasis[3], D. Liu[4].* 1) University of Texas Southwestern Medical Center, Dallas, TX; 2) Vanderbilt University; 3) University of Michigan; 4) Pennsylvania State University.

The advancement of technology has enabled medical researchers to collect an unprecedented amount of big data that are represented in three dimensions: the growing number of phenotypes (e.g., thousands of phenotypes via Electronic Medical Records or biobanks), the increasing sample sizes in large studies (e.g., the UK10K and UK Biobank projects), and the millions of genetic markers (e.g., imputation outcomes). Traditional analytical approaches can take months or years, which is computationally inhibitive. Thus there is an emerging need to efficiently analyze these big datasets in order to gain scientific insights. In order to solve this problem, we developed a new computationally efficient model and improved computational speed suited for phenome-wide analysis. The software implementation (RVTESTS) can be readily applied. It is benchmarked on the UK10K dataset, which includes 1854 samples, along with their 46 phenotypes and 43 million variants, and can be analyzed within one day. Further evaluated in a larger simulated dataset (100,000 individuals and 50 phenotypes), RVTESTS can analyze half a million variants per day. Our software can also be easily improved using parallel computations. We conclude that the computational speed and memory consumption of this tool outperform other popular approaches. In all, we have proposed a useful and efficient approach to tackle the needs of performing phenome-wide analysis of large-scale data sets.

496W

Efficient calculation of genetic correlations across 700,000 pairs of diverse phenotypes. *A. Kleinman, N. Furlotte, A. Auton, D. Hinds.* 23andMe, Inc., 899 W. Evelyn Ave., Mountain View, CA 94041.

The genetic correlation between a pair of traits is the proportion of phenotypic variance attributable to shared genetics. Specifically, assuming an infinitesimal model, two traits are said to be genetically correlated if effect sizes of SNPs associated with the traits are themselves correlated. This is not necessarily the same as phenotypic correlation, as two traits can be highly correlated in phenotypic space but not in genetic space if their correlation is solely induced by shared environmental factors. We used LD Score Correlation, a recently developed method that calculates genetic correlation from genome-wide association study (GWAS) summary statistics, to compute genetic correlations across more than half a million pairs of self-reported phenotypes. We replicated dozens of published correlations and found tens of thousands more across a diverse array of traits, including autoimmune, cardiovascular, infection, lifestyle, morphological, neurological, sleep, personality and reproductive phenotypes. Given the direct availability of phenotypic data, we showcased situations in which high phenotypic correlation is observed in the absence of high genetic correlation, and vice versa. Collectively, these correlations paint both a broad and detailed portrait of shared genetics across the spectrum of human traits and diseases.

497T

Multiple HCV amino acid variants associate with interferon lambda polymorphism in genotype 1a, 1b and 3a infections. *J. Fellay[1], N. Chaturvedi[1], K. Chodavarapu[2], M. Lin[2], A. Osinusi[2], E. Svarovskaia[2], H. Mo[2], D. Brainard[2], M. Subramanian[2], J. McHutchison[2], S. Zeuzem[3].* 1) School of Life Science, EPFL, Lausanne, VD, Switzerland; 2) Gilead Sciences, Foster City, USA; 3) Department of Medicine I, J.W. Goethe University Hospital, Frankfurt, Germany.

   A specific human genetic polymorphism in the interferon lambda (IFN-λ) region, tagged by SNP rs12979860, is associated with spontaneous clearance of HCV infection and with response to interferon-based antiviral treatment. A recent report also showed a higher prevalence of pretreatment NS5A variant Y93H in genotype (GT) 1b patients homozygous for the C allele at rs12979860. Here, we evaluate the associations between HCV amino acid variants and IFN-λ polymorphism in >7000 patients from various geographical regions and ethnicities, infected with HCV genotype 1a, 1b or 3a. We used logistic regression to search for associations between rs12979860 and amino acid variation in NS3, NS4A, NS5A and NS5B in a total of 7131 chronically infected HCV patients (GT1a, n=3623; GT1b, n=2010; GT3a, n=1498). HCV consensus sequences were obtained by next generation of sequencing on Illumina MiSeq, and binary variables were created for each non-conserved amino acid present in at least 0.3% of samples. We ran the analyses on all samples and in subgroups defined by HCV genotypes. To prevent spurious associations due to host and viral stratification, we included the following covariates in all regression models: gender, country of origin, self-reported ethnicity, cirrhosis status, prior treatment experience, SVR12 information and viral genotype. Associations with a p-value lower than the Bonferroni adjusted threshold ($p < 3.6 \times 10^{-5}$) were considered significant. A total of 73 HCV amino acids were significantly associated with rs12979860 (8 in NS3, 3 in NS4A, 44 in NS5A and 18 in NS5B). The strongest associations were observed in and around the highly variable interferon sensitivity-determining region of NS5A (top associated variant: I280V, $p = 2.8 \times 10^{-23}$). We also confirmed the association between the NS5A variant Y93H and IFN-λ polymorphism ($p = 5.4 \times 10^{-15}$). In the genotype-specific models, we detected 68, 38 and 26 significant associations for genotype 1a, 1b and 3a, respectively. The difference in the number of significant associations between subgroups most likely reflects an effect of sample size on statistical power. We also observed that patients with the rs12979860 CC genotype have a higher prevalence of multiple amino acid changes, including Y93H, a naturally occurring NS5A inhibitor resistance-associated mutation. This study provides strong evidence of pervasive viral adaptation to innate immune pressure from the host during chronic HCV infection.

498F

Frequency of arylsulfatase A pseudodeficiency in healthy Mexican individuals, haplotypes construction and enzyme activity determination. *A. Juárez[1], S. Mendoza[2], T. Da Silva[1,2], A. Porras[1,2], E. García[2].* 1) Instituto de Genética Humana "Enrique Corona Rivera" , Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Laboratorio de Bioquímica IB, División de Genética, Centro de Investigación Biomédica de Occidente, Guadalajara, Jalisco, México.

   Introduction. One of every 50 healthy individuals has a substantial deficiency of arylsulfatase A (ASA, EC 3.1.6.8). These individuals express even a tenth of the normal enzyme activity without developing the disease. The ASA pseudodeficiency (PD-ASA) is mainly associated to the presence of two alleles: the first; c.1055A> G (p.N352S; rs2071421), causes the loss of one of the three N-glycosylation sites in ASA, while the second; c. * 96A> G (rs6151429), leads to the loss of polyadenylation signal in exon 8, both in *ARSA* gene. This work was performed in order to determine the frequency of PD-ASA in Mexico. Material and methods. Samples of 200 healthy mexican individuals were included in the study. The determination of enzyme activity of ASA in leukocytes was accomplished by a modification of the Percy assay and 4-nitrocatechol sulfate substrate (Sigma® LifeScience). The polymorphisms analysis was performed by RFLPs, with *Dde*I and *Bsr*I (Time-SaverTM New England BioLabs® Inc.). The genotypes were confirmed by Sanger sequencing with Big Dye Primer Kit (Applied Biosystems) and ABI Prism 310 (Applied BioSystems). Finally, through Arlequin v.10.04 haplotypes were constructed and the analysis of linkage disequilibrium was performed with X Cube. Results and discusion. The results of enzyme activity of the total population showed a 95% CI of 1.74-2.09 nmol / mg protein / min. Allele, genotype and haplotype frequencies of c.1055A> G and c.*96 A> G polymorphisms were obtained, the mutated allele (G) was observed in 21% (86 of 400 alleles) for the first polymorphism and 4% (16 of 400 alleles) for the second one. 15 individuals were positive for GG haplotype (frequency 0.04, 4%) with ASA activity of 0.53-0.95 nmol / mg protein / min and no AG haplotype was found. The linkage disequilibrium statistics were: D '= 1, r2 = 0.1477. Conclusion. For the first time in Mexico, the presence of the alleles of PD-ASA is demostrated, the frequencies were in Hardy-Weinberg equilibrium. Alleles are found in linkage disequilibrium. The GG haplotype was found in 15 individuals, so the frequency of PD-ASA in Mexico is 7.5%, corresponding to the observed in other populations. No AG haplotype was observed. One individual has an ASA activity of 13%, that individual was homozygous for the G allele of both polymorphisms.

499W

Modeling the covariance of effect sizes in a meta-analysis. *D. Duong[1], S. Snir[2], E. Kang[1], B. Han[3], E. Eskin[1].* 1) Computer Science, UCLA, Los Angeles, CA. USA; 2) Evolutionary & Environmental Biology, University of Haifa, Israel; 3) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center. Seoul. Republic of Korea.

Meta-analysis can be traced back to the 17th century studies of astronomy, and over the years, meta-analysis has become one of the most widely used method in many fields of research including medical, social, and biological sciences. Meta-analysis is a method that pools results from different studies on the same topic in order to infer a general conclusion and to measure the consistency among study findings. Two most popular meta-analysis paradigms are the fixed and random effects models. The fixed effect model assumes that all studies share a true underlying mean effects, whereas the random effect model allows for heterogeneity among the true effect sizes of the studies. The shortcoming of the classical random effects model is that it models heterogeneity assuming independent and identically distributed (i.i.d.) effect sizes. A more realistic scenario happens when some studies have samples that share certain characteristics so that the effect sizes of the studies tend to be more similar than usual. For example, consider several studies in a meta-analysis that target a disease with some treatment where half of the studies contain only male individuals and the other half contain only female individuals. Study findings using only male (or female) participants would be more similar among themselves; more precisely, the effects of the treatment among these studies are more correlated than effects between a male-only and a female-only study. In this paper, we introduce a meta-analysis method that models the covariance of the effect sizes among the studies in a meta-analysis. Our method extends a likelihood ratio procedure in a meta-analysis so that the covariance among the effect sizes can be easily accounted for. We use our new method to analyze 17 mouse High-density lipoprotein (HDL) studies containing 4,965 distinct animals. Our method attains better statistical power than previous methods and identifies more significant loci.

500T

Phenotype similarity regression for identifying the genetic determinants of rare diseases. *E. Turro[1], D. Greene[1], S. Richardson[2], NIHR BioResource.* 1) University of Cambridge, Cambridge, United Kingdom; 2) MRC Biostatistics Unit, Cambridge, United Kingdom.

Rare genetic disorders, which can now be studied systematically with affordable genome sequencing, are often caused by high-penetrance rare variants. Such disorders are often heterogeneous and characterized by abnormalities spanning multiple organ systems ascertained with variable clinical precision. Existing methods for identifying genes with variants responsible for rare diseases summarize phenotypes with unstructured binary or quantitative variables. The Human Phenotype Ontology (HPO) allows composite phenotypes to be repre- sented systematically but association methods accounting for the ontological relationship between HPO terms do not exist. We present a Bayesian method to model the association between an HPO-coded patient phenotype and genotype. Our method estimates the proba- bility of an association together with an HPO-coded phenotype characteristic of the disease. We thus formalize a clinical approach to phenotyping that is lacking in standard regression techniques for rare disease research. We demonstrate the power of our method by uncovering a number of true associations in a large collection of genome-sequenced and HPO-coded cases with rare diseases.

501F

SMART: Statistical Metabolomics Analysis – An R Tool. *Y.-J. Liang[1,2], Y.-T. Lin[2], C.-W. Chen[2], C.-W. Lin[2], K.-M. Chao[1,3], W.-H. Pan[4], H.-C. Yang[2,5,6,7,8].*
1) Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan; 2) Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan; 3) Department of Computer Science and Information Engineering, National Taiwan University, Taipei 10617, Taiwan; 4) Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; 5) Institute of Public Health, National Yang Ming University, Taipei 11221, Taiwan; 6) Department of Statistics, National Cheng Kung University, Tainan 701, Taiwan; 7) Institute of Statistics, National Tsing Hua University, Hsinchu 30013, Taiwan; 8) School of Public Health, National Defense Medical Center, Taipei 114, Taiwan.

Metabolomics data provide unprecedented opportunities to decipher metabolic mechanisms by analyzing hundreds to thousands of metabolites. Data quality concerns and complex batch effects in metabolomics must be appropriately addressed through statistical analysis. This study developed an integrated analysis tool for metabolomics studies to streamline the complete analysis flow from initial data preprocessing to downstream association analysis. We developed Statistical Metabolomics Analysis – An R Tool (SMART), which can analyze input files with different formats, visually represent various types of data features, implement peak alignment and annotation, conduct quality control for samples and peaks, explore batch effects, and perform association analysis. A pharmacometabolomics study of antihypertensive medication was conducted and data were analyzed using SMART. Neuromedin N was identified as a metabolite significantly associated with angiotensin-converting-enzyme inhibitors in our metabolome-wide association analysis ($p = 1.56 \times 10^{-4}$ in an analysis of covariance (ANCOVA) with an adjustment for unknown latent groups and $p = 1.02 \times 10^{-4}$ in an ANCOVA with an adjustment for hidden substructures). This endogenous neuropeptide is highly related to neurotensin and neuromedin U, which are involved in blood pressure regulation and smooth muscle contraction. The SMART software, a user guide, and example data can be downloaded from http://www.stat.sinica.edu.tw/hsinchou/metabolomics/SMART.htm.

502W

Incorporate technical variation to assess reproducibility of genome-wide methylation data. *W. Guan[1], Y. Bai[1], J.S. Pankow[2], E.W. Demerath[2], J. Bressler[3], M. Fornage[3], M.L. Grove[3], M.Y. Tsai[4].*
1) Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN; 2) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 3) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 4) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

Genome-wide DNA methylation measures are now routinely used to investigate their association with various outcomes of interest. Inherent to the array-based DNA methylation measures, e.g., Illumina HumanMethylation450 (HM450) or Infinium MethylationEPIC chips, is the associated technical variation of non-biological interest. A commonly adopted strategy is to remove non-biological variation through preprocessing or normalization procedures. The processed data are then treated as the raw data (with only biological variation) and used in any downstream analysis. Intuitively downstream analysis can have improved statistical power and better accuracy of parameter estimates by accounting for technical variation. We have proposed a mixed effects model framework to incorporate technical variation, and developed an expectation–maximization (EM) algorithm to estimate the model parameters. We applied our proposed methods to the Atherosclerosis Risk in Communities (ARIC) methylation study. Specifically, we considered the bead-to-bead variation in the HM450 data. We calculated the intraclass correlation coefficients (ICC) for each CpG probe to evaluate reliability of the Illumina HM450 chip after accounting for the bead-to-bead variation, an extension of a previous study (Bose et al., 2014). We observed that the ICCs were improved by a median of 0.04, and ~6,000 probes have the updated ICCs close to 1, suggesting that most of the technical variation could be explained by the bead-to-bead variation at these sites. We think that accounting for bead-to-bead varation can improve power of association tests for future epigenome-wide association studies (EWASs) using methylation arrays.

503T

Multi-marker methylation predictors of clinical biomarkers. *M.W. Hattab[1], G.v. Grootheest[2], R.F. Chan[1], R. Jansen[2], A.A. Shabalin[1], L.Y. Xie[1], M. Zhao[1], S.L. Clark[1], B.W. Penninx[2], K.A. Aberg[1], E. J.C.G. van den Oord[1].* 1) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA, USA; 2) Department of Psychiatry, VU university Medical Center / GGZ ingest, Amsterdam, the Netherlands.

Biomarkers have become a standard approach to predict the risk of disease, diagnose patients, and monitor response to treatment. Their measurement typically involves collecting multiple biological samples (blood, saliva, urine and/or hair) and the use of a wide variety of laboratory assays or clinical tests. This makes the use of biomarkers typically time consuming, complex, and expensive. In this study we examine whether commonly used biomarkers have specific DNA methylation signatures in blood that could be used as alternative indicators. Methylation marks are potentially ideal as they involve the stable methyl-cytosine bond that can be measured in easy to collect genomic (histone-free) DNA. Furthermore, as different biomarkers may be captured by a single methylation assay, they could potentially provide a convenient and cost-effective one-stop alternative. To test our hypothesis we generated methylome-wide data in 1,159 samples using an optimized methyl-CG binding domain sequencing (MBD-seq) protocol that provide a cost-effective approach for assaying methylation status of almost all CpGs in the human genome. Rather than using individual sites, we use a method that combines all associated individual methylation markers into multi-marker predictors. Regularization regression methods such as elastic nets will be main tools for this purpose. Such methodologies enable the user to fit a model where the number of predictors exceed by far the number of observations. Coupled with a screening rule that eliminates sites that most likely have little or no effect, we have successfully applied them to millions of sites. To obtain an unbiased assessment of the overall predictive power, a nested *10*-fold cross validation is implemented. To provide proof of concept, we first show that MBD-seq leads to a considerably more precise, likely because it assays many more sites, prediction of biological aging compared to common used methylation arrays. Next we study an extensive battery of biomarkers including blood lipids (e.g., total cholesterol), inflammation and immune response markers (e.g., absCRP), hormonal function (e.g., parathyroid hormone, vitamin D25, testosterone, and insulin-like growth factor-1) as well as measures of oxidative stress (e.g., 8-OHdG). In addition to study correlations of multi-marker predictors and the biomarkers they intend to measure, comparisons are made in term of their ability to predict short and long term clinical outcomes. .

504F

Association of chromosome 15q25 (CHRNA5/A3/B4) with smoking behavior generalizes to a Hispanic/Latino cohort. *L.S. Emery[1], T. Sofer[1], A.M. Stilp[1], S.M. Gogarten[1], M.P. Conomos[1], T. Louie[1], E. Perez-Stable[2], K.D. Taylor[3], C.C. Laurie[1], L. Bierut[4], N.L. Saccone[5], R.C. Kaplan[6].* 1) Biostatistics, University of Washington, Seattle, WA; 2) National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD; 3) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Torrance, CA; 4) Psychiatry, School of Medicine, Washington University in St. Louis, St. Louis, MO; 5) Genetics, Human genetics, Washington University in St. Louis, St. Louis, MO; 6) Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY.

Genetic variants associated with nicotine dependence have previously been identified and characterized, primarily in populations of European ancestry. In particular, variants in the chromosome 15 gene *CHRNA5* (encoding the α5 cholinergic nicotinic receptor subunit) have been unequivocally associated with nicotine dependence and heavy smoking in European, African, and Asian ancestry cohorts. However, no large-scale genome-wide study has examined the genetics of smoking behaviors in Hispanic/Latino populations. We examined genetic associations with smoking behaviors in participants from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). HCHS/SOL participants were recruited from four U.S. field centers (Miami, San Diego, Bronx NY, and Chicago). Consented individuals were genotyped at ~2.3 million SNPs, and imputed to the 1,000 Genomes Project phase 3 reference panel (~49 million imputed variants). Current smokers reported the average number of cigarettes smoked per day (CPD) in structured questionnaires. We performed GWAS of heavy (CPD > 10) versus light (CPD ≤ 10) smokers in 5,085 current smokers from the HCHS/SOL. We used a mixed logistic regression model that accounted for population structure, sampling design, shared household, relatedness, sex, and age. Heavy smoking varies substantially among HCHS/SOL participants, and rates of non-daily smoking are particularly high as compared to other U.S. populations. The known association of the CHRNA5 region with smoking generalized to the HCHS/SOL, with multiple correlated SNPs reaching genome-wide significance ($p \leq 5 \times 10^{-8}$). The known functional SNP, rs16969968, had a p-value of 2.20×10^{-7} with minor allele frequency (MAF) of 0.22 and an odds ratio of 1.32 for the minor allele (A). In groups based on self-reported country of origin, the minor allele frequencies for rs16969968 were: Central American = 0.18; South American = 0.19; Mexican = 0.18; Puerto Rican = 0.25; Dominican = 0.22. Conclusion: The association of variants in a nicotine receptor gene with smoking behavior, previously reported in Europeans, generalized to Hispanics/Latinos despite different patterns of smoking behavior. The generalization of the functional SNP rs16969968 to an additional population may provide further insight into its mechanism of function.

505W

A quasi-likelihood approach for transmission-based association mapping using sibship data. *H. Kulkarni, S. Ghosh.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India.

The classical transmission disequilibrium test (TDT) [Spielman et al. 1993] based on thetrio design is an alternative to the population based case-control design to detect geneticassociation as it protects against population stratification. Most of the complex diseases are governed by quantitative precursor and hence, it may be more prudent to analyze these quantitative traits rather than converting them in to endpoint traits. The test for transmission disequilibrium based on logistic regression [Waldman et al. 1999, Haldar and Ghosh 2015] use the same approach as the classical TDT that is based on allelic transmission from heterozygous parents. However, in the presence of the sibship data,the marginal effect of linkage can result in transmission bias of parental alleles and hence,the test is valid only for linkage. We modify the classical TDT procedure using quasi likelihood [Wedderbur 1974] based on Generalized Linear Regression model [McCullagh and Neldar 1989] using a logit link. Parameters of the models are estimated using the Generalized Estimating Equation (GEE) [Gourieroux, Monfort, and Trognon 1984; Liang and Zeger 1986] but this method is highly influenced by the outliers. We use a modifiedResistance Generalized Estimating Equation approach (RGEE) [Preisser and Qaqish1999, Preisser and Qaqish, 1996; Hall, Zeger, and Bandeen-Roche,1996] to down weight the outliers. Moreover, information on genetic association is contained not only on the allele transmitted by a heterozygous parent, but also the allele transmitted by the other parent. We modify the quasi likelihood approach by including the transmission information from both parents. Based on extensive simulations under different genetic models as well as for different probability models for trait value we compute the type-I errors and the powers of the proposed test procedures and compare the powers with FBAT [Lake et. al.2000]. We observe that the powers of the tests are comparable across different distributions of trait values. The simulations also indicate that the quasi likelihood approach is more powerful than FBAT and inclusion of transmission information from both parents increases the powers of the test procedures. The proposed approach can be easily extended for multivariate phenotypes comprising a mixture of the categorical, count and continuous phenotypes.

506T

Clustering RNA-seq expression data using grade of membership models. *K.K. Dey, C.J. Hsiao, M. Stephens.* University of Chicago, Chicago, IL.

Grade of membership models, also known as admixture models, topic models or Latent Dirichlet Allocation, are a generalization of cluster models that allow each sample to have memberships in multiple clusters. These models are widely used in population genetics to model admixed individuals who have ancestry from multiple populations, and in natural language processing to model documents having words from multiple topics. Here we illustrate the potential for these models to cluster samples of RNA-seq gene expression data, measured on either bulk samples or single cells. Traditional clustering methods for this problem attempt to partition samples into distinct groups of similar expression patterns. However, it seems likely that oftentimes the structure of a typical gene expression data set will be too complex to be fully captured by such a partitioning. Motivated by this, here we analyse expression data using grade of membership (GoM) models, which generalize clustering models to allow each sample to have partial membership in multiple clusters. Our model allows that each sample has a proportion, or grade of membership in each cluster. We suggest the use of a "Structure" plot visualization, which represents the estimated membership proportions of each sample as a stacked bar-chart, with bars of different colors representing different clusters. Additionally, we also discuss ways to interpret the clusters by identifying the genes that are most distinctively expressed in each cluster. We applied these methods to several example RNA-seq applications to demonstrate its utility in identifying and summarizing structure and heterogeneity. Our applications include the GTEx V6 bulk RNA seq data and a couple of single cell RNA-seq data. Applied to the GTEx project data on 51 human tissues, the approach highlights similarities among biologically-related tissues and identifies distinctively-expressed genes that recapitulate known biology. Applied to single-cell expression data from mouse preimplantation embryos, the approach highlights both discrete and continuous variation through early embryonic development stages, and highlights genes involved in a variety of relevant processes from germ cell development, through compaction and morula formation, to the formation of inner cell mass and trophoblast at the blastocyte stage. The methods are implemented in the Bioconductor package **CountClust**.

507F

Novel genomic predictor of bleeding risk in African Americans treated with warfarin. *T. De[1], W. Hernandez[1], N. Nwanze[1], E. Smithberger[1], T. O'Brien[2], M. Tuck[3], J. Duarte[4], S. Bourgeois[5], R. Kittles[6], M. Perera[1].* 1) Department of Medicine, The University of Chicago, Chicago, IL; 2) The George Washington University, Department of Pharmacology and Physiology, Washington, DC; 3) Veterans Affairs Medical Center, Washington, DC; 4) University of Florida, Gainesville, FL; 5) Queen Mary University, William Harvey Research Institute, Clinical Pharmacology - Heart Centre, London; 6) University of Arizona, Tucson, AZ.

   Hemorrhage is the major concern for patients treated with warfarin. The relative risk of major bleeding from warfarin is 44% higher in patients of African descent than of other ethnicities. The commonly studied genetic variants associated with bleeding explain substantially less variability in African Americans (AAs). Uncovering the causal variants is critical for safer use of the drug and drug selection. We conducted the first genome-wide association study in 546 AAs to identify the genetic predictors of warfarin induced bleeding requiring hospitalization at International Normalized Ratio (INR) <4, a range when most physicians do not consider warfarin reversal. To determine if clinical factors could further help to predict bleeding risk, the predictive performance of HAS-BLED (hypertension, abnormal renal/liver function, stroke, bleeding history, labile INR, elderly, drugs/alcohol) scheme was evaluated by Cox proportional hazards analysis and C-statistics. Genome-wide significant signal was observed for rs62420645 on chromosome 6 (OR=4.43, p=2.20x10^{-08}). Validation in 201 AAs confirmed this association (p=0.002, combined p=3.456x10^{-10}). A HAS-BLED score of ≥3 was associated with 2-fold risk of major bleeding (*p*=0.02). rs62420645 inclusion to HAS-BLED improved major bleeding risk prediction by correctly reclassifying 12.5% of major bleeding, thereby increasing sensitivity of prediction from 77.5% to 90% (C-indices: 0.70 vs. 0.65 for HAS-BLED+rs.62420645 vs. HAS-BLED respectively). The number needed to genotype to prevent one bleeding event was 8. *CYP2C9* star alleles showed no significant association with warfarin related bleeds. Our findings can help to predict clinically avoidable bleeding events and improve the safety of anticoagulation therapy in AAs.

508W

Genetic variation and pulmonary tuberculosis susceptibility in Guinea-Bissau and The Gambia. *G. Tavera[1], A. Tacconelli[2], G. Morris[3], M. White[8], L. Sausville[8], C. Wejse[4], C. Ciccacci[7], P. Hill[5], C. Bisseye[3], W. Scott[6], G. Sirugo[2], S. Williams[1].* 1) Case Western Reserve University, Cleveland, OH, USA; 2) Unità di Genetica Medica, Ospedale San Pietro FBF, Rome, Italy; 3) MRC Laboratories, Fajara, The Gambia, West Africa; 4) Department of Infectious Diseases, Aarhus University Hospital, Skejby, Denmark; 5) Centre for International Health, University of Otago School of Medicine, New Zealand; 6) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Florida, USA; 7) Tor Vergata University, Rome, Italy; 8) Department of Genetics, Dartmouth College, Hanover, NH 03755, USA.

   The burden of tuberculosis (TB) in West Africa is responsible for more years of healthy life lost than any other infectious disease, apart from AIDS and malaria and is second only to AIDS in mortality. Treatment for TB is poor, with few novel therapies developed in the last 50 years. We analyzed two West African samples (Guinea-Bissau: n=289 cases and 322 controls; The Gambia: n=240 cases and 248 controls) to evaluate 60 single-nucleotide polymorphisms (SNPs) in 12 candidate genes that may be implicated pulmonary TB pathogenesis. Chi squared analysis identified rs10841847 in the C-Type Lectin Domain Family 4, Member E (*CLEC4E*) gene as significant after Bonferroni correction in the Guinea Bissau population (p=0.0008644). It was not significant in the Gambian population. *CLEC4E* is involved in diverse functions, including cell adhesion, cell-cell signaling and inflammation and immune responses. Although they did not pass Bonferroni correction, other SNPs reached significance in The Gambia samples alone (rs1800896, *IL10*, p= 0.0305 and rs163790, *CTSZ*, p=0.02166) and the Guinea Bisseau samples Analysis for epistasis using MDR and ViSEN software revealed weak epistatic interactions between the assayed SNPs. Further analysis of the ethnic backgrounds of study participants may clarify the significance of these SNPs.

509T

Utility and biases of multiple whole genome methods to estimate heritability and examine genetic architecture of complex traits. *L.M. Evans[1], R. Tahmasbi[1], P. Visscher[2], J. Yang[2], S. Vrieze[1], B. Neale[3], D. Bjelland[1], T. DeCandia[1], G. Abecasis[4], S. Das[4], M. Keller[1], Haplotype Reference Consortium.* 1) Institute for Behavioral Genetics, University of Colorado, Boulder, CO; 2) Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.

Statement of Purpose: Complex trait genetic architecture has important implications for disease genetics as well as providing fundamental understanding of human evolutionary genetics. Narrow-sense heritability (h^2) is a central concept to this, yet the h^2 estimated from family studies of most complex traits cannot be explained by significant genome-wide associated genetic markers— the so-called missing h^2 problem. Recently, a number of methods—GREML, LD score regression, LDAK, treelets, etc.—have begun to examine the source of this missing h^2 using genome-wide markers in unrelated individuals. We compared these methods to understand their performance across different genetic architectures, levels of stratification, and marker type (common SNPs from arrays vs. sequence data) using simulated phenotypes in real whole-genome sequence data. Methods Used: We used a subset of ~21K whole genome sequences from the Haplotype Reference Consortium to generate phenotypes (h^2=0.5) from 1K causal variants (CVs) chosen randomly from different minor allele frequency (MAF) categories from common to very rare. We then used either the whole genome sequence data or positions found on common SNP arrays to estimate h^2 and examine genomic architecture in mixed effects models (GREML) using genetic relationship matrices (GRMs) estimated from 12 different methods across levels of stratification. Summary of Results: Using SNP array genotypes, most methods estimated common CV phenotype h^2 well, but progressively underestimated h^2 for rarer CV phenotypes. Furthermore, with array SNPs, most methods cannot accurately estimate the variance explained by rare CVs when the phenotype is derived from a mix of common and rare CVs. In contrast, using whole genome sequence data, the use of multiple, MAF-stratified GRMs in the GREML model accurately estimated total h^2 and partitioned the variance between rare and common CVs. This was true regardless of sample stratification. When using a single GRM, even if adjusted to better estimate relatedness such as with Treelet Covariance Smoothing, estimates were highly dependent on both the genetic architecture and the degree of sample stratification and could be strongly biased. Thresholded GRMs also produced highly variable biases. Finally, GREML estimates based on GRMs from identical-by-descent segments detected from SNP array data in unstructured populations appeared to provide better estimation of both rare and common causal variants.

510F

A phenome-wide gene burden analysis to identify DrugBank genes associated with patient diagnoses. *S.A. Pendergrass[1], S.S. Verma[1], N. Josyula[1], D. Hartzel[1,2], D. Lavage[1,2], S. Mukherjee[3], O. Gottesman[3], H.L. Kirchner[1,2], F.E. Dewey[3], M.D. Ritchie[1].* 1) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 2) Phenotypic Analysis and Clinical Data Core, Geisinger Health System, Danville PA; 3) Regeneron Genetics Center, Tarrytown NY.

Genetic variation such as loss of function (LOF), can mimic the effect of drugs as well as increase risk or protection for specific diseases. There are ~800 unique genes in DrugBank (http://www.drugbank.ca/) that have been identified in pathways that most drugs target. To identify DrugBank genes associated with a comprehensive series of diseases we performed a gene burden based phenome-wide association analysis (PheWAS) using low frequency (minor allele frequency < 1%) variants. We used whole exome sequencing data from 38,568 unrelated European American adults (>18 years of age) from DiscovEHR, and 541 ICD-9 based diagnoses from the Geisinger Health System electronic health record. We contrasted binning rare variants across genes two ways: all rare variants within the genes and only LOF and non-synonymous variants via different filters. We collapsed low frequency variants by gene using a software tool Biobin, and for association used a Madsen Browning style weighted burden based approach with logistic regression. All models were adjusted for the covariates of age, sex and the first 4 principal components reflecting global ancestry. Among our many highly significant results, the most significant was between low frequency variants in the gene *CASR* associated with hypercalcemia (ICD-9 275.42, p=1.43x10^{-22}), passing a Bonferroni significance threshold of 1.18x1^{-7}. This gene is a calcium receptor, and the drug cinacalcet targeting this gene is a treatment for calcium homeostasis. With *CASR,* testing for all variants regardless of function did not show a significant association, showing the association is driven mainly from LOF variants. The gene *GUCY1A2* was significantly associated with orthostatic hypotension (p=2.41x10^{-9}). Drugs that target this gene are nitric oxide and isosorbide mononitrate, and are treatments for the respiratory and cardiovascular system. We also found a significant association between *TACR1* and chronic sinusitus (p=8.10x10^{-10}), however the diagnosis is unrelated to drug treatments targeting this gene. Our current work includes contrasting our associated diagnoses for these genes with why specific drugs are prescribed including direction of effect of the associations, and pathway based analyses. Also important is identification of potential pleiotropy showing contrasts in increased risk or protection for traits for these genes. This study represents a unique gene-based rare variant PheWAS with a series of novel associations.

511W

Using maximum allele frequencies across populations greatly increases power for Mendelian gene discovery. *T. Mori[1,2], S. Uchida[2], D.A. Nickerson[3], M.J. Bamshad[1,3,4], J.X. Chong[1], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Department of Nephrology, Tokyo Medical and Dental University, Tokyo, Japan; 3) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 4) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA, USA.

Identification of a putative causal variant in exome/genome sequence data from an individual with a Mendelian condition (MC) typically proceeds by filtering for variants that follow the expected inheritance pattern and are rare (i.e. alternative allele frequency falls below a specified cutoff frequency in population controls). Data from large population control datasets, including the 1000 Genomes Projects (1KG), Exome Variant Server (EVS), and the Exome Aggregation Consortium (ExAC), are widely used, and each provides global allele frequencies (globalAAF) as well as allele frequencies within each constitutive subpopulation (i.e. Europeans, East Asians, Africans, etc.). Intuitively, causal variants for highly penetrant MCs should be rare across almost all human populations (excluding small founder populations) because the vast majority of pathogenic variants for MCs are pathogenic regardless of ancestry and should thus be rare in every population. Yet many published analyses [99% (80/81) of all Mendelian gene discoveries published in the American Journal of Human Genetics in the past year] continue to filter on globalAAF instead of the maximum population frequency across all subpopulations (maxAAF). To quantify the increased power obtained by using maxAAF instead of globalAAF, we used standard analytical parameters (e.g. filtering for variants that are non-synonymous, splice, or UTR) to analyze data from 2,504 unrelated individuals in 1KG under common Mendelian inheritance models, comparing the number of candidate genes identified using globalAAF versus maxAAF as the allele frequency filter. Under an autosomal dominant model, maxAAF decreased the median number of candidate genes per individual by nearly 40% (477 with globalAAF vs. 280 with maxAAF); under a homozygous recessive model, the median number of candidate genes was 2 vs. 1, respectively; and under a compound heterozygous model, the median number of candidate genes was 17 and 7. These differences were highly significant (Wilcoxon rank-sum test: $p < 2.2 \times 10^{-16 \text{ to } -12}$). Using maxAAF rarely eliminates variants classified as pathogenic/likely pathogenic by ClinVar and consistently results in a ~50-80% decrease in the number of candidate genes identified compared to using globalAAF. Adoption of maxAAF as the gold standard for frequency cutoff during variant filtration can substantially reduce the number of candidate genes for a given MC and should facilitate successful discovery of the underlying gene.

512T

Trajectory of new variants requiring pathogenicity assessment as potential secondary findings across 50,000 exomes in the DiscovEHR cohort. *U.L. Mirshahi, A.H. Wardeh, K. Manickam, B. Moore, T. Mirshahi, D.J. Carey, M.F. Murray, the Geisinger-Regeneron DiscovEHR Collaboration.* Weis Center for Research, Geisinger Clinic, Danville, PA.

Any project that seeks to deliver incidental or secondary genomic findings to participants in a large cohort faces pipeline bottlenecks related to the pathogenicity assessment for newly encountered variants in a gene of interest. The Geisinger GenomeFIRST return of results program intends to return secondary findings in 76 genes (G76) to a cohort of 50,000 participants who have undergone whole exome sequencing (WES) as part of the Geisinger-Regeneron DiscovEHR collaboration. In this study, the hypothesis is that the number of *new variants requiring pathogenicity assessment* (NVRPA) would decrease as the cohort increases. From our 50,000 WES, 5 groups of 10,000 samples were randomly selected. We examined all G76 non-synonymous coding and splicing variants characterized as NVRPA in each 10,000 exome increment. In a second examination, NVRPA for just one of the G76, *LDLR*, were identified. In the case of this gene, efforts to identify pathogenic/likely pathogenic (P/LP) variants had already taken place across the 50,000 (Abul-Nusn et al. abstract); therefore, the number of NVRPA called P/LP could also be ascertained.   In the G76, there were 6574, 3092, 2284, 1697, and 1380 NVRPA in the first, second, third, fourth, and fifth 10,000 exomes, respectively. For *LDLR* alone there were 152, 56, 49, 32, and 32 NVRPA in the first, second, third, fourth, and fifth 10,000 exomes, respectively. Finally for *LDLR*, there were 17, 6, 5, 5, and 2 newly encountered P/LP variants in the first, second, third, fourth, and fifth 10,000 exomes, respectively. In the first 10,000 samples, 44% of all G76 NVRPAs, 47% of all *LDLR* NVRPAs, and 49% of all P/LP *LDLR* NVRPAs were identified for assessment. Thus, most NRVPA will be encountered in the first few thousands of WES samples; the majority of NVRPA in the latter 10,000s are likely novel variants (no functional or family data). Thus, they will likely be classified as variant of unknown significance with a very small number of benign or putative loss of function variants.   In the predominantly European-American cohort, the NVRPAs are expected to become less of a bottleneck to return results to patients as we progress through the cohort. Our data suggests that the majority of unique P/LP variant calls are likely to be made early in the analysis of a large cohort. A more diverse cohort will likely see greater magnitudes of NVRPA; however, a similar pattern of decreasing NVRPA will likely emerge as the cohort grows.

513F

Haplotype-based predictors for complex trait association. *R. Brown[1], B. Pasaniuc[1,2,3].* 1) Bioinformatics IDP, UCLA, Los Angeles, CA; 2) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 3) Department of Human Genetics, UCLA, Los Angeles, CA.

A large portion of current genetic research looks for associations between phenotype and genotype. One of the most common methods for this is a standard linear regression of a phenotype on single nucleotide polymorphisms (SNPs). An assumption of this method is that underlying causal architectures disproportionally fall on the set of haplotypes defined by the presence of one of the SNP alleles. While this assumption may be valid for some loci if there is only a single causal architecture in a locus, recent work has shown that there are often multiple causal variants per locus and non-linear causal architectures. In such cases, no SNP necessarily tags the set of haplotypes that contains all the architectures positively affecting a trait. We develop a method that tags the greatest number of haplotype sets for a genetic region as opposed to the greatest number of SNPs in that region. We use the 1000 Genomes data and a list of DNase 1 hypersensitivity sites (DHSs) (Gusev, AJHG 2014) to determine all possible sets of common haplotypes (>5% frequency in Europeans) occurring in a DHS. We develop set predictors to tag the haplotype sets. Using the same number of set and SNP predictors, we show that set predictors tag (with an $r^2 > 0.8$) a greater proportion of the possible haplotype sets than do SNP predictors. Since the set predictors are tagging more of the possible haplotype combinations, it should increase power to detect loci that contain multiple causal variants or non-linear causal architectures.

514W

Interrogating gene-by-environment interactions using genetic risk scores. *N. Furlotte, V. Vacic, G. Benton, A. Auton, D. Hinds.* 23andMe Inc., Mountain View, CA.

The effect of genetic variation is dependent on the environment in which it is expressed. This interaction between genetic and environmental variation is of fundamental importance to our understanding of basic biology as well as the etiology and progression of disease. As such, the ability to quantify gene-by-environment interactions will play a key role in the application of personalized medicine. By understanding the effect of genetic risk factors in differing environmental contexts, clinicians will be better able to tailor therapeutic recommendations to specific patients. For example, the effectiveness of certain dietary and exercise recommendations may depend on underlying genetic propensity for obesity. Unfortunately, gene-by-environment interactions (GxE) remain notoriously difficult to identify, due in part to the low statistical power for discovery resulting from small effect sizes. However, recent work has demonstrated that it is possible to detect interactions by aggregating genetic effects, even when the power to detect individual GxEs is low. In this approach, GxE interactions are identified by constructing a genetic risk score (GRS) and testing for interaction between the score and an environmental condition. Here, we investigate interactions between a GRS computed using over 350 BMI-associated SNPs and different environmental conditions. Using a cohort of over 300,000 consented research participants from the 23andMe customer base, we show that there is a statistically significant interaction between the GRS for BMI (grsBMI) and exercise frequency (p = 1.32e-17) as well as fast food intake (p = 6.34e-18). To quantify the effect of the GRS on true BMI, we estimate the amount of increase in true BMI expected from a one unit increase in grsBMI and then compare estimates under different environmental conditions. The effect size among those who exercise is 0.97, increasing to 1.7 for those who do not exercise. Similarly, we find that the effect among those who frequently eat fast food is 1.87, while only 1.21 in those who consume infrequently. This discrepancy becomes more extreme depending on how you define "frequently." Additionally, we explore how environment affects heritability estimates. Taken together, our observations imply that the expected effects of diet and exercise are dependent on genetic predisposition.

515T

Trans-ethnic comparison of partitioned heritability reveals shared cell-type specific enrichment between East Asian and European genome-wide association studies. *M. Kanai[1], M. Akiyama[1], Y. Okada[1,2], M. Ikeda[3], N. Iwata[3], M. Kubo[4], Y. Kamatani[1].* 1) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan; 3) Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan; 4) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Genome-wide association studies (GWAS) have identified thousands of loci associated with human diseases and traits. There has long been a strong interest in the so-called "missing heritability", where genome-wide significant SNPs could only explain a small portion of heritability expected. Recent works in the LD score regression (LDSC) enable us to estimate the total variance explained by all or partial common SNPs, even when only summary statistics are available, which facilitates understanding the complex genetic backgrounds of human traits. With a sufficient sample size, the LDSC approach is theoretically applicable to any GWAS of any population. However, previous works were primarily based on the result of European GWAS, since currently fewer massive GWAS have been conducted for other populations, and there lacks both summary statistics and LD score references for them. Here, we report a trans-ethnic comparison of LDSC analysis (genetic heritability, genetic correlation and partitioned heritability) based on East Asian GWAS for 10 complex traits (including rheumatoid arthritis, body mass index, adult height, schizophrenia, bipolar disorder, and type 2 diabetes; mainly from the BioBank Japan project) and publically available European GWAS for the corresponding traits. We found that both overall and partitioned genetic heritability were largely consistent between populations where substantial heritability was observed. For example, we observed a similar proportion of estimated heritability (h_g^2 = 0.326 [s.e. 0.016] and 0.313 [0.014] for East Asians and Europeans, respectively) and significant correlations of categorical enrichment coefficient z-scores in 53 functional categories (r = 0.72, P = 1.22 × 10^{-9}), 10 cell-type-group specific categories (r = 0.96, P = 1.47 × 10^{-5}), and 220 cell-type specific categories (r = 0.92, P < 2.2 × 10^{-16}) between East Asian and European GWAS of adult height. Remarkably, cell-type specific enrichment showed a distinct pattern among related traits regardless of populations, which allowed us to cluster traits based on the earned profile of cell-type specificity. Our results suggested that the proportion of heritability was often shared between East Asians and Europeans, illuminating uniquely shared cell-type specificity of related traits across populations. In addition, we evaluated the trans-ethnic genetic correlation between East Asian and European GWAS, using the *popcorn* software.

516F

The effect of varying selection criteria at each stage of tiled regression on power and type I error. *J.A. Sabourin, H. Sung, A.J.M. Sorant, A.F. Wilson.* Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore MD, 21224, USA.

In staged modeling procedures such as tiled regression, which may use stepwise selection on subsets of genetic variants at hotspot-based tile, chromosome, and genome-wide stages, the choices for analysis parameters at each stage can affect the final outcome. The impact that these choices have on the SNPs selected at the final stage has not been fully explored. In this study, we investigated the effects of different critical value settings in tiled regression on both type I error and power with a simulated quantitative trait with a causal model based on the Trinity Student Study GWAS data comprised of 757533 SNP genotypes in 2232 unrelated individuals. The simulation model was based on additive effects of 5 independent causal SNPs on chromosome 4, each with a locus specific heritability of 0.01. The causal SNPs were selected such that there were no missing values in the data set and the minor allele frequency was approximately 0.05. Two-hundred replicates from this model were generated and analyzed with tiled regression, as implemented in TRAP v2.0, under two critical value settings: (1) a fixed critical value (10^{-5}) for each of the 3 stages and (2) a set of critical values beginning with a liberal value (10^{-3}) at the tile stage and becoming more conservative at each consecutive stage (10^{-4} at the chromosome stage and 10^{-5} at the genome stage) so that the critical value at the final stage matched that of setting 1. Observed type I error rates were estimated based on the 712,830 SNPs remaining after exclusion of SNPs on chromosome 4 to ensure an estimate based on SNPs that were independent of the causal SNPs. The observed type I error rate when using critical value setting 1 was 3.25x10^{-6} and when using critical value setting 2 was 5.72x10^{-6}, which, although still conservative, was closer to the nominal rate of 10^{-5}. Power was estimated as the proportion of the 5 causal SNPs selected at the final stage, averaged over replicates. Analysis using critical value setting 1 had observed power 0.386, slightly lower than that using setting 2, which was 0.395. In our simulation experiments, we found that the use of a liberal critical value at the first stage of tiled regression with successively more conservative values at later stages (setting 2) achieved a type I error rate closer to the nominal value along with increased power when compared to the setting using a fixed critical value at all stages (setting 1).

517W

Contrasting regulatory architectures of complex traits. *A. Sarkar[1,2], Y. Park[1,2], M. Kellis[1,2].* 1) Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Several computational models (GREML, PCGC regression, and LD score regression) have been developed to address the missing heritability problem, revealing common variants explain much of the heritability of complex traits. These models have provided numerous insights about the genetic architecture of complex traits, but they make a critical, non-realistic assumption which calls into question their use in partitioning heritability across regulatory annotations. Specifically, these models assume every genetic variant has an infinitesimal effect on phenotype. This assumption makes model inference feasible, by replacing an ill-specified regression problem with a simpler optimization problem. However, the resulting class of models cannot distinguish whether the observed enrichment in regulatory annotations stems from a higher density of causal variants in regulatory annotations, or simply larger effect sizes of causal variants within regulatory annotations relative to causal variants outside regulatory annotations. Additionally, this class of models does not apply to higher-resolution regulatory annotations, where they become mis-specified, leading to incorrect inferences. Here, we present an efficient genome-scale Bayesian sparse regression model that overcomes these limitations by introducing three key ideas. (1) We use recent advances in variational methods to perform approximate Bayesian inference on the high dimensional imputed dosage matrix and directly fit the underlying regression problem. (2) We formulate a novel re-parameterization of the approximation so the inference algorithm scales in the number of individuals rather than the number of variants. (3) We use biologically-motivated priors to enforce sparsity and directly model the effect of regulatory annotations on the prior probability of each variant being causal and its causal effect size. We systematically apply this model to both simulated and real genomic datasets to demonstrate its scalability and its accuracy, using: (1) idealized simulations on synthetic genotypes and annotations; (2) realistic simulations on imputed genotypes and true annotations from the Roadmap Epigenomics consortium; and (3) true annotations and imputed genotypes from the Wellcome Trust Case Control Consortium. We show that we can accurately recover non-infinitesimal genetic architectures, lending new insights into the regulatory architectures of seven complex diseases.

518T

Estimating the frequency dependence of SNP effect sizes in human complex traits. *A.P. Schoech[1], P. Loh[1,2], P.F. Palamara[1,2], L. O'Connor[1], H.K. Finucane[1,3], A.L. Price[1,2].* 1) Harvard T.H. Chan School of Public Health, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Massachusetts Institute of Technology, Cambridge, MA.

Understanding the role of rare variants is of crucial importance to elucidating the genetic basis of human complex traits and diseases. It is widely hypothesized that negative selection may cause rare variants to have larger per-allele effect sizes than common variants, but the extent of this trend is currently unknown. Here, we present a statistical method based on linear mixed models to estimate the frequency dependence of SNP effect sizes in sequenced samples. We estimate the parameter α in a model in which per-allele effect sizes have variance $[p(1-p)]^{\alpha}$, where p is allele frequency and negative values of α imply larger effect sizes for rare variants. We employ a profile likelihood approach, using restricted maximum likelihood to estimate the phenotypic variance explained by the genotype data for a given α. This computation is repeated for different α values over a biologically plausible range, choosing the value that maximizes the overall likelihood as the estimate of α. Our simulations show that this method produces unbiased estimates under a wide range of genetic architectures. We applied the method to 3,567 samples from the UK10K cohort with whole-genome sequencing data and measurements for 13 quantitative phenotypes: height, body mass index, waist to hip ratio, heart rate, diastolic blood pressure, FVC:FEV1 ratio, hemoglobin levels, homocysteine, fasting glucose, adiponectin, HDL, LDL and total cholesterol. We included SNPs down to a minor allele frequency of 0.03% in our analyses. We obtained an estimate of $\alpha = -0.76 \pm 0.28$ across phenotypes, which is significantly different from zero, confirming that rare variants have larger per-allele effect sizes than common variants in this data set and implicating significant negative selection on the SNPs affecting these traits. To our knowledge, this is the first study quantifying the frequency dependence of allelic effects in whole-genome sequencing data. Application of the method to larger data sets will enable trait-specific estimates of α, providing a quantitative comparison of the impact of negative selection on different traits and informing the design of genetic association studies.

519F

Identity-by-descent as a proxy rare variant association test in large-scale genome-wide association studies. *J.N. Torres[1], J.R. Homburger[1], G.L. Wojcik[1], G. Belbin[2], S. Prabhu[1], C.D. Bustamante[1], E.E. Kenny[2], C.R Gignoux[1].* 1) Center for Computational, Evolutionary Human Genomics, Stanford, Stanford, CA; 2) Charles Bronfman Institute for Personalized Medicine, New York, NY.

In the second decade of genome-wide association studies, researchers are increasingly turning to sequencing as a strategy to identify associated variants with rare minor allele frequencies (MAF<1%), motivated by the prediction that rare variants are often found to have larger effect sizes than common variants, and contribute to the heritability often missed by top hits identified in GWAS. While sequencing technologies are developing rapidly, genotyping arrays continue to provide a low-cost snapshot of genomic variation. These will not capture the entire frequency spectrum in the study as rare variants may not be in linkage disequilibrium with the array's tag SNPs. However, in this context, leveraging Identity-by-Descent in a GWAS framework (IBDGWAS) may provide an alternative to capturing patterns of trait-associated rare variants. Population genetic theory and recent empirical studies have shown that rare variants tend to have arisen more recently in human history and thus should be correlated with IBD tracts, and therefore detectable with genotyping arrays. The power to identify IBD segments for association testing scales quadratically so large sample sizes, as is becoming the standard with large-scale consortia and biobank studies, provide an emergent opportunity to uncover genomic variation. Therefore examining rare variant associations via IBDGWAS may provide a low-cost alternative to sequencing. To examine the relative power of IBDGWAS versus sequencing at capturing rare variant association, we employed both a coalescent-based framework for large-scale case-control simulation followed by examination in several biobank datasets. We compared IBDGWAS to exome and whole-genome sequencing assuming a fixed budget and evaluated the tradeoff between sequencing's resolution and the large boost in sample size possible with IBDGWAS. We first evaluate the effect of a population bottleneck on levels of IBD detectable in mega-scale datasets. We describe demographically realistic scenarios in which IBDGWAS outperforms sequencing strategies, in particular, whole genome sequencing given the relative expense. Based on these findings, we can integrate the findings from both IBDWAS and GWAS to learn about the relative contributions of rare and common variants to genomic architecture, as well as the portability of findings across populations with diverse demographic histories..

520W

Testing family association with rare and common variants via wavelet transformation. *H. Wang, Y. Lin, D. Chen, S. Wu, C. Chang, C. Fann.* Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Many common diseases are known to cluster in pedigrees which have become a rich resource for detecting association between rare variants and complex traits. In recent years, many researchers have modified existing population-based methods for detecting family-based rare variant association. These tests including collapsing/burden test and variance-component tests focus on identifying a group of rare variants in a region associated with a disease, thereby effectively reducing the degrees of freedom as compared with individual rare variant test. In addition, both common and rare variants may be the causal variants in a region and therefore should be considered simultaneously. It is inevitable that the power may be loss if neutral variants are incorporated into test statistics. With the aims of improving the power of rare variants association in family design and reducing the noise caused by neutral variants, we extended the wavelet-based test to identify both common and rare causal variants within a region. The wavelet transformation which can suppress the noise data has been applied to rare variants association in case-control data but has not been used in family design. In our study, pedigree sequence datasets were simulated. We evaluated type I error rates and compared the power of family-based wavelet test (famwave) with two other family-based association tests, family-based weighted sum test (famWSS) and family-based SKAT (famSKAT). The results demonstrated that type I error rates under different scenarios were well-controlled and had a comparable or better power (0.4~0.98) than the other tests. In order to illustrate the feasibility and performance of the proposed methodology in realistic setting, the Genetic Analysis Workshop 17 datasets were utilized. The qualitative trait (affected/ unaffected) was used to test with 228 variants in 15 genes. The percentage of the number of significant results among the 200 replicates was used to evaluate the performance of these methods. For all 15 candidate genes, the results indicated that our proposed method detected the causal genes more frequently than the other tests, for example a large number variants in gene PIK3C2B (71 variants).

521T

Improving imputation accuracy by inferring causal variants in genetic studies. *Y. Wu[1], F. Hormozdiari[1], J. Wha J Joo[1], E. Eskin[1,2].* 1) Computer Science, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA.

Genotype imputation has been widely utilized for two reasons in the analysis of Genome-Wide Association Studies (GWAS). One reason is to increase the power for association studies when the causal SNPs are not collected in the GWAS. The second reason is to aid in the interpretation of a GWAS result by predicting the association statistics at untyped variants. In this project, we show that the prediction of association statistics at untyped variants which do in fact have an influence on the trait is overly conservative. Current imputation methods assume that none of the variants in a region (locus consists of multiple variants) affect the trait which is often inconsistent with the observed data. In this paper, we propose a new method, CAUSAL-Imp, which can impute the association statistics at untyped variants taking into account that some of the variants in the region may affect the trait. Our method builds on recent methods that impute the marginal statistics for GWAS utilizing the fact that the marginal statistics follows a multivariate normal distribution. We utilize both simulated and real data sets to assess the performance of our method. We show that traditional imputation approaches underestimate the association statistics for variants involved in the trait while our approach provides less biased estimates of these association statistics.

522F

Adjusting missing confounders in epigenome-wide association studies. *C. Wu[1], J. Pankow[2], E. Demerath[2], J. Bressler[3], M. Grove[3], M. Fornage[3], W. Guan[1].* 1) Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA; 2) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA; 3) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA.

DNA methylation is a commonly studied epigenetic mechanism that may be involved in the development of common diseases. Unlike inherited changes in genetic sequence, variation in site-specific methylation varies by tissue, developmental stage, and may be impacted by aging and exposure to environmental factors. Failing to control for these non-genetic factors in epigenome-wide association studies (EWASs) may lead to false discoveries and loss of statistical power. These factors, if available, can be directly included in a regression model, or can be inferred through decomposition of the methylation matrix, e.g., using principal component analysis (PCA). Here we introduce a novel statistical method based on standard and supervised principal component analysis (SPCA) to correct for unobserved confounders in EWASs. Compared to standard PCA, SPCA will assist to capture confounders that only contribute to a small proportion of CpG sites among the genome-wide measures. Through simulation studies, we demonstrate that our proposed method can maintain acceptable false positive rate and improve statistical power of epigenetic association test. We also compare our results to existing methods, such as ReFACTor and FaST-LMM-EWASher, and apply our method to an EWAS of smoking in the atherosclerosis risk in communities (ARIC) study. We propose that our method will help to control for potential confounding and increase statistical power of epigenetic association studies.

523W

Detection of recessive selection identifies non-additive components of complex disease. *D.M. Jordan[1], D.J. Balick[2], I.A. Adzhubei[2], S.R. Sunyaev[2], R. Do[1].* 1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mt. Sinai, New York, NY; 2) Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Methods to identify genes and variants for complex diseases generally assume that effects of alleles are additive, meaning that a homozygous genotype confers twice the risk of a heterozygous genotype. Indeed, the majority of genome-wide association studies (GWAS) for complex diseases have assumed additive models of risk, but the few studies that have tested recessive models have discovered recessive associations that were not detectable using additive models[1]. Genes under recessive selection exhibit detectably different population dynamics[2]. Leveraging this property, we have developed a novel method to quantify the strength and recessivity of selection of all protein-coding genes across the human genome by comparing European population sequencing data from the Exome Aggregation Consortium (ExAC) (N=35,000) with simulated evolutionary histories for both additive and recessive alleles. This method could inform model choice by identifying genes and pathways likely to be under recessive selection. We find a variety of biologically meaningful categories enriched in the predicted recessive class, including glycoproteins (Benjamini P=6.3x10^{-13}), immunoglobulin domains (P=0.023), and inflammatory response (P=0.0052). The enrichment for inflammatory genes in particular suggests that many complex diseases with inflammatory and autoimmune components may be under recessive selection, such as Crohn's disease and rheumatoid arthritis (RA). In the case of Crohn's disease, we find that genes with large and well-validated effects are predicted to be under recessive selection, while genes implicated by GWAS studies have no enrichment for recessive selection. Similarly, RA loci discovered by GWAS show no enrichment for recessive selection, despite the fact that RA is known to involve inflammatory and immune pathways that are highly enriched for recessive selection according to our method. These examples highlight the potential usefulness of our catalog of recessive selection as a tool for variant and gene prioritization, and more broadly, for gaining biological insight into complex disease. 1. Nikpay M, Goel A, Won H-H, et al. (2015) A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genet* 47, 1121–1130. doi:10.1038/ng.3396 2. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. *PLoS Genet* 11(8): e1005436. doi:10.1371/journal.pgen.1005436.

524T

A new sparse Bayesian method for multi-trait genome wide association studies. *K. Sharp[1], V. Iotchkova[2], A. Dahl[3], J. Marchini[1,3].* 1) Statistics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Welcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 3) Welcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Genetic association studies of complex traits have concentrated primarily on testing for association between a SNP and each trait separately. However, when measurements of multiple, correlated phenotypes are available, this approach ignores useful information. In such cases, statistical power can be increased by testing all traits simultaneously. Existing multi-trait approaches look for evidence of association by comparing the fit of just two models: a null model of no association at all and a saturated model incorporating non-zero effects of a SNP on all traits. However, we rarely expect such a dichotomy especially as the number of traits increases. We have developed a Bayesian model designed to detect association when correlations exist between high-dimensional phenotypes and between individuals due to relatedness. We impose a sparsity assumption through the use of a 'spike and slab' prior on the genetic effect sizes. This allows us to learn the subset of traits that are associated with each SNP or genetic region. In this way our method is able to perform model inference (association testing) and model selection simultaneously. To allow for relatedness and population structure, our method generalizes single-trait, linear mixed model analysis to high-dimensional traits. We use MCMC to fit the model at SNPs showing some marginal evidence of an effect. On simulated datasets of unrelated individuals, when the true signal is sparse, our approach has significantly more power to detect association than simpler approaches based on fitting a saturated model (SM) or single phenotype tests (minP). For 10 simulated phenotypes, of which only 1 had a true association, we observed a mean increase in power to detect association of 25% and 32% over SM and minP respectively (fpr=0.01). We applied our method to 15 glycomics phenotypes for individuals from 3 isolated populations. Two previous meta-analyses discovered associations at 19p13.3, 14q23.3 and 12q24.31. The second, larger, study also identified a QTL at 2q21. Our method is able to uncover strong evidence of association at these SNPs (log10 BF > 9.5) and produces a stronger signal at the 2q21 locus in the Korcula and Vis cohorts than p-value based approaches. We also find strong evidence of association with additional phenotypes at each locus. For example, the larger meta-analysis reported associations with 4 phenotypes at 19p13.3, and with only 1 at 2q21. We find associations with 7 and 3 phenotypes respectively.

525F

The impact of genotype imputation on the statistical power of rare variant association studies. *D.M.H. Tong, R.D. Hernandez.* University of California San Francisco, San Francisco, CA.

Genome-wide association studies have discovered a large number of associations for many complex human diseases. Despite these successes, much of the heritability of these traits remains unexplained. Low frequency and rare variants are a common explanation, and genotype imputation has been widely used to introduce low frequency variants into GWAS to boost power. In this study, we evaluated the power of gene-based rare variant association methods on simulated whole genome sequences against simulated imputed sequences. We simulated DNA sequences for 20,000 individuals under a demographic model representing three different ancestries (African, European, and Asian) with evolutionarily conserved and neutral sequence variation. Phenotypes for each individual were simulated under a range of models encompassing a diverse set of genetic architectures. Each individual genome was then downsampled to match common genotyping arrays, and imputed using minimac and IMPUTE2 with a subset of held-out individuals as a reference panel. The statistical power of a diverse set of rare variant association methods was evaluated and compared on both the entire set of simulated sequences as well as the imputed sequences.

526W

A unified approach to estimating population structure and relatedness. *B.S. Weir[1], J. Goudet[2].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Ecology and Evolution, UNIL-Sorge, Switzerland.

Evolutionary studies, association mapping and forensic identification rely on estimates of population structure or relatedness. Estimates of quantities depending on allelic identity by descent cannot be estimated in an absolute sense, but only relative to a reference set of alleles. The usual Fst reference set is alleles among independent populations. The EIGENSTRAT kinship estimates reference set is for other unrelated pairs of individuals in a sample. KING-robust uses non-ibd alleles within individuals as a reference. However, migration leads to dependent populations, all pairs of individuals in a study may have low kinship levels and consequent inbreeding. We have therefore re-cast population structure, relatedness and inbreeding estimation to make explicit what is being estimated by standard estimates, and to provide simple alternatives. We use allelic matching, or ibs. If a series of populations i have reference allele sample frequencies P_i then within-population matching proportions are $P_i^2+(1-P_i)^2$, with average M(W). For populations i,j the matching proportions $P_iP_j+(1-P_i)(1-P_j)$ have average M(B). The corresponding ibd parameters are theta(i), theta(ij) with averages theta(W), theta(B). We avoid confounding by the number of populations by using M(B) instead of the usual total heterozygosity. Our population-specific estimates [M(i)-M(B)])/[1-M(B)] are unbiased for [theta(i)-theta(B)]/[1-theta(B)]. In effect, standard methods set theta(B)=0. The new estimates are good indicators of natural selection, they are sensitive to population divergence times and migration rates for rare variants, and they give preliminary indicators of association when the populations are cases and controls. The usual Fst quantity is [theta(W)-theta(B)]/[1-theta(B)], and for pairs of populations drifting apart this is proportional to the time since divergence from an ancestral population. For population sample sizes of 1, the matching proportions are $M(i)= (1-X_i)^2$ within individual i, or $M(ii)=[1+(1-X_i)^2]/2$ for an individual with itself, and $M(ij)=[1+(1-X_i)(1-X_j)]/2$ for individuals i,j where X_i is the reference allele dosage for i. Now [M(ij)-M(B)]/[1-M(B)] provides estimates of [theta(ij)-theta(B)]/[1-theta(B) where theta(ij) is the coancestry for i,j. When i=j, $theta(ii)=(1+F_i)/2$ where F_i is the inbreeding coefficient of individual i. The estimates of theta(ij) for distinct i,j sum to zero by construction, as do large-sample expected values of EIGENSTRAT estimates.

527T

Prediction accuracy of combined genetic and non-genetic risk scores. *F. Dudbridge.* Non-communicable disease epidemiology, London Sch of Hygiene and Tropical Medicine, London, United Kingdom.

The substantial heritability of most complex diseases suggests that genotypic data could be a useful predictor of individual disease risk. Such risk prediction could identify individuals who would benefit from preventive medication, such as statins, and could improve the efficiency of screening programs for early detection of diseases such as breast cancer. To date the performance of genetic risk scores has fallen short of the potential implied by heritability, but this shortfall can be explained by inadequate sample sizes used to estimate effect sizes in highly polygenic models. However, as sample sizes approach hundreds of thousands, more accurate genetic risk prediction is becoming feasible. When there are existing risk predictors based on environment or lifestyle, the main questions are to what extent can these predictors be improved by adding genetic information, and what is the ultimate potential of combined genetic and non-genetic risk scores? A key issue is the correlation between genetic and non-genetic risk, which may occur when the non-genetic risk factors mediate the genetic risk, or when one of the risk factors is family history. Here I extend previous work on the predictive accuracy of polygenic risk scores to allow for a non-genetic score that may be correlated with the polygenic score. I derive measures of predictive accuracy including liability R^2, area under the ROC curve and net reclassification improvement, as functions of the sample size used to discover the polygenic score, chip heritabilities of the disease of interest and of the non-genetic score, and genetic correlation between disease and non-genetic score. I consider simple addition of the two scores and a weighted sum that adjusts for their correlation. Using current examples from cardiovascular disease and breast cancer, I show that improvements in AUC are generally small but reasonably large NRI could be obtained with current sample sizes. Correlation between genetic and non-genetic scores turns out to have only minor effects on numerical results. Since existing risk predictors tend to explain less variation than the heritability, the long term prospect might be that polygenic scores become the default predictor and the burden is on epidemiologists to find sufficiently informative additional risk factors.

528F

Gene-based segregation method for identifying rare variants in family-based sequencing studies. *D. Qiao[1], C. Lange[3], N.M. Laird[3], S. Won[4], B. Hobbs[1], S.M. Lutz[5], J.D. Crapo[6], T.H. Beaty[7], E.K. Silverman[1,2], M.H. Cho[1,2].* 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 4) Department of Public Health Science, Seoul National University, Seoul, Republic of Korea; 5) Department of Biostatistics, University of Colorado, Aurora, CO, USA; 6) Department of Medicine, National Jewish Health, Denver, CO, USA; 7) John Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

Background: Family-based whole-exome sequencing data have been used to identify rare coding variants of large effect and complete or near-complete penetrance associated with Mendelian diseases or complex diseases with Mendelian subtypes. A filtering approach based on variant novelty, functionality, and segregation with phenotype within families has led to successful discoveries for some diseases; however formal statistical approaches are limited. **Methods**: We propose a GEne-based SEgregation Test (GESE), constructed using the marginal probabilities of segregating events under the null hypothesis of Mendelian transmission. This test takes into account different degrees of relatedness in the families and genetic background variation in each gene. In addition, a weighted version of this test allows incorporating additional subject phenotypes and functional annotations to improve the power for identifying the causal genes. **Results:** We showed via simulations that the GESE and weighted GESE tests maintain appropriate type I error rate, and have greater power than several commonly used association methods. We applied the methods to whole-exome sequencing data of 347 subjects from 49 extended pedigrees with severe, early-onset chronic obstructive pulmonary disease (COPD) in the Boston Early-Onset COPD Study (BEOCOPD) and proposed several candidate genes for future validation. **Conclusions**: Our proposed methods show great promise in identifying rare coding variants of large effect and high penetrance for family-based sequencing data. The proposed tests are implemented in an R package that is publicly available. .

529W

Metrics to assess imputation quality in the presence of complex relatedness. *C. McHugh[1], B.K. Cornes[1], B.J. Chen[1], M.C. Zody[1], D.E. Housman[2], N.S. Wexler[3], R.B. Darnell[1].* 1) New York Genome Center, New York, NY; 2) Massachusetts Institute of Technology, Department of Biology, Cambridge, MA; 3) Columbia University Medical Center, New York, NY.

With the decreasing cost of genotyping technology, researchers can afford to generate SNP chip array genetic data for a large number of samples, resulting in a set of relatively common variants in all study individuals. On the other hand, obtaining rare variant calls for all samples using whole genome sequencing (WGS) is more expensive. Oftentimes a cost-effective approach is used where only a small number of samples are submitted for WGS, then imputation is performed in all samples using the directly sequenced samples as references. Currently, some imputation methods specifically take into account pedigree structure while others are pedigree-agnostic. In family-based studies, we may sequence only a select number of relatives, but want to choose them intelligently in order to provide the most information for imputation quality and accuracy. A few tools are available that offer algorithms to select samples for WGS but they have a large computational burden, especially for complex pedigrees. Furthermore, they do not offer comprehensive assessment of imputation quality in the context of a pedigree. Here, we develop metrics to assess imputation quality and accuracy in related individuals, by sample and by variant. We perform imputation using various software packages, including IMPUTE2 and MaCH, in a large, complex pedigree where we have SNP array data on all subjects but have only directly sequenced a subset of the study samples. From sample-based metrics, we can determine expected imputation accuracy, taking into account the position of the individual within the pedigree. One sample-based metric calculates the concordance between imputed and sequenced genotype calls. Other metrics are devised to analyze characteristics of discordant sites. We assess the sample-based metrics for all individuals stratified by pedigree characteristics such as how many relatives they have. Quality is also evaluated by variant, taking into account various characteristics such as minor allele frequency and call rate. With this comprehensive assessment of imputation output, we gain insight into why some samples and variants perform better than others in the presence of complex relatedness. In the future, such metrics may be used to select the most informative samples for WGS, an important design decision for many family-based studies.

530T

Linkage analysis and regional heritability vs GWAS in family-based cohorts. *R. Nagy[1], P. Navarro[1], C. Hayward[1], J.F. Wilson[1,2], C.S. Haley[1,3], V. Vitart[1].* 1) Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 2) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom; 3) Roslin Institute and Royal (Dick) School of Veterinary Studies, Edinburgh, United Kingdom.

Alternatives to traditional GWAS are needed to tackle complex associations of inherited haplotypes and allelic heterogeneity. Here, we tested how using family-based data and haplotype-based methods can provide additional insights into genetic analysis. Analyses were performed in four population isolate cohorts of Croatian or Scottish heritage (CROATIA-Vis, CROATIA-Korcula, ORCADES and Viking Health Study Shetland) and Generation Scotland: The Scottish Family Health Study (GS:SFHS), a family-based biobank from the Scottish general population. These cohorts all have extensive pedigree information, making them ideal for linkage analysis. Some of these cohorts are also population isolates, which means individuals share longer haplotypes derived from common ancestors. We performed variance component-based linkage analyses, regional heritability and single SNP GWAS in these cohorts, for measured quantitative traits of public health importance. Using linkage analysis and regional heritability, we identified regions that were missed by GWAS (in our own analyses and in published large consortium GWAS meta-analysis). Some regions harbour genes whose expression patterns and functions make them good candidates for follow-up. We will present two such examples resulting from scans of alanine transaminase, a liver enzyme associated with metabolic disease and intra-ocular eye pressure, a trait associated with glaucoma. The studied cohorts range in size from 1000 to 20000 individuals with average family sizes ranging from 3 to 43 individuals. Due to their family structures, both the 20000-individual Generation Scotland cohort (average of 6 people per family) and the 2000-individual ORCADES cohort (average of 43 people per family) have similar power to detect linkage, showing that simply increasing sample size is not always the best solution to increasing the power of these analyses.

531F

TreeLMM: Modelling heterogeneity of genetic effects. *R. Moore[1,2], F. Casale[2], I. Barroso[1], O. Stegle[2].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom.

Genetic effects can vary between individuals. For example, the effect of a genetic variant may depend on environmental exposure such as smoking status, life style covariates, etc. However, widely used approaches to analyse genotype-phenotype associations commonly assume that genetic effects are identical across all samples within a cohort. One approach to address these differences in genetic effects is to divide the samples into distinct groups and then either perform analyses for individual groups followed by meta analysis or conduct multi-trait modelling, for example using linear mixed models. Such approaches can result in a prohibitive numbers of groups, with small sample sizes, that are not fully independent of one another. Hence, both independent analyses and existing multi-trait models do not fully leverage the structure in the data. Here, we propose TreeLMM, an approach that explicitly accounts for heterogeneity in effect sizes and group structure. This is achieved by defining a prior on effect sizes that encodes the similarity between the different groups of individuals; these groups can contain just one individual. The method uses score tests ensuring that it is efficient and in principle enables the analysis of thousands of samples and arbitrary numbers of groups. To illustrate our approach, we have applied the method to perform eQTL mapping in 44 tissues made available through the GTEx project. We have leveraged expression profiles to estimate similarities and mapped *cis*-eQTLs across the cohort of 7,051 samples. Compared to independent modelling and methods that ignore heterogeneity in effect size, we identify a 2-fold and 4-fold increase respectively in the number of independent variants that are associated with gene expression changes. Additionally, our model allows for estimating posterior probabilities of the effect of genetic variants in individual tissues. This approach gives a powerful handle on dissecting tissue-specific effects as well as secondary association signals.

532W

Missing genetic interactions result in prediction bias at the extremes of risk for time-to-event models. *D. Soave[1,2], L.J. Strug[2,1].* 1) University of Toronto, Toronto, Ontario, Canada; 2) The Hospital for Sick Children, Toronto, Ontario, Canada.

Background: Risk prediction models can translate genetic association findings to inform clinical decision-making. Most models are evaluated on their ability to discriminate, and the calibration of risk-prediction models is largely overlooked in applications. Models that demonstrate good discrimination in training datasets, if not properly calibrated to produce unbiased estimates of risk, will generally perform poorly in new patient populations. Poorly calibrated models arise due to missing covariates, such as genetic interactions that in many instances are unknown or not measured. Methods: We demonstrate that models omitting interaction effects can lead to increased bias in predicted risk for patients at the tails of the risk distribution; i.e those patients who are most likely to be affected by clinical decision making. We propose a new calibration test for Cox risk-prediction models for time-to-event data that leverages additional power from bias in risk estimates at the extremes of the risk distribution. Our test aggregates martingale residuals for subjects from extreme high and low risk groups with a test statistic maximum chosen by varying which risk groups are included in the extremes. We show how to estimate the empirical significance of our test statistic by simulating from a Gaussian distribution using the covariance matrix for the grouped sums of martingale residuals. Results: A simulation study shows our new test maintains control of type 1 error and demonstrates improved power over conventional goodness-of-fit tests when risk prediction is poorest at the tails of the risk distribution. We apply our method in the development of a prediction model for onset of cystic fibrosis-related diabetes. Conclusions: Our study highlights the importance of assessing both calibration and discrimination in the development of predictive time-to-event models, and we show our new test to be a useful, complimentary tool in the assessment of risk model calibration.

533T

An efficient gene-gene interaction test for rare variants in case-control studies. *R.H. Chung, C.Y. Kang.* National Health Research Institutes, Zhunan, Taiwan.

Gene-gene interaction can play an important role in complex disease etiology. With the advancement of next-generation sequencing, many rare variants have been discovered. Current gene-gene interaction tests for disease studies are mainly developed for common variants. A few gene-gene interaction tests accounting for rare variants are available. However, they tend to be computationally intensive and may not be applicable to genome-wide gene-gene interaction analysis. Therefore, an efficient gene-gene interaction test for rare variants in disease studies is desirable. We developed a powerful and efficient gene-gene interaction test based on the goodness-of-fit (GOF) test statistic (iGOF). For a pair of SNPs, two 3 by 3 genotype tables are created with the observed genotype counts in cases and controls, respectively. Based on the 3 by 3 tables, the GOF statistic is calculated as the sum of differences in counts between the observed and expected genotype counts, each squared and divided by the expectation, under the null hypothesis of no interaction. The observed and expected counts from cases and controls both contribute to the statistic. For a pair of genes, the algorithm includes several steps: (1) assign SNPs with minor allele frequencies less than 5% to the genes; (2) exclude SNP pairs with minor allele counts in cases less than those in controls; (3) calculate the GOF statistics for the remaining pairs of SNPs between the two genes; (4) calculate the iGOF statistic, which is the sum of the GOF statistics; (5) permute the case and control status and calculate the p-value. We used simulations to demonstrate that the iGOF test has correct type I error rates under various scenarios. We also used simulations to compare the power of iGOF with other interaction tests considering rare variants, including SPAr, KBAC, and SKAT. The results suggest that iGOF has more power than other tests under most of the scenarios. To improve the performance of the iGOF test, the iGOF statistic is approximated by a chi-square distribution; permutations will only be performed if the p-value from the chi-square distribution is less than a pre-specified threshold. Moreover, permutations are performed based on the genotype tables stored in memory, which significantly reduces the run time. In conclusion, we developed a powerful and efficient gene-gene interaction test for rare variants in case-control studies using sequencing data.

534F

A novel and efficient study design to test parent-of-origin effects in family trios. *R. Feng, X. Yu.* Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, PA.

Increasing evidence has shown that certain genes affect various prenatal, neonatal, and pediatric diseases differently depending on their parental origins. Parent-of-origins have been accommodated in genetic models to improve the power of detecting disease-associated genes and partially recover missing heritability. In many genetic studies, children's DNA samples were collected for initial testing and their parents contributed their own DNA samples for further investigations. Nowadays, next-generation sequencing (NGS) is often done on children's DNA samples to search for common and novel variants responsible for diseases. However, the cost of NGS of their parents is generally not affordable to all investigators, making the assessment of parent-of-origin effects impossible. Motivated by the reality, we proposed a new study design collecting children's NGS data, in combination with the parental genotypes obtained from the traditional genotyping array, at a fractional cost of sequencing all family members. We developed a powerful and efficient likelihood-based method, which incorporates both sequence alignments and linkage disequilibrium into inferring the parental origins of children's alleles and estimating their effect on the disease phenotype. We evaluated the performance of our method and compare with existing genotype-based methods using simulations. With a common read length of 100 base pairs, our method showed improved power over a haplotype-based method using SNP data only. When either read depth or pair-end length increased, our method achieved optimal power closer to an ideal test that assumed known parental origins of children's alleles. With its flexibility to accommodate missing genotypes, our method is also helpful to recover the power lost due to missing single parent's DNA.

535W

Heterozygosity rate is not a good quality check metric in association studies of admixed populations. *A. Gaye[1], S. Davis[1], G. Gibbons[2].* 1) NHGRI, NIH, Bethesda, MD., USA; 2) NHLBI, NIH, Bethesda, MD., USA.

BackgroundHeterozygosity rate and genotype failure rate are routinely used to identify samples with low DNA quality at the data quality control stage of genetic association studies. Excessive heterozygosity rate may indicate sample contamination whilst a reduced heterozygosity rate could indicate inbreeding. Samples with heterozygosity > 2-3 standard deviations from the mean heterozygosity are routinely excluded from genetic case-control studies. We show that for admixed populations this leads to the exclusion of good quality samples because outlying heterozygosity might just be due to particular ancestry composition.MethodsWe conducted admixture and principal component analyses of two African American datasets together with data from the 1000 Genome Project. We first calculated the observed heterozygosity rate per individual and identified heterozygosity outliers. We the used the algorithm ADMIXTURE, a model-based approach, to estimate the fraction of global ancestry from each of the populations (admixed, African and European populations). We subsequently plotted the PCA results and identified heterozygosity outliers whose ancestry composition causes them to cluster away from the admixed group.ResultsIn this work we confirmed that heterozygosity rate is valid a QC metric to identify sample that should be excluded from a genetic association study. However, we have also showed that heterozygosity rate can be misleading if used in studies of admixed groups because samples that have a genetic diversity relatively different from that of the majority of the study participants are flagged as outliers to exclude. Therefore we will recommend carefully checking the ancestry composition of samples away from the mean heterozygosity at the QC stage to ensure valid samples are not incorrectly discarded particularly in settings where sample size is limited and where each exclusion has a greater impact on power than in large studies.

536T

Single-marker tests using a new joint modeling approach of multiple traits outperform current single- and multi-marker tests in rare variant association studies. *S. Konigorski[1,2], Y.E. Yilmaz[2,3,4], T. Pischon[1,5,6].* 1) Department of Molecular Epidemiology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany; 2) Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, Canada; 3) Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada; 4) Discipline of Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada; 5) Charité Universitätsmedizin Berlin, Germany; 6) DZHK (German Center for Cardiovascular Research) partner site Berlin, Germany.

Investigating the role of rare variants in complex phenotypes has received increasing interest in recent years. Yet, the low statistical power of rare variant association tests remains one of the main challenges and it is an open question which statistical methods are the most powerful. Much focus has been on multi-marker tests, which combine the rare variants' information in a given region and test the association of the region with the phenotype. There is a common belief that they are more powerful than single-marker tests, which test the association of each marker separately. However, different hypotheses are tested in the two approaches, which has to be considered for any comparison and conclusion as to which method has highest power, but is lacking from the literature. In this study, based on the underlying hypotheses considered for each type of test, we describe a framework to compare their empirical type I error and power and conduct extensive simulation studies. In addition, we propose a new single-marker test obtained through a copula-based joint analysis of multiple traits of a phenotype (C-JAMP) to increase the power of standard single-marker tests. The results demonstrate that the power to identify causal loci depends largely upon whether the research interest is limited to conclusions on the gene/region level or seeks to reveal causal variants, and that previous comparisons have been misleading. We show that in a number of realistic situations, standard single-marker tests lead to more powerful association tests than multi-marker tests when the same genetic locus is investigated. Furthermore, C-JAMP allows increasing the power for identifying functional loci significantly, and is able to identify 2 or 3 times more causal loci compared to both standard single-marker tests and popular multi-marker tests. It uses the information contained in the association between different traits and allows building more biologically meaningful modeling approaches. Hence, we suggest modeling multiple dependent traits and performing single-marker tests instead of aggregating rare variants in a region and performing multi-marker tests.

537F

Increasing the probability that a finding is genuine for next-generation sequencing studies. *W. Lin, J. Li.* Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan.

With the advancement of next-generation sequencing technology, searching for rare causal variants (minor allele frequency < 1%) gradually becomes possible. However, the statistical power for detecting individual rare causal variants is very low due to the extreme sparsity. Most analysis approaches strengthen the association signals by aggregating the information of multiple rare variants in a gene/region. The BURDEN test and the sequence kernel association test (SKAT) are two commonly used methods for rare-variant association analyses. If we find an association for a gene/region with a phenotype, the next important step is to pinpoint individual rare causal variants from among a large number of variants within the region. To this end, statistical approaches such as the BE (backward elimination) procedure and the ADA method ('adaptive combination of *P*-values method'), have been proposed. It has been shown that the signal-to-noise ratio of variants identified by ADA is larger than that of variants identified by BE. In this study, we develop an ADAQ approach ('adaptive combination of *Q*-values method') to further increase the probability that a finding is genuine. With synonymous / non-synonymous annotations for variants, we first allocate all variants into a non-synonymous group and a synonymous group, and transform two groups of per-site *P*-values into Benjamini-Hochberg *Q*-values, respectively. We then remove the variants with *Q*-values larger than the optimal *Q*-value truncation threshold, which are more likely to be neutral. The optimal *Q*-value truncation threshold is searched through our ADAQ algorithm. Comprehensive simulations have shown that ADAQ produces an even larger signal-to-noise ratio than ADA. Moreover, we applied ADAQ to the Genetic Analysis Workshop 17 (GAW 17) data sets. It controls the number of false positives more effectively and generates a larger signal-to-noise ratio than ADA. Therefore, we recommend using ADAQ to pinpoint individual rare causal variants, when synonymous / non-synonymous annotations for variants are available.

538W

Polygenic scores using summary statistics via penalized regression. *T.S.H. Mak[1], R.M. Porsch[2], S.W. Choi[2], X. Zhou[2], P.C. Sham[1,2,3].* 1) Centre for Genomic Sciences, University of Hong Kong, Hong Kong; 2) Department of Psychiatry, University of Hong Kong, Hong Kong; 3) State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong.

Polygenic scores (PGS) summarize the genetic contribution of a person's genotype to a disease or phenotype. They are useful in a wide variety of analyses of genetic data. Many possible ways of calculating polygenic scores have been proposed, and recently there is much interest in methods that incorporate information available in published summary statistics. As there is no inherent information on linkage disequilibrium (LD) in summary statistics, a pertinent question is whether we can make use of LD information available elsewhere to supplement such analyses. To answer this question we propose a method for constructing PGS using summary statistics and a reference panel in a penalized regression framework, which we call lassosum. We also propose a general method for choosing the value of the tuning parameter in the absence of validation data. We then assessed the performance of lassosum, with a particular emphasis on whether the use of a ``wrong'' reference panel can negatively impact the predictive power of the PGS, and whether it is beneficial to combine lassosum with the method of clumping when faced with summary statistics of a large number of SNPs. Results from our study suggest that lassosum is faster and more robust than other similar methods in almost all scenarios. They also suggest that accounting for LD with a reference panel is beneficial only when the signals from the data are strong. Using a ``wrong'' reference panel can impair the performance of the PGS when signals are strong and correlated with one another. In the presence of summary statistics from a large number of SNPs, clumping may both enhance or decrease the performance of standard PGS, although its effects on lassosum is attenuated. lassosum combined with pre-filtering by clumping appears to be a robust and reliable option for calculating predictive PGS.

539T

Properties of global and local ancestry adjustments in genetic association tests in admixed populations. *E. Martin[1], I. Tunc[2], Z. Liu[1], A. Beecham[1], G. Beecham[1].* 1) Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) Bioinformatics and Systems Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda MD.

Population substructure can lead to confounding in genetic association tests and failure to adjust properly can result in spurious findings. Genetic ancestry can be measured on a global level, e.g., with principal components, or on a local level for regions across the genome. Recent reports have considered whether it is preferable to adjust for local or global ancestry, but the conclusions depend highly on the model for population structure and trait/disease model. Here, we formally address the question of confounding by ancestry, express regression parameters in terms of population genetics parameters to clarify hypotheses from adjusted and unadjusted models, and consider questions of relative power. We examine theoretical expectations to show population substructure scenarios under which we might expect confounding, and to show the consequence of ancestry adjustment to genetic association tests. Theoretical calculations were verified and generalized using simulations. We show that in a single admixed population, testing the trait locus itself does not lead to confounding by ancestry so it is not necessary to adjust, and in fact adjusting tends to decrease power. However, when testing a marker locus, both global and local ancestry can be confounders. We show that the genetic parameter of adjusted and unadjusted models provides different tests for linkage disequilibrium (LD): the unadjusted model tests for LD in the admixed population, while the local-ancestry adjusted model tests for LD in the ancestral populations. As such, it is recommended that global ancestry adjustment be used for initial screening and that local ancestry for fine-mapping. Similar conclusions hold in stratified admixed populations. We further show how parameters relate to the fixation index F_{ST} (a measure of population diversity), which allows us to discuss results in the context of existing admixed populations.

540F

BinomiRare: An efficient and robust test of the association of a rare-variant with a disease, for pooled and meta-analyses. *T. Sofer.* Biostatistics, University of Washington, Seattle, WA.

Most regression-based tests of the association between a low-count variant and a binary outcome do not protect type 1 error, especially when tests are rejected based on a very low significance threshold. Exception is the Firth test that corrects for biases in the asymptotic properties of the logistic regression model. However, even the Firth test may not control type 1 error well when combining multiple case-control studies via meta-analysis, and it may suffer a substantial loss of power. The problem is exacerbated when the case-control proportions differ between studies. We propose a novel test, BinomiRare, which only requires modeling of the disease risk, independently of the variant, and avoids the calibration problem of other regression-based tests. Disease probabilities are adjusted to confounders using regression models. We quantify the strength of association between a variant and the disease based on the departure of the number of diseased individuals carrying the variants from the expected distribution of disease probability under the null hypothesis of no association between the disease and the variant. We use the Poisson-Binomial distribution to obtain p-values. To meta-analyze the test across multiple cohorts, each cohort provide the disease probabilities of all carriers of the variant of interest and the number of diseased individuals among the carriers. BinomiRare is computationally quick, and faster than the Firth test. It controls type 1 error in both pooled and meta-analysis, even when the case-control proportions differ between the studies. It is more robust than regression-based tests due to fewer assumptions, and thus less prone to false positives. We demonstrate the test in studying the association of rare variants with asthma in the Hispanics Community Health Study/Study of Latinos (HCHS/SOL). The proportion of asthma cases varies substantially between the HCHS/SOL's ethnic groups, from 4% in Mexicans, to 26% in Puerto-Rican. The BinomiRare test results are similar to the results of the Firth test on the pooled cohort, both when the entire pooled sample is analyzed together, and when the dataset is stratified by ethnic group and than meta-analyzed (correlation 0.93 between Firth and BinomiRare p-values). In comparison, meta-analyzing Firth test results across ethnic group results in substantial differences (correlation 0.30 between Firth on stratified and on pooled data), and evidence of inflation (genomic control lambda = 1.32).

541W

Smooth-threshold multivariate genetic prediction with unbiased model selection. *M. Ueki[1], G. Tamiya[2]. 1) Biostatistics Center, Kurume University, Kurume, Fukuoka, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Aoba-Ku, Sendai, Miyagi, Japan.*

We develop a new genetic prediction method, smooth-threshold multivariate genetic prediction, using single nucleotide polymorphisms (SNPs) data in genome-wide association studies (GWASs). Our method consists of two stages. At the first stage, unlike the usual discontinuous SNP screening as used in the gene score method, our method continuously screens SNPs based on the output from standard univariate analysis for marginal association of each SNP. At the second stage, the predictive model is built by a generalized ridge regression simultaneously using the screened SNPs with SNP weight determined by the strength of marginal association. Continuous SNP screening by the smooth thresholding not only makes prediction stable but also leads to a closed form expression of generalized degrees of freedom (GDF). The GDF leads to the Stein's unbiased risk estimation (SURE), which enables data-dependent choice of optimal SNP screening cutoff without using cross-validation. Our method is very rapid because computationally expensive genome-wide scan is required only once in contrast to the penalized regression methods including lasso and elastic net. Simulation studies that mimic real GWAS data with quantitative and binary traits demonstrate that the proposed method outperforms the gene score method and genomic best linear unbiased prediction (GBLUP), and also shows comparable or sometimes improved performance with the lasso and elastic net being known to have good predictive ability but with heavy computational cost. Application to whole-genome sequencing (WGS) data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) exhibits that the proposed method shows higher predictive power than the gene score and GBLUP methods.

542T

Quantifying the extent to which index event biases influence large genetic association studies. *H. Yaghootkar[1], M.P. Bancks[2], S.E. Jones[1], A. McDaid[3,4], R. Beaumont[1], L. Donnelly[5], A.R. Wood[1], A. Campbell[6], J. Tyrrell[1], L.J. Hocking[7], M.A. Tuke[1], K.S. Ruth[1], E.R. Pearson[5], A. Murray[1], R.M. Freathy[1], P.B. Munroe[8,9], C. Hayward[10], C. Palmer[5], M.N. Weedon[1], J.S. Pankow[2], T.M. Frayling[1], Z. Kutalik[3,4]. 1) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, U.K; 2) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, U.S.A; 3) Institute of Social and Preventive Medicine, Lausanne University Hospital, 1010, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland; 5) Division of Cardiovascular & Diabetes Medicine, Medical Research Institute, University of Dundee, Dundee, Scotland, U.K; 6) Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, U.K; 7) Institute of Medical Sciences, University of Aberdeen, Aberdeen, U.K; 8) Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, U.K; 9) NIHR Barts Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine, Queen Mary University of London, London, U.K; 10) Generation Scotland, MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, U.K.*

As genetic association studies increase in size to 100,000s of individuals, subtle biases may influence conclusions. One possible bias is "index event bias" (IEB), also called "collider bias", caused by the stratification by, or enrichment for, disease status when testing associations between gene variants and a disease-associated trait. We first provide a statistical framework to analytically quantify and account for IEB in a study, where only summary statistics are available. We then use a combination of real and simulated data to identify and quantify real examples of IEB, including a single large study (120,000 individuals from the UK Biobank) and a meta-analysis of independent studies. We observed evidence of biased associations for some disease alleles and genetic risk scores, even in population-based studies. For example, a genetic risk score consisting of type 2 diabetes variants was associated with lower BMI in 113,203 type 2 diabetes controls from the population based UK Biobank study (-0.010 SDs BMI per allele, P=5E-4), entirely driven by IEB. Three of 11 individual type 2 diabetes risk alleles, and 10 of 25 hypertension alleles were associated with lower BMI at p<0.05 in UK Biobank when analyzing disease free individuals only, of which all but six hypertension alleles remained associated at p<0.05 after correction for IEB. Our formula (implemented in R) suggested that the associations between *CCND2* and *TCF7L2* diabetes risk alleles and lower BMI could (at least partially) be explained by IEB. Variants remaining associated after correction may be pleiotropic and include those in *CYP17A1* (allele associated with hypertension risk and lower BMI). The formula suggested IEB could also mask a true pleiotropic association: the type 2 diabetes risk allele at *MTNR1B* was associated in the UK Biobank with higher BMI and this result strengthened on correction for IEB. The extent of IEB depends on several factors including the association strength between the trait being analysed and the disease, disease prevalence, sample size, and the effect size and minor allele frequency of the disease associated variant. In conclusion, IEB may result in false positive or negative associations in very large studies stratified or strongly enriched for/against disease cases. Studies examining the joint effect of multiple variants (for example in Mendelian randomization studies) will be more prone to index event biases than those of single variants.

543F

Statistical approaches to replicability crisis in genetic epidemiology. *D. Zaykin[1], O. Vsevolozhskaya[2].* 1) Dept Biostatistics, NIEHS/NIH, Res Triangle Park, NC; 2) Dept. Biostatistics, U Kentucky.

Lack of replicability in science has been a recent concern. While the causes are diverse, genetic epidemiology studies in particular are susceptible to biases and errors introduced by the very nature of the statistical approach. Despite the best intentions, important factors may be unknown, mis-measured, or modeled incorrectly. Still, there is also a question of whether most commonly used approaches to analysis, such as significance testing are inherently flawed. P-values go hand in hand with testing the unrealistic "point null" hypothesis of complete lack of effect, and there are calls that their usage should be abandoned. P-values are often misinterpreted as measures about the hypothesis credibility while another line of criticism is itself statistical and in the context of replicability refers to great variability of P-values in future replication experiments, given a specific small P-value in a discovery study. Despite their flaws P-values are ubiquitous and are here to stay. Fortunately, when augmented with minimal additional information, P-values that result from common statistical tests employed in genetic epidemiology can yield full posterior distributions for the usual measures of effect sizes, such as odds ratios and relative risks. The proposed method alleviates many shortcomings of P-value and their usage, not the least of which is testing an implausible hypothesis of complete absence of effect.

544W

Imputation of low frequency and rare coding variants in founder population improved by usage of whole exome sequence based reference panel. *I. Surakka[1,2], A.P. Sarin[1,2], K. Karczewski[3,4], R. Durbin[5], D. MacArthur[3,4], V. Salomaa[2], A. Palotie[1,4,6,7,8], S. Ripatti[1,5,9], SISu project group.* 1) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA; 4) Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, USA; 5) Wellcome Trust Sanger Institute, Hinxton, UK; 6) The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA; 8) Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Department of Public Health, University of Helsinki, Helsinki, Finland.

Population specific reference panel improves imputation accuracy especially for rare and low frequency variants, particularly in population isolates, such as Finns. However, the gain in imputation precision has not been quantified for sequence-based reference panels. We compared the imputation quality between a Finnish population based sequence reference panel and global reference panels from 1000 Genomes project (1000G) and the Haplotype Reference Consortium (HRC). In addition, we evaluated the utility of extending the Finnish low-coverage whole genome sequencing (WGS) panel (N = 1,941) with a jointly called whole exome sequencing (WES) based reference panel (N = 4,932) when imputing coding variation into our test dataset of 10,489 Finns with Illumina CoreExome genotypes. We masked all variation with minor allele frequency (MAF) < 1%. Using the Finnish WGS reference panel we saw a ~100% increase in the sensitivity (the number of high confidence imputed exonic variants) in 0.1-0.5% allele frequency range compared to the 1000G panel. The HRC reference panel provided 7% more variants compared to the Finnish WGS panel but with the cost of introducing 0.8% false positive variant calls compared to 0.1% with the Finnish WGS reference panel. The highest sensitivity we observed for the combined panel of 1000G and WGS. However, the use of the combined reference panel introduced high number of false positive variants with high imputation quality measures; over 3% of the monomorphic variants observed in the test dataset imputed as polymorphic with info > 0.7. By combining the WES panel with the population specific WGS reference panel, we observed over 20% more high confidence imputed coding variants in the 0.1-0.5% allele frequency range compared to the Finnish WGS only panel. When looking at the imputation results of the polymorphic masked variants we observed that especially in the lowest allele frequency range (MAF < 0.1%) the population specific panels induced less false negative calls and the percentage of well-imputed variation was larger than when using the admixed reference panels. Our results highlight a problem of introducing falsely polymorphic low frequency and rare variants when using dual panels with different ancestries. In addition, our study shows that a population specific WES based reference panel boosts considerably the imputation of rare and low frequency coding variants.

545T

Correcting subtle stratification in summary association statistics. *G. Bhatia[1,2], N. Furlotte[3], X. Liu[1], H.K. Finucane[1,4], A. Gusev[1,2], A.L. Price[1,2].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) 23andMe, Inc. Mountain View, CA; 4) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

   Population stratification is a well-documented confounder in GWAS. While inclusion of PC covariates computed from common SNPs (SNP-PCs) often provides an effective solution to stratification, it is not applicable post-hoc to meta-analyses without access to individual-level data. In addition, this strategy will not correct for subtle stratification that is not captured by SNP-PCs. In analyses of summary statistics from 39 GWAS (mean sample size n=89k), including 21 GWAS from 23andMe that used SNP-PC covariates, we observed a significantly inflated LD score regression (LDSC) intercept for several traits—suggesting that subtle, uncorrected stratification can be a major concern, even when SNP-PC covariates are included. Here we propose a new method, PC loading regression, to correct for stratification in summary statistics by leveraging SNP loadings for PCs computed in a large reference panel. In addition to SNP-PCs, the method can be applied to haploSNP-PCs, i.e. PCs computed from a larger number of rare haploSNPs (haplotype variants constructed using a 4-gamete test) that better capture subtle structure. Using simulations based on real genotypes from 54k GERA individuals of diverse European ancestry, we show that application of PC loading regression to summary statistics computed without PC correction can effectively correct for stratification along top PCs. Correcting for the top 4 SNP-PCs in GERA data, we did not observe a significant reduction in LDSC intercept for 23andMe summary statistics, which already included SNP-PC covariates. However, when correcting for additional haploSNP-PCs in 23andme GWAS, inflation in the LDSC intercept was eliminated for eye color, hair color, and skin color and substantially reduced for height (1.34 to 1.16; n=430k individuals). After correcting for SNP-PCs and haploSNP-PCs in GIANT height summary statistics (Wood et al. 2014 Nat Genet), the LDSC intercept was reduced from 1.27 to 0.97 (n=250k individuals) consistent with no confounding after correction. Many association signals were eliminated by this correction including all 11 genome-wide significant SNPs at the LCT locus, which is highly differentiated among European populations and widely known to produce spurious signals (Campbell et al. 2005 Nat Genet). Overall, our results suggest that subtle stratification is a concern in GWAS of large sample size and that PC loading regression can correct for this stratification.

546F

Improving transcriptome-wide association studies with gene-network aided prediction of gene expression. *D. Chang, A. Wuster, T. Behrens, R. Graham, T. Bhangale.* Genentech, South San Francisco, CA.

   Several new methods have been developed to impute gene expression data into GWAS cohorts where expression data is not routinely collected. These methods use reference data, where genotype and expression data are collected in the same individuals, to train models that predict expression from genotype data. After applying the learned models to GWAS samples, imputed expression can be associated with disease risk. These methods both lower the multiple testing burden, and test a more direct hypothesis –is genic transcription associated with disease risk? However, expression cannot be reliably predicted from SNPs for many genes (genes with low heritability), limiting the utility of these methods. Gene expression is often influenced by (or correlated with) several other genes. We thus developed a new method that uses SNPs predictive of expression for a given gene's interacting partners to improve overall prediction accuracy. We applied the method to an RNA-sequencing dataset of monocyte derived dendritic cells (Lee *et al.* Science 2014) stimulated with influenza virus (N=183, training data) and interferon-beta (N=165, test data). Genotypes were obtained from Illumina one million SNP array, imputed using 1000 Genomes haplotypes. We first applied regularized regression (Gamazon *et al.* Nature Genetics 2015), to obtain predictive models for each gene based on *cis* SNPs. Next we applied graphical lasso (*huge* R package) to estimate the gene network from the gene expression data. Due to the computational burden of this estimation, we first separated genes into clusters based on correlated expression (*WGCNA* R package) and estimated networks separately for each cluster. Finally, the expression level of a gene was predicted based on *cis* SNPs of the gene and its interacting partners, and the estimated correlation between genes in the network. Of the 7,118 robustly expressed genes (RPKM>5), 6,379 genes were assigned to a cluster. Overall, for genes with *cis*-predictive SNPs, we found the average correlation between observed and predicted expression from *cis* SNPs alone to be 0.256, while for our method it was 0.349. For 4,396 genes, expression could not be predicted using *cis* SNPs as the regularized regression failed to include any SNP in the model. However, their expression could be predicted using our method (mean correlation=0.339). In conclusion, incorporating gene networks can improve prediction accuracy and power when imputing gene expression into GWAS datasets.

547W

Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *J.P. Cook[1], A. Mahajan[2], A.P. Morris[1].* 1) Department of Biostatistics, Farr Institute, Liverpool, Liverpool, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Linear mixed models (LMMs) are becoming increasingly popular for the analysis of genome-wide association studies (GWAS) because they account for relatedness and population structure via a kinship matrix. However, these models assume that the outcome of interest is quantitative, and their properties for binary traits have not been widely investigated, particularly in the context of meta-analysis. In this study, we performed simulations to compare the performance of linear and logistic regression models under alternative weighting schemes in a fixed-effects meta-analysis framework, considering designs that incorporate variable case-control imbalance, confounding factors and population stratification. Fixed-effects meta-analysis was undertaken with inverse-variance weighting of effect sizes (implemented in GWAMA) and effective sample size weighting of directed Z-scores (implemented in METAL). We simulated three scenarios for case-control imbalance: (i) no imbalance (1:1 ratio across all studies); (ii) moderate imbalance (variable ratio of 3:1 to 1:3 across studies); and (iii) extreme imbalance (variable ratio of 19:1 to 1:19 across studies). Within each set of simulations, comparisons were performed at effect allele frequencies of 0.5, 0.2, 0.1, 0.05 and 0.01. Our results show that linear models can be used for the analysis of binary traits, without loss of power compared to logistic models, irrespective of the extent of case-control imbalance, provided that one of the following meta-analysis schemes is used: (i) effective sample size weighting of *Z*-scores; or (ii) inverse-variance weighting of allelic effect sizes after conversion onto the log-odds scale. However, in the presence of extreme case-control imbalance, there is a substantial loss in power through application of inverse-variance weighting of allelic effect sizes from the linear model, without conversion onto the log-odds scale. Our conclusions are valid in the presence of binary confounder with moderate effect size (odds-ratio <5) and extreme population structure through application of LMMs. Our study has important implications for the design and analysis of GWAS, and the utility of linear modelling approaches in large-scale meta-analysis of complex binary traits.

548T

A fast and accurate algorithm to test for binary phenotypes and its application to PheWAS. *R. Dey, E. Schmidt, G. Abecasis, S. Lee.* Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI.

We propose a fast and well-calibrated single variant test for binary phenotypes with an application to Phenome-wide Association Studies (PheWAS). The availability of electronic medical record (EMR)-based phenotypes allows for genome-wide association analyses in thousands of traits, and has great potential to identify novel genetic variants associated with clinical phenotypes. We can interpret the 'PheWAS' result for a single genetic variant by observing its association across a landscape of phenotypes. Since PheWAS can test 1000s of phenotypes, and many of the binary phenotypes are relatively unbalanced (case:control=1:5) or often extremely unbalanced (case:control=1:600), existing methods cannot efficiently analyze them. For example, score test, which is computationally efficient, can have highly inflated type I error rates. Firth's penalized likelihood ratio test is more robust for controlling the type I error, but it is not scalable to handle large PheWAS datasets. Here we propose a computationally fast score test based method that estimates the distribution of the test statistic using the saddlepoint approximation. Our method is much faster than Firth's test (~10-100 times). It can also adjust for covariates and control type I error rates even when the case-control ratio is extremely unbalanced. Further, we extend our method to allow meta-analysis by identifying a minimal set of statistics that can be shared across different studies. We demonstrate superior performance of our method through extensive simulation studies and apply it to a real data example from the Michigan Genomics Initiative.

549F

Genome-wide association of 2,500 structural and functional brain imaging phenotypes in the UK Biobank study. *L.T. Elliott[1], F. Alfaro-Almagro[2], M. Jenkinson[2], K.L. Miller[2], S.M. Smith[2], J. Marchini[1,3].* 1) Department of Statistics, University of Oxford, 24-29 St. Giles', Oxford, OX1 3LB; 2) Oxford University Centre for Functional MRI of the Brain (FMRIB), Oxford, OX3 9DU; 3) The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN.

Brain function, connectivity and structure are known to be heritable and are important for the study of many human disease traits, but the localization within the genome of the genetic variants driving this heritability has remained elusive due to the sample size and quality of data required for this task. The UK Biobank has collected genetic data on around 500,000 UK individuals, and (to date) brain imaging data on 10,000 individuals, later to rise to 100,000. We will present the results of a genome wide association study on the intersection of the genetic and imaging datasets, which consists of 2500 brain imaging derived phenotypes and up to 10,000 individuals genotyped at around 10 million polymorphic SNPs. The phenotypes include connection strengths between brain areas assessed during a resting state analysis, volumes of brain structures and tissues, and microstructural white matter properties. Along with univariate tests for association, we will report results from multi-phenotype approaches that work on the phenotype-by-SNP summary statistic matrix, thereby uncovering latent phenotypes with strong associations.

550W

Is selective inference more powerful in meta-, gene-based and pathway analyses? *E.T. Goossens[1], L. Sun[2,3].* 1) Department of Statistics, Purdue University, West Lafayette, IN; 2) Department of Statistics, University of Toronto, Toronto, ON; 3) Dalla Lana School of Public Health, University of Toronto, Toronto, ON.

Statistical analyses of high dimensional data, including genetic association studies, are often interested in combining evidence across multiple sources. The most common setting is meta-analysis, in which p-values (or other types of summary statistics) are combined across all studies to test the global null hypothesis that there is no association. This agnostic (either fixed-effect or random-effect) approach, however, may not be powerful when a proportion of the studies have truly null effects. Selective inference (appropriately testing parameters after some selection algorithm) has the potential to address this issue, but requires explicitly accounting for inherent selection bias. We propose a novel ordered-subset approach that simultaneously performs signal selection and adjusts for selection bias. We derive the approximate distribution for the test statistic under the null, and show the proposed method is accurate in finite samples. The performance of this method depends on a combination of factors: k, the total number of variables to be combined; k1 (k0), the number of true (null) variables; and the corresponding signal strength. Compared with traditional meta-analysis, our method improves power in the presence of heterogeneity when k0/k1 is large and each signal is relatively weak. In the absence of a mixture of null and non-null signals, we show the loss of power is limited via a simulation study. We also employ our method in the regression setting, illustrating its suitability for joint analysis of multiple SNPs in gene-based association studies and multiple genes in pathway analyses.

551T

An independent component analysis framework for modeling confounding factors in expression quantitative trait loci analysis. *J. Ju[1,2,3], S. Shenoy[3], J. Mezey[2,3,4].* 1) Physiology, Biophysics and Systems Biology Graduate Program, Weill Cornell Graduate School of Medical Sciences, New York, NY; 2) Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY; 3) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 4) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Expression Quantitative Trait Loci (eQTL) studies provide valuable insights into genetic mechanisms by identifying associations between genetic variations and gene expression. These types of analyses are complicated by non-genetic factors that influence expression levels, such as sample-specific environmental effects or technical effects related to differences between laboratories and procedures, which often remain unaccounted for and can potentially cause spurious correlations and reduce power to detect true associations. The importance of correcting for confounding factors is increasing with the recent rise of consortium scale studies in which samples from various populations are processed across multiple laboratories and have more potential sources of variation. Here we introduce CONFETI: CONfounding Factor Estimation Through Independent component analysis, a method to correct for non-genetic confounding variance in eQTL analysis. CONFETI uses Independent Component Analysis (ICA) to estimate generative sources from the observed data under the assumption that the data is a linear combination of statistically independent components. These components can be individually inspected for exploratory data analysis and can be further characterized as genetic or non-genetic components. The similarity between samples is calculated using a lower dimensional representation of the observed data constructed with non-genetic independent components, which is subsequently incorporated in a linear mixed model framework for eQTL analysis. To evaluate the performance of CONFETI, we compared it with published methods in synthetic eQTL data and multiple human datasets. In simulations we introduced a mix of sparse and Gaussian confounding factors and found that CONFETI most accurately corrected the effects of sparse confounding factors, while being able to separate genetic components from non-genetic components. We analyzed multiple tissues from the Genotype-Tissue Expression (GTEx) consortium, and evaluated the performance of different methods by measuring eQTL replication. We found the highest number of replicating eQTLs in similar tissues and across different tissues using CONFETI.

552F

A powerful framework for combining heterogeneous data in genetic association studies. *C. Lee[1], E. Eskin[2,3], B. Han[1].* 1) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, Korea; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, California 90095, USA; 3) Department of Computer Science, University of California, Los Angeles, Los Angeles, California 90095, USA.

Meta-analysis has become essential for combining genome-wide association studies (GWASs). Recent study designs of meta-analysis of GWASs now combine different layers of information, such as different phenotypes (multiple autoimmune diseases, psychiatric diseases, or expression levels from different tissues), different ethnicities, and different environmental interventions. In these study designs, it is important to account for heterogeneity, which refers to the effect size differences between datasets. Previously developed meta-analysis methods including RE2 (Han and Eskin, AJHG 2011) and ASSET (Bhattacharjee et al., AJHG 2012) can account for heterogeneity, but it is unclear if they achieve optimal power. Here, we propose a powerful method called RE3, which accounts for heterogeneity to obtain high power in meta-analysis of GWASs. The core idea of RE3 is to utilize the observation that the magnitude of heterogeneity likely lies within an expected range in meta-analysis of GWASs. By focusing on the expected range of heterogeneity in the newly designed statistic, RE3 can dramatically improve power compared to our previous approach, RE2. We simulated different patterns of heterogeneity, and RE3 outperformed RE2 and ASSET in all simulated scenarios. Then we performed a full literature search for publications that cited RE2 method, and obtained meta-analysis input data from them, which can represent the real patterns of heterogeneity. When we applied the methods to this real data gathered without bias, RE3 consistently produced more significant p-values compared with competing methods for all associated variants that we tested (5.4 times smaller p-values on average). These results underscore that RE3 can potentially be a useful tool for future study designs that aggregates different layers of information. Unlike other competing methods, RE3 is highly efficient even for combining a large number of datasets; it only takes a day for meta-analyzing a hundred GWASs. RE3 is freely available at http://software.buhmhan.com/RE3.

553W

Modeling ancestry-dependent phenotypic variance reduces bias and increases power in genetic association studies. *S. Musharoff[1], D. Park[1], S. Huntsman[1], C. Eng[1], E. Burchard[1,2], N. Zaitlen[1].* 1) Lung Biology, University of California, San Francisco, San Francisco, CA; 2) Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Many complex human phenotypes vary dramatically in their distributions between populations. Genetic association studies typically use estimates of ancestry, such as principal components (PCs), as fixed-effect covariates to prevent confounding caused by a dependence of phenotypic mean on ancestry. However, the current gold standard approach approach of including PC covariates in linear regression models (LR+PC) assumes that different populations have the same phenotypic variance, which may not hold for recently admixed populations. In this work we consider the possibility that populations with differences in phenotypic mean also have differences in phenotypic variance. First, we show this is the typical case under an additive genetic architecture. Then, we develop ADGLM, a likelihood-based method based on a double generalized linear model, to account for relationships between ancestry and phenotypic variance in genetic association studies. In simulations, our test ADGLM has better power than several linear regression tests that assume equal variance across groups. We observe power increases of 12 - 66% and obtain unbiased parameter estimates for data simulated with realistic effect sizes and minor allele frequency differences of 0.45. Furthermore, we show that the standard approach LR+PC can lead to inflation or deflation of p-values for tests of genetic association when population phenotypic variances differ. For example, simulated populations with minor allele frequencies of 0.05 and 0.5 produce test statistics with an inflation factor (lambdaGC value) of 1.56, which ADGLM fixes. When applied to the Study of African Americans, Asthma, Genes and Environments (SAGE), ADGLM find significant associations of baseline lung function (FEV1) with global ancestry proportion when either sex or BMI is a sole covariate. By contrast, LR+PC requires additional covariates (age, sex, height, and weight) for significant associations. When applied to GALA II Puerto Ricans, ADGLM finds ancestry significantly associated with mean methylation at 8 of ~320K genome-wide probes while LR+PC finds 1 significant association. ADGLM also finds 44 probes significantly associated with methylation variance, which may be due to ancestry-associated environmental effects. Overall, ADGLM finds more significant associations than linear regression with PCs possibly because ancestry affects phenotypic variances as well as phenotypic means, and is promising for other association studies.

554T

Statistical correction of the winner's curse explains replication variability in quantitative trait genome-wide association studies. *C. Palmer[1], I. Pe'er[1,2].* 1) Department of Systems Biology, Columbia University Medical Center, New York, NY, USA; 2) Department of Computer Science, Columbia University, New York, NY, USA.

Genome-wide association studies (GWAS) have identified hundreds of candidate SNPs putatively responsible for variation in human quantitative traits. However, effect sizes are often small and replication can appear inconsistent, leading to pervasive questions about the utility of the field as a whole. We seek to validate the performance of GWAS discovery of SNPs by systematically evaluating replication efficiency across the field of human quantitative trait genetics. We extract all 332 studies of quantitative, normally-distributed traits from the NHGRI GWAS database that attempted replication of results, and apply filters to find studies providing enough published data to perform power calculations. We find that 229 of these studies fail to provide full information of sample sizes, allele frequencies, or effect sizes for replicated and non-replicated SNPs. We find that, in 87 studies attempting replication of 1269 loci in samples with similar continental ancestry to the discovery population, discovery and replication additive effect estimates are statistically indistinguishable once adjusted for the winner's curse. Replication of 385 loci in 13 studies across populations of different continental ancestry shows a detectable attenuation of replication effect estimate for replication p-values below 1e-7, attributable to linkage disequilibrium variation across distinct ancestries. We see no systematic evidence for distinct behaviors of a given allele across different samples beyond those attributable to LD interactions with tagged true variants. Outlier detection using predicted replication rate estimates can successfully detect studies with clear study design failures unrelated to GWAS itself. Given the evident reliability of replication in human quantitative trait biology, we provide a tool that can diagnose systemic failures of study design by detecting significant deviations from expected replication when considering any variable replication p-value threshold.

555F

Cost-effective strategies for rare-variant genome-wide association studies in the age of large reference panels. *C. Quick, C. Fuchsberger, M. Boehnke. Department of Biostatistics, University of Michigan, Ann Arbor, MI.*

Rare variants (RVs) are hypothesized to be a major source of missing heritability for complex diseases, establishing a need for efficient RV genome-wide association study (GWAS) designs. While array genotyping and imputation using a publicly available reference panel is cost-effective for detecting association at common variants, this strategy is less effective for RV GWAS. In particular, imputed genotyping array data only include RVs that are present in the reference panel or assayed by the array, and imputation accuracy for RVs is typically lower than for common variants. Here, we consider a hybrid GWAS strategy in which a subset of study participants is whole-genome sequenced and the remainder are array genotyped and imputed with a reference panel comprised of the sequenced study participants and an external reference panel (for example from the 1000 Genomes Project (1000G) or the Haplotype Reference Consortium (HRC)). We use analytic and simulation approaches to compute GWAS power as a function of the numbers of sequenced and array-genotyped study participants given a specified external panel and complex disease model. We applied our approach to find optimal numbers of study participants to sequence and array-genotype for a range of complex disease models and external panel sizes given current sequencing and array-genotyping costs. Our results show optimal hybrid strategies can substantially increase GWAS power relative to sequence-only and array-only strategies under a budget constraint when risk RVs of large effect contribute substantially to disease heritability. Gains in GWAS power over an array-only strategy decrease as external panel size increases; we discuss implications of small versus increasingly large publicly available reference panels (e.g., 1000G with ~1K samples versus HRC with ~33K samples) for cost-effective RV GWAS strategies. Finally, we introduce a web-based power calculator with flexible disease model and reference panel specification to facilitate identifying cost-effective strategies and estimating sample size for RV GWAS.

556W

Combining patient electronic health information with genetic data to gain novel biomedical insights. *E. Schmidt[1], L. Fritsche[2], S. Lee[1], P. VandeHaar[1], C. Scheller[1], C. Brummett[3], S. Kheterpal[3], G. Abecasis[1]. 1) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI.*

The Michigan Genomics Initiative (MGI) is a collaborative research effort among physicians and researchers at the University of Michigan (UM) with the goal of combining patient electronic medical records (EMR) with genetic data to gain novel biomedical insights. Patients undergoing surgery at the UM Health System are invited to participate. The surgical procedural period provides a unique opportunity to collect patient biospecimens and enriched health information. Together, these data offer a powerful resource for innovative data analysis and discovery. Here we interrogate 7.7M common variants and 1,448 ICD-9 code-defined disease states to establish relationships between the genome and a diverse set of EMR-based outcomes, providing a comprehensive resource for the biomedical and genetic research communities. Blood samples from consenting surgical patients are collected and genotyped at 270K common markers on the Illumina HumanCoreExome array, which contains exome and custom GWAS content. Genotypes are imputed using the Haplotype Reference Consortium reference panel, providing dense mapping at over 7.7M common genetic markers. Phenotypes inferred from ICD-9 (International Classification of Disease, 9th edition) billing codes assigned to patients during their hospital or clinic visits are translated into broader 'PheWAS' (phenome-wide association study) code groups of shared disease etiology. The MGI cohort is enriched for cases with neoplasms including skin cancer (n=2,359) and breast cancer (n=1,160), reflecting the specialties of the UM hospital. In total, we perform genome-wide analyses in 18,267 unrelated European individuals for 1,448 PheWAS codes with at least 20 cases, across common variants with minor allele frequency > 0.01. We replicate several well-known genetic associations at genome-wide significance including Factor 5 and thrombosis (P=4x10^{-39}), *TCF7L2* and type 2 diabetes (P=2x10^{-11}), *PITX2* and atrial fibrillation (P=3x10^{-9}), and *FGFR2* and breast cancer (P=4x10^{-8}). In addition, we examine the shared genetic effects among different related and apparently unrelated traits, providing insight into relationships between traits not previously studied. Finally, we expose potential limitations of using electronic billing data for research such as the misclassification of type 1 and type 2 diabetes. Associations for a total of 1,448 traits can be explored in dynamic GWAS-to-PheWAS landscapes via our interactive browser at http://pheweb.sph.umich.edu.

557T

Assessment and control of batch effects in genotyping Million Veteran Program (MVP) subjects. *N. Sun[1,2], M. Li[3,4], Y. Shi[3], H.S. Hunter-Zinck[3], J. Huang[3], C. Pan[5,6], J. Gelernter[1,7], T.L. Assimes[5,6], P.S. Tsao[5,6], C.J. O'Donnell[3,8], E.R. Hauser[9,10], S. Pyarajan[3,8], H. Zhao[1,2] for MVP Genomic Working Group.* 1) VA Connecticut Healthcare System, West Haven, CT; 2) Yale School of Public Health, New Haven, Connecticut; 3) VA Boston Healthcare System, Boston, MA; 4) University of Utah, Salt Lake City, UT; Johns Hopkins University, Baltimore, MD; 5) VA Palo Alto Health Care System, Palo Alto, CA; 6) Stanford University School of Medicine, Stanford, CA; 7) Yale School of Medicine, New Haven, CT; 8) Harvard Medical School, Boston, MA; 9) Durham VA Medical Center, Durham, NC; 10) Duke University, Durham, NC.

The Million Veteran Program (MVP) is an observational cohort study and mega-biobank in the Department of Veterans Affairs (VA) health care system. Like the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and UK biobank, MVP integrates electronic health records with genomic laboratory testing data for all individuals. The first set of the biological samples of MVP subjects were collected over a 3-year period from over 50 different recruitment sites at the VA and processed at a central biorepository using standard sample quality metrics for DNA quality to allow genotyping. Duplicated samples were included to assess genotyping quality. A custom Affymetrix Axiom MVP_1.0 array with 723,305 probe sets was genotyped for 200,846 samples by two independent laboratories. In order to produce high quality genotyping data, we analyzed evidence for batch effects, which are well known to have a significant impact on genotype calling results. The large MVP dataset provides the opportunity to identify the most important experimental variables needed to adjust for technical and experimental variation when genotyping is performed over time and in different laboratories. We performed ANOVA analysis on a total of 17 experimental conditions to identify time and plate lot as the main factors affecting genotype calling variation. Various batching strategies were compared using the genotyping calls obtained from applying Affymetrix Power Tools (APT) and SNPolisher. The size of each batch was controlled at 4000 to 5000 samples to optimize the calling quality and the computational time of APT. The final 45 batches were identified to maximize the number of recommended probe sets in common across multiple batches, increase the concordance for duplicates within lab (>97%) and between labs (>96%), and reduce the discordance fractions per paired genotypes with at least one minor allele over duplicates. In this presentation, we will describe our strategies to minimize batch effect for the mega-cohort of MVP samples.

558F

Novel nonparametric test for multiple traits of sequencing data. *X. Wang[1], H. Huang[2].* 1) University of North Texas, Denton, TX; 2) University of Wisconsin-Milwaukee, Milwaukee, WI.

Pleiotropy, the effect of one variant on multiple traits, is a widespread phenomenon in complex diseases. Joint analysis of multiple traits can not only increase statistical power to detect disease susceptible genetic variants but also improve our understanding of the etiology of complex diseases. Although the cost of next generation sequencing (NGS) has been reduced, it is expensive to detect rare variants, and typically requires large samples. Moreover, most of the existing methods are parametric. These methods often assume parametric models and particular probability distributions. However, for NGS data, those assumptions may not be correct and it is typically difficult or impossible to ascertain whether or not certain parametric assumptions are justifiable. Therefore, parametric statistical methods can be very misleading. We developed novel nonparametric statistical methods to efficiently detect rare variants for multiple traits, which can not only be applied in large samples, but also work well for small samples. To perform the proposed nonparametric method, we first classified subjects into two categories; whether they carry rare variants or not in a gene. We then employed a nonparametric test to find out whether the two different categories differ significantly with regard to their risk. In this study, we propose three different nonparametric tests: Dempster-ANOVA type, Lawley-Hotelling type, and Bartlett-Nanda-Pillai type. To evaluate the performance of the nonparametric tests, we conducted extensive simulation studies with different underlying population distributions, namely, multivariate normal traits, t-distributed traits, and Cauchy type of traits. We accessed the type I error rates and compared the power of proposed tests with the existing test MSKAT. The type I error rates of all three tests are under control with different underlying population distributions, while MSKAT has inflated type I error rates for t-distributed traits and Cauchy type of traits. When sample size is less than 800, the nonparametric methods are more powerful than MSKAT in most scenarios for multivariate normal traits and t-distributed traits. The proposed tests are more powerful than MSKAT for Cauchy type of traits when the sample size is 1000 though MSKAT has inflated type I error rate. Therefore, our new methods are valid and powerful tests to detect rare variants for multiple traits in next generation sequencing data.

559W

Imputation quality scores are poorer predictors of true quality for rare than for common variants. *P. Yajnik[1], C. Fuchsberger[2], M. Boehnke[1].* 1) Biostatistics - School of Public Health, University of Michigan, Ann Arbor, MI; 2) Center for Biomedicine, EURAC, Bolzano, Italy.

The advent of high-throughput DNA sequencing has made it possible to assay nearly the complete spectrum of genetic variation. However, genome sequencing remains expensive for large epidemiologic studies. Genotyping on dense arrays followed by imputation is a cost-effective alternative to sequencing. Since genotype imputation is imperfect, imputation algorithms also output a variant level quality score that is typically used to filter out variants believed to have been imputed poorly. As with genotypes, the true quality of imputation is unknown and the quality scores are predictions of the true quality. The quality scores are known to work well as classifiers of true imputation quality for common variants (MAF ≥ 5%). We study their classification properties for low-frequency ($0.5\% \le$ MAF $< 5\%$) and rare variants (MAF $< 0.5\%$). We imputed variants on chromosome 20 in 8800 participants recruited from Finland (METSIM study) using Minimac3 with the 1000 Genomes Phase 3 reference panel and Illumina OmniExpress array as the backbone. Participants were also genotyped with the Illumina Exome array. Exome array genotypes were assumed to be true. 1206 imputed variants were also present on the Exome array of which 54.1%, 24%, and 21.9% were rare, low-frequency, and common, respectively. True imputation quality (defined as the squared correlation between imputed dosages and true genotypes) decreased with decreasing MAF. The classification accuracy of the Minimac3 quality score (MachRSQ) also decreased with decreasing MAF. The conventional cutoff of 0.3 used for common variants resulted in a positive predictive value (probability that the true r-squared exceeded 0.5 given that MachRSQ > 0.3) of 100%, 93.4%, and 71.4% for common, low-frequency and rare variants, respectively. We also observed that posterior genotype probabilities were miscalibrated for rare variants; the observed proportion of heterozygotes and alternate homozygotes was smaller than the respective imputed posterior probabilities. Finally, the mean squared error between true quality and MachRSQ was larger for rarer variants. The poorer performance of quality scores for rarer variants is partly due to the larger proportion of poorly imputed variants. However, the larger variance of prediction and bias due to miscalibrated posterior probabilities also decreased classification efficiency. We use our results to recommend best practices for evaluating and interpreting imputation quality metrics.

560T

Integrated enrichment analysis of genetic variants and biological pathways using GWAS summary statistics. *X. Zhu[1], M. Stephens[1,2].* 1) Department of Statistics, The University of Chicago, Chicago, IL; 2) Department of Human Genetics, The University of Chicago, Chicago, IL.

Statistical analyses that integrate genome-wide association studies (GWAS) with external genomic information, such as biological pathways, can increase power and yields new biological insights. We develop a regression-based approach to these analyses using GWAS summary statistics. Specifically, we perform Bayesian multiple regression analysis by combining a likelihood based on summary data (Zhu and Stephens, 2016) with a prior that incorporates pathway information (Carbonetto and Stephens, 2013). This combination not only allows the discovery of enriched pathway, estimation of enrichment level and prioritization of associated variants simultaneously, but also removes the need for individual-level data. Within the regression framework, our method automatically accounts for linkage disequilibrium, whereas other methods address this issue with extra complexity. We design a parallel variational algorithm for large-scale inference. Indeed, using moderate computational resources, a single analyst can finish the integrated analysis of 1.1 million SNPs and 3915 pathways within 1.5 day. We apply our method to the summary data of 21 complex human traits. Our analyses confirm many published links between pathways and phenotypes, including hedgehog signaling and adult height (BF=2.4E36), ERK MAPK signaling and body fat distribution (BF=3.2E67), IL23-mediated signaling and inflammatory bowel disease (BF=1.0E29), T helper cell surface molecules and rheumatoid arthritis (BF=3.3E8), adipocytokine signaling and type 2 diabetes (BF=2.7E55). Our analyses also implicate several links that were not highlighted in earlier GWAS, such as endochondral ossification (EO) and height (BF=3.0E63), S-Adenosyl-L-methionine biosynthesis and Alzheimer's disease (BF=3.3E88), synthesis of phosphatidylserine and heart rate (BF=1.6E244), thyroid hormone metabolism and myocardial infarction (BF=3.8E191). Moreover, our method uncovers biologically relevant genes in the enriched pathways that were not identified in previous analyses of the same data. For example, prioritization of EO genes shows that a putatively novel locus contains at least 1 height-associated SNP with posterior probability 1.00; in contrast, this probability is only 0.41 without considering the pathway. This locus is not near any of the previous 697 GWAS hits (distance>3.9Mb), and covers the gene *PTH1R*, which is involved in bone formation (Datta et al, 2010) and osteocyte survival (Maycas et al, 2015).

561F

A comparison study of multivariate fixed models and Gene Association with Multiple Traits (GAMuT) for next-generation sequencing. *R. Fan[1], C. Chiu[1], J.L. Mills[1], J.S. Jung[2], Y.F. Wang[3], A.F. Wilson[4], B.E. Bailey- Wislon[4], D.E. Weeks[5], C.I. Amos[6], M. Boehnke[7], M.M. Xiong[8].* 1) BBB, DIPHR, NIH, NICHD, Rockville, MD; 2) National Institute on Alcohol Abuse and Alcoholism; 3) Center for Drug Evaluation and Research, FDA; 4) National Human Genome Research Institute, NIH; 5) University of Pittsburgh; 6) Dartmouth Medical College; 7) The University of Michigan, Ann Arbor; 8) University of Texas – Houston.

In this project, extensive simulations are performed to compare two statistical methods to analyze multiple correlated quantitative phenotypes: (1) approximate F-distributed tests of multivariate functional linear models (MFLM) and additive models of multivariate analysis of variance (MANOVA), and (2) Gene Association with Multiple Traits (GAMuT) for association testing of high-dimensional genotype data. It is shown that approximate F-distributed tests of MFLM and MANOVA have higher power and are more appropriate for major gene association analysis (i.e., scenarios in which some genetic variants have relatively large effects on the phenotypes); GAMuT has higher power and is more appropriate for analyzing polygenic effects (i.e., effects from a large number of genetic variants each of which contributes a small amount to the phenotypes). MFLM and MANOVA are very flexible and can be used to perform association analysis for: (i) rare variants, (ii) common variants, and (iii) a combination of rare and common variants. Although GAMuT was designed to analyze rare variants, it can be applied to analyze a combination of rare and common variants and it performs well when (1) the number of genetic variants is large and (2) each variant contributes a small amount to the phenotypes (i.e., polygenes). MFLM and MANOVA are fixed effect models which perform well for major gene association analysis. GAMuT can be viewed as an extension of sequence kernel association tests (SKAT) of mixed models. Both GAMuT and SKAT are more appropriate for analyzing polygenic effects and they perform well not only in the rare variant case, but also in the case of a combination of rare and common variants. Data analyses of European cohorts and the Trinity Students Study are presented to compare the performance of the two methods.

562W

Bayesian evaluation of variant involvement in Mendelian disease. *D. Greene[1,2,3], E. Turro[1,2,3], S. Richardson[3].* 1) Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK, CB2 0XY; 2) NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK, CB2 0PT; 3) Medical Research Council Biostatistics Unit, Cambridge Biomedical Campus, Cambridge, UK, CB2 0SR.

Rare diseases are often caused by high-penetrance rare variants. Because of genetic heterogeneity and low numbers of cases, statistical procedures testing each variant for marginal association with phenotype are often underpowered. Procedures which aggregate rare variants across regions also sacrifice power because they are likely to include non-pathogenic rare variants in the aggregation. Therefore, models are required which account for a mixture of pathogenic and non-pathogenic rare variants explicitly. Ideally, the composition of this mixture would be informed by additional information such as sequence conservation across species and population allele frequency. It is also important to model the modes of inheritance typical in Mendelian disease, including autosomal dominant, autosomal recessive, and X-linked recessive inheritance. We present a new method, `Bayesian Evaluation of Variant Involvement in Mendelian Disease' (BeviMed), which assesses the evidence of association between a case-control label and presence of one or more genetic configurations of alleles at rare variant sites. The configurations depend on a latent partitioning of variants into pathogenic and non-pathogenic groups and can be informed by external information. Different modes of inheritance are modelled by conditioning on the number of pathogenic alleles carried by each individual (and gender, in the case of X-linked inheritance). Thus compound heterozygosity and homozygosity can be treated equivalently. Inference is performed by obtaining posterior probabilities of the different models of association and a baseline model of no association. Our approach performs as well or better than existing methods in terms of sensitivity and specificity at the region level (which could correspond to a gene) and at the individual variant level. BeviMed can analyse 10,000 samples genotyped at 100 rare variant sites with minor allele frequencies averaging 2.5% in under a second, enabling fast genome-wide rare variant evaluation. We show the results of analysing a dataset of 5,000 whole-genome sequenced samples from patients with diverse rare diseases.

563T

On the difference between genealogical and genetic ancestors. *M. Jeanpierre[1,2].* 1) Université Paris Descartes, Paris, France; 2) Genetics Dept, Assistance Publique-Hôpitaux de Paris, Paris, France.

   The correct interpretation of genetic variation requires an accurate estimation of the penetrance of disease-causing variants. Such estimates are dependent on the context in which they are obtained and are sensitive to sampling biases due to the clustering of cases in highly affected families. Low-penetrance variants may be analyzed as negatively selected traits, through ancestral graphs representing the common ancestry of variant carriers. The expected number of ancestors for a given individual doubles with each generation, until the simple dichotomous division is interrupted by the occurrence of a common ancestor. However, this simple pattern does not reflect the highly variable size of the chromosome segments actually inherited. The distribution of the mosaic of inherited segments can be visualized with a simple analytical model. The highly asymmetric distribution of the lengths of genetic contributions implies that some ancestors may not actually have contributed anything to the genetic make-up of contemporary chromosomes. The difference between the mean numbers of genetic ancestors (defined as those actually contributing to distant offspring) and genealogical ancestors (some of whom may have made no contribution to the genome of the descendants considered) becomes more marked the further back in time we look, because the number of genetic ancestors increases linearly with the number of generations, whereas the number of genealogical ancestors increases exponentially. New variants are annotated with existing pathogenicity prediction tools. The reuse of recently identified variants as a training set thus presents a real risk of circularity. Assessments of pathogenicity from the network of ancestral links may provide an independent source of experimental evidence, particularly for the weakest variants, because some of these ambiguous variants may reach moderate frequencies. However, it remains challenging to understand the birth and death of mutations subject to weak selection, because human populations are not in equilibrium. Here is shown that algebraic expressions exploiting subtree structures provide an analytical representation of pathogenicity that is useful in practice, because stochastic methods are not suitable for the analysis of large datasets. Contemporary chromosomes are mosaics of ancestral genomes, with highly asymmetric distributions of information that become increasingly apparent the further back into the past we delve. .

564F

Reference-based phasing using the Haplotype Reference Consortium panel. *P. Loh[1,2], P. Danecek[3], P. Palamara[1,2], C. Fuchsberger[4,5], Y.A. Reshef[6], H.K. Finucane[1,7], S. Schoenherr[8], L. Forer[8], S. McCarthy[3], G.R. Abecasis[5], R. Durbin[3], A.L. Price[1,2,9].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK; 4) Center for Biomedicine, European Academy of Bozen/Bolzano (EURAC), affiliated to the University of Lübeck, Via Galvani 31, Bolzano 39100, Italy; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 6) Department of Computer Science, Harvard University, Cambridge, MA; 7) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA; 8) Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck 6020, Austria; 9) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

   Haplotype phasing is a central problem in human genetics. Over the past decade, phasing has most commonly been performed via statistical methods applied within a genotyped cohort. In general, the accuracy of statistical phasing methods increases steadily with sample size; as such, we and others have recently developed methods that achieve very high statistical phasing accuracy in very large cohorts ($N>$100,000). However, for smaller cohorts, accuracy of cohort-based statistical phasing is fundamentally limited by the quantity of data available. Here, we explore an alternative paradigm, reference-based phasing, which can achieve high accuracy even in smaller cohorts by leveraging information from an external reference panel. We present a new reference-based phasing algorithm, Eagle2, which we have incorporated into the Sanger Imputation Service and the Michigan Imputation Server to perform free reference-phasing using the 32,470-sample Haplotype Reference Consortium (HRC) panel. This approach achieves >2x improved phasing accuracy over publicly available alternatives when used to phase small European-ancestry cohorts, with smaller improvements for larger cohort sizes. The Eagle2 algorithm represents a substantial computational advance over existing reference-based phasing algorithms: Eagle2 achieves a 20x speedup over SHAPEIT2—i.e., genome-wide phasing in 1.5 minutes per sample—with a 10% improvement in accuracy across a range of ancestries. (SHAPEIT2 is not available for reference-based phasing on public HRC servers due to licensing restrictions.) Eagle2 achieves this performance via two key ideas that distinguish it from previous phasing algorithms: a new data structure based on the positional Burrows-Wheeler transform and a rapid search algorithm that explores only the most relevant phase paths through a hidden Markov model (HMM). We have released Eagle2 as open-source software. We note that Eagle2 targets a complementary user group compared to very recent work on phasing very large cohorts. In particular, our Eagle1 method (Loh et al. 2016 Nat Genet) is targeted at phasing very large ($N>$100,000) cohorts and achieves much lower accuracy than both Eagle2 and previous methods when used to phase smaller cohorts. Likewise, the SHAPEIT3 method (O'Connell et al. in press Nat Genet), which is unreleased at the time of this writing, is targeted at phasing "biobank scale datasets.".

565W

Estimating regional heritability in the presence of linkage disequilibrium.
L. Gai[1], S. Mangul[1], E. Kang[1], S. Eyheramendy[2], E. Eskin[1,3]. 1) Department
of Computer Science, UCLA, Los Angeles, CA; 2) Department of Statistics,
Pontificia Universidad Catolica de Chile; 3) Department of Human Genetics,
UCLA, Los Angeles, CA.

The study of expression quantitative trait loci (eQTL) has emerged as a
promising area of research for understanding the mechanisms behind many
complex diseases and traits. Recently, there has been increased interest in the
genetic architecture of gene expression, which can be studied by estimating
the contribution of genomic regions to heritability. However, methods for par-
titioning heritability typically assume variants in each region are not in linkage
disequilibrium (LD) with variants in other regions. This assumption is violated
if proximal variants are partitioned into different regions, as is often the case
when partitioning by functional annotation or variant frequency. This can lead
to inaccurate estimates of a region's contribution to heritability. We present an
alternative partitioning scheme, Local Heritability using Linear Mixed Models
(locLMM) which accounts for the inherent ambiguity arising from the presence
of LD. We demonstrate that locLMM provides accurate estimates of the lower
bound of regional heritability and its application to eQTL data.

566T

**Using shared IBD haplotypes in unrelated samples to obtain near unbi-
ased estimates of the full narrow-sense heritability.** *M. Keller[1], L. Evans[1],
R. Tahmasbi[1], T. de Candia[1], D. Bjelland[1], J. Yang[2], M. Goddard[3], P. Visscher[2].*
1) Inst Behavioral Genet, Univ Colorado, Boulder, Boulder, CO; 2) Queensland
Brain Institute, University of Queensland, Brisbane, 4072, Queensland, Austra-
lia; 3) Faculty of Veterinary and Agricultural Science, University of Melbourne,
Parkville, Victoria, Australia.

Despite 50 years of estimates of narrow-sense heritability (h^2) from twin
and family studies, and more recently from unrelated individuals using
genome-wide SNP data, the true levels of h^2 remain elusive. Twin and family
studies over-estimate h^2 to the degree that non-additive genetic and common
environmental factors inflate phenotypic similarity between close relatives,
while models based on SNP similarity only capture the effects of (typically
common) causal variants tagged by SNPs on arrays. The use of IBD sharing
has been proposed as a method to derive estimates of h^2 due to both common
and rare causal variants; if this method is done on a sample of unrelated
individuals, the estimates are not biased by factors shared by close relatives.
Here, we use real whole genome sequences to thoroughly investigate the
behavior of h^2 estimates based on IBD sharing. We simulated phenotypes
of different allelic spectra (from common to very rare) and different levels of
population stratification from ~21K whole genome sequences from the Haplo-
type Reference Consortium. We simulated SNP array data by extracting from
this data ~500K common SNPs found on the UK Biobank Axiom Array. We
calculated genomic relationship matrices (GRMs) from IBD shared segments
estimated from computationally phased Axiom SNPs, and estimated h^2 using
random effects models. GRMs based on longer IBD segments provide
increasingly unbiased h^2 estimates. Once IBD segments are ~4 centimorgans
(cM) in length, this method accurately estimates all of the h^2 except that due
to the rarest causal variants (those observed < 3 times in the sample). GRMs
based on IBD segments > 4 cM become increasingly less reliable and, in strat-
ified samples, lead to over-estimates of h^2. We conclude that estimates based
on IBD segments ~4 cM in homogeneous samples provide unbiased estimates
of the h^2 due to all but extremely rare causal variants, and the method provides
sufficiently reliable h^2 estimates to be used in large sample SNP array data
in outbred populations (predicted standard errors of ~.03 in 120K unrelated
individuals), such as that currently being collected by the UK Biobank. We
believe that this is the most thorough investigation to date of the behavior of h^2
estimates from IBD sharing, and our results provide a clear roadmap for how
near full h^2 can be estimated in unrelated samples in a way that is not biased
by non-additive genetic and environmental factors shared by close relatives.

567F

Calculating the statistical significance of rare variants causal for Mendelian and complex disorders. *A.R. Rao, S.F. Nelson.* David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA.

To calculate the significance of finding a rare protein-altering variant in a given gene, one must know the frequency of seeing a variant in the general population that is at least as damaging as the variant in question. Based on data from 2504 individuals in the 1000 Genomes Project dataset, we calculated the number of individuals who have a rare variant in a given gene for numerous filtering threshold scenarios, which may be used for calculating the significance of an observed rare variant being causal for disease. Examining the dataset, we find that fewer individuals tend to carry rare, heterozygous loss-of-function (LOF) variants in genes that are known to be causal for Mendelian disorders, and fewer yet carry variants in genes essential for the survival of human cell lines. Additionally, we provide data on the number of individuals with rare variants in genic regions mapping to protein domains, which is useful information to have when vetting putative variants found in certain regions of a gene. We also apply our methods to previous findings in 19 next-generation sequencing studies and demonstrate that incorporating mutational burden data would have provided statistical evidence in support of the causal gene in all studies, with the benefit of being easy to compute from our dataset, as opposed to running simulations of populations and estimated mutation rates. If we use the frequency counts to rank genes based on intolerance for variation, the ranking correlates well with Z scores derived from the ExAC dataset ($\rho = 0.485$), with the benefit of being directly interpretable and useful for developing a quantitative, statistics-based approach for presenting clinical findings. We describe methods to use the mutational burden data for calculating the significance of observing rare variants in a given proportion of sequenced individuals, and we make the datasets, collectively named Significance of Rare VAriants (SORVA), freely available for download or querying at https://sorva.genome.ucla.edu.

568W

Maternal-fetal genotype interaction testing: Human birth weight and reproductive immunity. *M.M. Clark[1], O. Chazara[2,3], E.M. Sobel[4], H.K. Gjessing[5], P. Magnus[5], A. Moffett[2,3], J.S. Sinsheimer[1,4,6].* 1) Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA; 2) Department of Pathology, University of Cambridge, Cambridge, United Kingdom; 3) Center for Trophoblast Research, University of Cambridge, Cambridge, United Kingdom; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 5) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; 6) Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA.

The interplay between mother and offspring gene products is often neglected in epidemiological studies of human complex traits. One likely example of a maternal-fetal genotype (MFG) interaction is due to the direct contact of semiallogenic cells at the site of placentation. During pregnancy, invasive trophoblast cells emerge from the placenta into the uterine wall and assist in remodeling the uterine vasculature to provide an adequate blood supply to the fetus. Killer cell immunoglobulin-like receptors (KIRs) on the surface of uterine NK cells produced by the maternal immune system bind to HLA-C molecules expressed by the invasive trophoblasts from the fetus. Previous work (Hiby *et al.*, 2014) found evidence of an association between variation in normal birth weight and an interaction of maternal *KIR* and maternal and offspring *HLA-C*, but models involving the joint but independent effects of *KIR* and *HLA-C* could not be compared. Although a number of statistical approaches exist to test for the association of maternal-fetal genotype (MFG) interactions and disease, very few approaches exist to test for their association with quantitative traits. These approaches have been limited to case-parent trios, cannot easily account for covariates, and parameter interpretations can be unclear. To address these limitations and motivated by the hypothesis that both *KIR* and *HLA-C* jointly affect birth weight, we developed the multi-locus Quantitative-MFG (QMFG) test, using an underlying linear mixed effect model (LMM). We demonstrate the QMFG test's statistical properties with simulation studies. We also show that standard methods that focus only on the offspring genotype will miss these effects. We then reexamine the association of maternal *KIR*, maternal *HLA-C*, offspring *HLA-C*, and birth weight using data from a cohort of mothers and babies from a United Kingdom cohort study and from the Norwegian Mother and Child Cohort (MoBa) study. We find evidence for a *KIR-HLA-C* interaction along with main effects of maternal *KIR* and maternal and offspring *HLA-C*, indicating that these two loci may provide independent effects in addition to their interaction or they may be involved in a more complex pathway than a simple multiplicative interaction. More generally, our results show the value of using a LMM with maternal and offspring genotypes as predictors when these data are available.

569T

How to combine sequence data: Recommended meta-analysis strategies of sequencing data for association testing of rare variants. *Z. Chen, C. Fuchsberger, M. Boehnke.* Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Recent technological advances in high-throughput DNA sequencing have dramatically increased the abundance of whole-genome sequence data. These data offer new opportunities to discover disease variants but are also less standardized and depend heavily on the sequencing and variant calling protocols employed. The lack of uniformity is accentuated in the context of studying rare and low-frequency variants where an aggregated dataset is commonly used to increase the power of association tests. For array-based association studies, joint and meta-analysis strategies are well-developed but it is not yet known how best to combine sequencing studies of varying protocols. Here, we seek to quantify the differences in terms of variant identified and variant quality between joint and by-study genotype calls that used different standard protocols and compare the relative power of joint and meta-analysis of single variant tests considering these technical differences. We ran low-coverage (~5X) whole-genome sequencing data from 2772 individuals in the GoT2D study through the GotCloud and GaTK variant calling pipelines using default settings jointly and then separately in 5 sub-cohorts determined by geographic location. We compared the variants discovered between the two pipelines and then assessed the quality of their genotype calls by calculating their non-reference concordances with a set of highly confident ("gold standard") genotypes from deep exome sequence and Illumina HumanOmni2.5 array data. We then conducted single-variant association tests of type 2 diabetes (T2D) status for each of the joint and by-study call sets and combined the by-study results using fixed-effects meta-analysis. 85% of SNPs discovered in each joint and by-study call set were found by both GotCloud and GaTK pipelines while the rest were found by one pipeline but not the other. SNPs discovered by both pipelines have >99% non-reference concordance with our gold standard genotypes while SNPs found by GotCloud alone or GaTK alone have on average 97% and 71% non-reference concordances, respectively. Meta-analysis of by-study single variant association results show reduced power compared with joint analysis when covariates (sex, principal components, etc.) are included in the regression model (genetic variants versus T2D status). We propose strategies for combining sequencing studies that will offer improved power for analysis of rare and low-frequency variants.

570F

A polygenic risk modeling technique based on penalized estimation approach to incorporate linkage disequilibrium and functional information of SNPs using genome-wide association study summary-level data. *T. Chen[1], J. Shi[2], N. Chatterjee[3].* 1) Département de mathématiques et de statistique, Université Laval, Québec (Québec), Canada; 2) Biostatistics Branch, NCI/DCEG, Bethesda, MD, USA; 3) Depts Biostatistics and Medicine at the Johns Hopkins University Bloomberg School of Public Health and Johns Hopkins School of Medicine, Baltimore, MD, USA.

Polygenic risk score (PRS) has been a popular tool for genetic risk prediction of complex diseases. It can be constructed based on the summary statistics from genome-wide association studies (GWAS). Several methods have been proposed to improve the performances of PRS. The standard version is based on a set of independent SNPs with p-values less than a predefined significance level. Moreover, winner's-curse adjustments and external functional/annotation on a set of independent SNPs have been proposed to improve the performances of PRS. Another type of modification is to incorporate LD information based on Bayesian modelling technique, but without the usage of the functional knowledge. We propose a new method to incorporate both LD and functional annotation based on a penalized estimation approach. We applied our methods to GWAS summary-level data of 6 traits including height, BMI and complex diseases. Overall, the incorporation of LD enhances the prediction efficiency as 10-75% augmentation in prediction R squares compared to the standard PRS. For instance, the improvements of BMI and type 2 diabetes are from 4.3% to 6.6% and 2.1% to 3.6% respectively. The additional information of functional annotation of SNPs has only benefited certain traits such as type 2 diabetes to 4.2%. Extensive simulation studies have been conducted to illustrate the performances of the proposed method. In addition, we provide the prediction R squares obtained from other competing methods in both real data and simulation studies and the results demonstrate that our proposed methods have significant advantages.

571W

Bayesian meta-analysis for cross-phenotype genetic association study.
A. Majumdar[1], S. Bhattacharya[2], J. Witte[1]. 1) University of California, San Francisco, San Francisco, CA, USA; 2) Indian Statistical Institute, Kolkata, India.

Simultaneous analysis of genetic association across multiple traits may reveal shared genetic susceptibility among traits (pleiotropy). While assessing pleiotropy, alongside measuring the evidence of overall pleiotropic association, it is crucial to identify the traits associated with a risk locus since only a subset of the traits may have the true genetic effects. We propose a Bayesian cross-phenotype meta-analysis using a spike and slab type of prior that provides a Bayes factor measuring the global pleiotropic association and an optimal subset of traits associated with the risk locus. In the spike and slab type of prior, the spike corresponding to the null genetic effects can either be a positive mass at zero (Dirac spike), or a normal distribution with zero mean and a variance smaller than that of the normal distribution (with zero mean) representing the slab corresponding to the non-null genetic effects (continuous spike). Gibbs samplers are designed for these two types of priors in the case of both uncorrelated and correlated summary statistics corresponding to multiple traits. We demonstrate by simulations for multiple normally distributed traits that a continuous spike is an overall better alternative than Dirac spike for drawing inference. This meta analysis approach allows heterogeneity in both the direction and magnitude of the genetic effects across traits. Unlike a frequentist method, it also estimates the posterior probability of a trait being associated with a risk locus. It is applicable to both cohort data and separate studies of multiple traits having overlapping or non-overlapping subjects. For strongly correlated summary statistics, the Gibbs sampler can suffer from poor mixing due to multi-modality of the posterior distribution of the model parameters. To address this issue, we propose an empirical Bayes version of the meta analysis by leveraging the information obtained by a false discovery rate (FDR) controlling procedure using the same data. Simulations show that in common scenarios, our method produces better accuracy in the selection of traits than a subset-based meta analysis ASSET (Bhattacharjee et al., 2012) with respect to both specificity (correctly discarding null genetic effects) and sensitivity (correctly including non-null genetic effects). Finally we analyze 22 traits in the large Kaiser cohort using our proposed approach.

572T

Region-based association tests for sequencing data on survival traits. *Y. Chiu[1], L. Chien[2], D. Bowden[3].* 1) National Health Research Institutes, Zhunan, Miaoli 35053, Taiwan, ROC; 2) Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, ROC; 3) Center for Diabetes Research, Center for Genomics and Personalized Medicine Research, Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.

Identification of functional variants has been one of the important steps to understand the genetic basis of complex diseases. Family-based designs enriched with affected subjects and disease associated variants can increase statistical power for identifying functional rare variants. However, few rare variant analysis approaches are available for time-to-event traits in family designs. We developed novel pedigree-based burden and kernel association tests for time-to-event outcomes with right censoring for pedigree data, referred to FamRAS (Family-based Rare variant Association tests for Survival traits). Cox proportional hazard models were employed to relate a time-to-event trait with rare variants (MAF < 0.05) with flexibility to encompass all ranges and collapsing of multiple variants. In addition, the robustness of violating proportional hazard assumptions was investigated for the proposed and four current existing tests including the conventional population-based Cox proportional model and the burden, kernel and sum of squares statistic (SSQ) tests for family data. The proposed tests can be applied to large-scale whole-genome sequencing data. They are appropriate for the practical use under a wide range of misspecified Cox models, as well as for population-based, pedigree-based or hybrid designs. In our extensive simulation study and data example, we showed that the proposed kernel test is the most powerful and robust choice among the proposed burden test and the existing four rare variant survival association tests. When applied to the Diabetes Heart Study, the proposed tests identified exome variants of the *JAK1* gene on chromosome 1 to be associated with age at onset of type 2 diabetes ($P < 4.82 \times 10^{-5}$, the Bonferroni significance threshold for testing multiple genes on chromosome 1).

573F

A novel approach to genetic association testing: The Conditional on Opposite Alleles Test (COAT). *B. Maher.* Johns Hopkins School of Public Health, Baltimore, MD.

The topic of allelic versus genotypic models for genetic association testing using the usual statistical models has been explored by a number of investigators over the past few decades. Generally, the results of these investigations have lead to widespread use of genotypic dosage tests in spite of some advantages of allelic tests. Presented here is a set of tests that is a middle ground between genotypic and allelic tests, referred to as the Conditional on Opposite Alleles Tests (COAT). The basis for COAT is to model alleles at a SNP as paired data, not as fully indistinguishable (as in genotype tests) or as fully independent (as in allelic tests). This is accomplished using mixed models, allowing for random intercepts, to account for the effect of the opposite allele in the pair, or slopes, to allow for non-additive effects or by adjusting for the effect of the opposite allele in a linear model. COAT is compared via simulation to common existing allelic and genotypic tests and to other novel approaches. Under the models tested, COAT using a random intercept maintains or exceeds the power of the usual approaches under additivity and exhibits increases in power to detect association when dominance is present. Overall, COAT is an important alternative to existing tests, is valid for discrete and continuous traits, and is easily extended to family-based data. .

574W

Effects of imputation on combined admixture and association mapping.
D. Yorgov[1], S.A. Santorico[1,2,3]. 1) Mathematical and Statistical Sciences, University of Colorado, Denver, USA; 2) Human Medical Genetics and Genomics Program, University of Colorado School of Medicine, Aurora, Colorado, USA; 3) Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado, Aurora, Colorado, USA.

For populations like African Americans and Latinos, the long-range correlations introduced by the admixture process allow for gene mapping methods that combine genotype with ancestral origin information (local ancestry). In several studies, combining admixture and association information resulted in findings that were not detectable with standard association testing, suggesting that for a genome-wide association study (GWAS) power can be gained from combining two different sources of genetic information. Using extensive simulations based on real Latino genomes, we investigate the possible gains in power from different ways of incorporating local ancestry into single variant association testing at GWAS chip resolution. We compare such approaches to imputation followed by a standard association test. We simulate polygenic traits with a single causal common variant per locus, and a causal allele with the same effect regardless of its ancestral origin. Markers with sufficient degree of allele frequency differentiation in the ancestral populations that do benefit from the admixture information were (1) those with differential linkage disequilibrium patterns in the ancestral populations (for combined tests) and (2) those with causal allele originating largely from one of the ancestral populations (for admixture mapping). For this simulation scenario our results suggest that, at GWAS chip resolution, there is limited benefit from incorporating local ancestry in admixture mapping or in combined admixture and association testing since higher power can be achieved by the imputation approach. Specifically, imputation of the causal marker followed by association testing was the best approach with respect to power both on average across all regions containing causal markers and individually at each region. Imputation and association yielded increases of the average power compared to the best powered admixture and combined tests by factors of 5.3 and 2.13, respectively. We further show that the standard linear mixed model approach, without local ancestry adjustment through a fixed effect, controls well for Type I error in both the association test and in the combined admixture and association tests. Additional simulation scenarios that allow for allelic heterogeneity are being considered to assess the utility of a combined approach as a complementary scan when high resolution genotypes are available via sequencing or imputation.

575T

A statistical method for phenotype-genotype association that is robust to sequencing misclassification. *L. Zhou[1], D. Londono[1], A. Musolf[2], T.C. Matise[1], D. Gordon[1].* 1) Department of Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Regions of the human genome where genotype frequencies significantly differ among cases and control groups may be regions that harbor disease loci. The purpose of this study is to develop a statistical test of association between multi-locus genotype (MLG) frequencies and a disease phenotype in a case/control study utilizing Next Generation Sequencing (NGS) data. This statistic is a likelihood ratio test whose asymptotic null distribution is central chi-square (degrees of freedom dependent upon the number of MLGs). We designed the statistic to be robust to differential sequencing misclassification (sequencing errors) in NGS variant calls. The parameters utilized in the test are (for each individual): (i) the observed alternative read counts at the set of SNPs; (ii) corresponding sequencing coverage at the SNPs; and (iii) phenotypes. Maximum likelihood estimates of MLG frequencies, sequencing errors and log-likelihoods are determined by an expectation-maximization (EM) algorithm. We apply permutation and bootstrapping to assess the type I error and power. We also use the factorial design to determine the performance of this statistic under specific genetic-model parameter situations. We find that this method maintains the correct type I error rate in simulation data and the 1000 Genome data for permutation with significant levels of 1%, 5%, and 10%.

576F

Quality control analysis of the 1000 Genomes Project Omni2.5 genotypes. *N.M. Roslin[1], W. Li[1], A.D. Paterson[1,2,3], L.J. Strug[1,2,4].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 3) Epidemiology and Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 4) Biostatistics and Statistics, University of Toronto, Toronto, ON, Canada.

The 1000 Genomes Project genotyped 2318 individuals (48.1% male) from 19 populations in 5 continental groups on the Illumina Omni2.5 platform. The data are publicly available, and will prove a valuable resource to obtain ethnic-specific allele frequencies, as well as exploring population histories through principal components analysis (PCA), estimation of inbreeding coefficients, and admixture analysis. As in any study, the data should be cleaned prior to analysis, to remove individuals or markers of questionable quality. Furthermore, a thorough understanding of the relationships between individuals must be established. Here we report our findings after comprehensive examination of the data for quality control. The basic quality of the genotypes was assessed using standard procedures. KING version 1.4 was used to confirm the relationships in the provided pedigrees, and also to detect undeclared relationships. PCA was used to examine the similarities and differences between individuals among and between population groups. In general, the data was found to be of high quality. No samples were removed due to low call rate (<97%), or excess heterozygosity. Sex chromosome genotypes showed one individual with a discrepancy between reported and inferred sex, and were unable to determine sex in an additional 20 individuals; the sex for these was changed to unknown. Relationship checking found discrepancies between first-degree relationships in the provided pedigrees and the genotypes in 9 families, including one instance where a reported parent/child pair was unrelated, two instances where full sibs were unrelated, and one set of three individuals who formed a newly defined trio. A set of 1756 individuals who were inferred to be no closer than 4th degree relatives was extracted and used in PCA. These individuals clustered in a pattern that is consistent with other published reports of global populations. We identified 4 individuals whose genotypes clustered more closely with a different geographic region than the one in the provided data. Although the genotype data is of high quality, errors exist in the publicly available dataset that require attention prior to using the genotypes. A list of SNPs with good quality metrics, as well as a revised pedigree file and a list of unrelated individuals is available at http://tcag.ca/.

577W

Quantifying the reproducibility/repeatability of a multivariate assay. *R. Makowsky, B.J. LaFleur.* Biostatistics, HTG Molecular, Tucson, AZ.

Establishing diagnostic Reproducibility / Repeatability (RR) of highly multivariate assays is a challenging statistical problem. One common approach is to do multiple pairwise comparisons and assess the correlation coefficients across various conditions (e.g., instruments, lots, etc.). This approach is statistically unsatisfactory in that pairwise correlation is a measure of linear association, and not a measure of reproducibility. Interclass correlation coefficients derived from variance components are a suitable alternative, but it is not obvious how to utilize them in a highly multivariate setting. We propose a general strategy for how to conduct, visualize, and evaluate such an analysis with high-dimensional outcomes. We exemplify our approach using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA). This assay measures the expression of 2,084 human miRNA transcripts using next generation sequencing. For this study, we were interested in quantifying the RR of the assay when run on different days and processors using eight replicates of three different Sample types per combination of Day and Processor. We show that the ICC values for Day and Processor were consistently low across probes (each < 1% on average), with Sample Type accounting for the majority of the observed variance. Additionally, we demonstrate how this approach is invariant to standard transformation (CPM) and normalization (Median, as proposed by Ander and Huber, 2010, and implemented in Bioconductor packages DESeq / DESeq2) approaches. In summary, we found that this analytic approach provides a useful quantification of RR, is easy to visualize and summarize for high dimensional data, and correctly accounts for the intended use population's variability.

578T

A unifying statistical test for privacy in the era of large-scale genomic and medical data sharing. *S. Prabhu, C. Bustamante.* Department of Genetics and Biomedical Data Science, Stanford University, Stanford, CA.

Widespread sharing of genomic and clinical datasets between researchers is necessary to unleash the personalized medicine revolution. This rationale has motivated large-scale international projects like the NIH's PMI-CP and Genomics England's 100K genomes, among others. However, proposals to disseminate sensitive biological data have raised pressing questions about privacy – particularly the risk of study partyicipant re-identification. The concerns are valid, and they need to be addressed immediately. Current guidelines for mitigating risk largely amount to a two-pronged approach: (a) permit broad access to study-wide summary statistics, while (b) control and limit access to individual-level data. From a statistical privacy standpoint, these are known to be inadequate. Multiple studies have already established that at least some quantitative summaries (allele frequency in a dataset, allele carrier status in a dataset) are vulnerable to attack, while the safety of other summaries (effect sizes, p-values) remains tentative. Changes to federal policy in response to these attacks have been ad-hoc and reactionary (e.g. stop sharing allele frequencies), rather than quantitative, principled and measured. This confusion stems from multiple definitions of privacy, utility, and attack techniques. In light of these anecdotal demonstrations, there is an urgent need for an overarching statistical framework in which various publication schemes might be assessed, risk to participant privacy (from both existing and future attacks) might be measured, and appropriate policies might be devised. Here, we present such a unifying framework. Specifically, we unify statistical attacks based on Hypothesis Tests and Bayesian formulations (popular in the genomics and statistics communities), with the widely adopted Differential Privacy definition (popular in the computer science community). We present the Model Indistinguishability Criterion (MIC), a new information theoretic metric that conveys a natural and intuitive meaning of risk to a study participant. The MIC is almost assumption free, and offers a practical means for data stewards to test whether their publications are secure. We characterize the connection between MIC and well-established model selection criteria like AIC (Akaike) and BIC (Bayes). To the best of our knowledge, this is the first body of work to comprehensively integrate the major frameworks of statistical privacy analysis used across multiple disciplines.

579F

Characterizing the health phenome in communities of distantly related individuals in a multi-ethnic biobank in New York City. *G.M. Belbin[1,2], B.S. Glicksberg[1,3], N.D. Beckmann[1,3], G. Nadkarni[1,2], D. Park[4], M.C. Yee[5], S. Ellis[1,2], A. Auton[6], J. Cho[1,2,3], R.J.F. Loos[1,2], N. Abul-husn[1,2], N. Zaitlen[7], E.E. Kenny[1,2,3].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mt Sinai, New York, NY; 3) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mt Sinai, New York, NY; 4) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA; 5) Carnegie Institution for Science, Dept. of Plant Biology, Stanford, CA; 6) Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Bronx, New York, NY; 7) Department of Medicine, University of California, San Francisco, CA.

Shifting patterns of demography may profoundly shape population health landscapes. Yet in US healthcare systems demography is often crudely captured, and some populations are often misclassified *e.g.* Hispanic/Latinos. Here we propose using genomic data in a healthcare system in New York City (NYC) to detect fine-scale population structure and evaluate whether it is informative for the population health phenome. We utilized the Mount Sinai Bio*Me* Biobank, a multi-ethnic biobank, where 21665 participants with genotype data are linked to Electronic Health Records (EHR). To detect finescale population structure we created a network of distant relatedness between pairs of individuals in Bio*Me via* haplotypes shared identical-by-descent (>3cM). Using a machine learning method for community detection we find that 83% of participants fall into one of 36 distinct communities. Using ethnicity and place of birth data provided by participants, we find that these communities reflect recent, shared ancestry at country and regional levels. We showed >90% sensitivity and specificity to detect several groups including self-reported Jewish (N=808), participants born in Puerto Rico (PR; N=2702), Dominican Republic (DR; N=1695) and Japan or Korea (N=138). Other communities showed high specificity (>95%) but lower sensitivity (<60%) *i.e.* participants born in Mexico (N=188) or Colombia (N=138). We find multiple communities enriched for people born in Ecuador and India, perhaps reflecting substructure in those populations. We also predict finescale demographic ancestry for 9,600 participants born in NYC *via* their community membership. Next we show that some communities are significantly enriched for EHR derived billing codes. Top signals recapitulate known differential prevalences of traits. *e.g* sickle cell trait (p<5e-17) in African-Americans (AA); and diseases, *e.g* chronic kidney disease (p<2e-23) in AA, asthma (p<4e-193) in PR, type 2 diabetes (p<0.03) in Mexican, and Hepatitis B (p<1e-24) in East Asian communities. Other signals indicate previously unreported community-specific disease risks that may be due to shared genetic or environmental factors *e.g.* enrichment of peripheral vascular disease in the DR community (p<1e-46). Integrating genomic data into healthcare systems can improve our understanding of population health phenomes, lead to previously unrecognized health conditions enriched within communities and will be an important step forward for precision medicine.

580W

Kinship estimation based on extremely low-coverage sequencing data. *J. Dou[1], S. Chothani[1], X. Sim[2], J. Hughes[3], D. Reilly[3], E. Tai[2], J. Liu[1], C. Wang[1].* 1) Genome Institute of Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 3) Merck Research Laboratories, Kenilworth, New Jersey, United States.

Estimation of kinship is important to genetic association studies, both for control of cryptic relatedness to avoid spurious associations and for estimation of trait heritability. However, estimation of kinship is challenging for target sequencing studies, where sequencing efforts focus on small regions of the genome. Existing methods often assume accurate genotypes at a large number of markers across the genome. We show that these methods, without accounting for the genotype uncertainty in shallow sequencing data, can yield a strong downward bias in kinship estimation. We develop a novel method that explicitly models the uncertainty and thus reduces the bias. Our method includes kinship estimators for homogeneous samples and for samples with population structure and admixture. We tested the method on a target sequencing dataset of Singapore Chinese and Malays, which includes many admixed individuals. We show that our method, with imputation to the 1000 Genomes data, can accurately estimate kinship coefficients using the off-target sequencing data at ~0.1x depth. Our method enables control of cryptic relatedness in target sequencing studies without additional genotyping data. We have implemented the method in a fast and memory efficient C++ program called SEEKIN. .

581T

DNAForest: Rapid genome-wide reconstruction of coalescent trees from thousands of phased genomes. *H.M. Kang[1], F. Zhang[2].* 1) Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, MI; 2) Department of Computational Medicine & Bioinformatics, Univ Michigan, Ann Arbor, Ann Arbor, MI.

Recent advances of sequencing technologies and statistical phasing methods now allow us to accurately phase diploid genomes. Even with accurate haplotype map, inferring coalescent tree of diploid genome is still one of the most challenging problems in human genetics due to stochastic mutation and recombination procedures make the time complexity of the algorithm to increase exponentially to the sample sizes. If coalescent tree of individual haplotypes could be accurately reconstructed for every individual genomic position, it will allow us to expand our ability to precisely understand the history of individual genomes and individual genetic variants. We propose a novel algorithm DNAForest, that allows us to rapidly approximate coalescent tree at every variant position in a scalable way. DNAForest leverages the data structure implemented for the Positional Burrows-Wheeler Transform (PBWT) algorithm. PBWT enables extremely compact encoding of haplotypes by compressing prefix- or suffix-sorted haplotypes. In DNAForest, we use both prefix- and suffix- ordered haplotypes together to rapidly approximate the coalescent tree using 2-dimensional hierarchical clustering. The pairwise distance between haplotypes are calculated from matching haplotype lengths, and the the time complexity of hierarchical clustering can be reduced from $O(n^2)$ to $O(n \log n)$ the sorted nature of PBWT. Because PBWT provides compact storage of haplotypes, the actual computational gain is even larger due to reduce storage and memory bottleneck. Adjacent coalescent trees can be reconciled together to refine the colescent tree or to generate ancestral recombination graph (ARG). We constructed coalescent tree of ~5,000 phased haplotypes from the 1000 Genomes project. As expected, a subset of African population most belongs to an outgroup clades while the other almost always clustered together with Admixed Africans. We also observed many clades consisting of hundreds of haplotypes mixed between European, Asian, and African ancestry, suggesting shared ancestry within the specific region of genome. As expected, we observed that time to most recent common ancestor (MRCA) are extremely small for most pairs of individual within the same clade, but they are orders of magnitude distant between individuals from different clades. We also observe that DNAForest improves statistical phasing beyond existing reference panel. DNAForest also allows us to estimate the age of each variant precisely.

582F

RVMMAT: Rare-Variant Mixed Model Association Tests for binary traits in structured and related samples. *H. Chen, X. Lin.* Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

We develop the computationally efficient Rare-Variant Mixed Model Association Tests (RVMMAT) for binary traits in structured and related samples. With the advance in next-generation sequencing technology, statistical methods for testing genetic association with rare variants have been proposed and widely applied to unrelated samples. These methods are also known as gene-based or SNP-set tests, since rare variants are often grouped by genes or genomic regions in the analysis. The burden test and sequence kernel association test (SKAT) are two widely applied rare variant tests. Here we propose and implement the burden test, SKAT and a novel combined test in RVMMAT. All tests share the same null generalized linear mixed model, which only needs to be fitted once in a whole-genome analysis. We show in simulation studies that the proposed tests control correct type I error rates in the presence of population stratification and cryptic relatedness, in both single-cohort studies and meta-analysis. We compare the power of these tests in various scenarios and illustrate how they can be used to test a broad class of different scientific hypotheses in large-scale sequencing studies. We also apply RVMMAT to a real data whole-exome analysis.

583W

Exploring ancestry of your sample using genotypes or sequence read data: The LASER server. *D. Taliun[1], S.P. Chothani[2], S. Schönherr[3], L. Forer[3], M. Boehnke[1], G. Abecasis[1], C. Wang[2].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Computational and Systems Biology, Genome Institute of Singapore, Singapore; 3) Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria.

Recent genetic association studies focus on detecting association with rare variants, an analysis that typically requires large samples to achieve reasonable statistical power. Augmenting case samples with controls from other studies is an efficient, cost effective way to improve statistical power. However, external controls that have different ancestries from cases can lead to false positive association results. Ancestry matching between cases and external controls can be challenging due to possible differences in available array genotyping or sequencing data and data sharing restrictions. We introduce the LASER server that allows ancestry comparison for samples from multiple studies without pooling raw data together and even when samples are genotyped with different arrays, sequenced at only a few genes or the exome, or genome sequenced at different depths. The server is based on our LASER method (Wang et al., 2014, 2015) that uses a set of reference samples with known ancestry to compute a reference principal component space. LASER places all study samples in the same reference principal component space enabling direct comparison of ancestries. LASER allows studies to share only principal components coordinates for ancestry matching as long as the same ancestry reference panel was used. Our most recent implementation of the LASER server allows the user to choose between three ancestry reference panels: 1) the Human Genome Diversity Panel covering 53 worldwide populations; 2) the POPRES panel covering 37 European populations; and 3) an Asian panel covering 43 Asian populations. Users can upload genotype data or sequence reads in a web browser, or by giving restricted access to their SFTP server. Ancestry estimation results can be quickly and easy explored using integrated 2D and 3D interactive web visualizations, where users can instantaneously get information about selected individuals and the ancestry composition of their K nearest-ancestry neighbors. The LASER server can be accessed at http://laser.sph.umich.edu.

584T

Matching controls for independently ascertained case cohorts in population based association analyses. *C. Van Hout, C. O'Dushlaine, B. Ye, F. Dewey, A. Baras, A. Shuldiner, I. Borecki.* Regeneron Genetics Center, Tarrytown, NY.

Canonical design for case-control genetic association studies includes recruitment of cases and controls from the same population. In practice, patient care providers and specialty clinics often focus on specific diseases, collecting rich data on affected individuals, cases, but often not from controls. To enable genetic discovery using cases from independently ascertained cohorts, it is essential to identify suitable controls from other studies or public resources. Controlling for confounding due to population stratification to limit type 1 error in matching controls to cases is crucial. While increasing the ratio of controls to cases offers increased power to detect truly associated variants, sub-optimal matching may introduce population stratification. Furthermore, practices to control inflation may differ between common and rare variants, and may depend on case:control ratios. We compared several approaches to matching controls to cases: random matching, IBS matching, and principal components (PCs) based matching. We evaluated the performance of these approaches under the null hypothesis of no true association using using exome sequence data from DiscovEHR cohort from Geisinger Health System (GHS) in Northern Pennsylvania, and Dallas Heart Study (DHS) cohorts. We selected 500 individuals at random from 1,329 putatively unrelated participants of European ancestry from DHS as cases, and matched controls from 38,333 GHS participants of European ancestry. We compared matching approaches at case:control ratios ranging from 1:1 to 1:10, and measured control of inflation using regression based λGC within various Minor Allele Count (MAC) bins. To evaluate sampling variation, we evaluated 100 replicates for each configuration. Control of inflation was similar using IBS and PC based matching, offering substantial improvement over random matching. In general, we observed increased inflation for higher ratios of controls:cases. However, there was no superior matching method across all MAC bins. Importantly, by comparing case:control matching within GHS, i.e., no true population stratification, we observed systematic decrease in λGC in rarer MAC bins, likely driven by data sparseness. Accounting for these features and confounders in evaluations of matching approaches will be important as high throughput sequencing in independently ascertained case cohorts becomes increasingly available.

585F

Estimating population-specific FST. *Q. Zhang[1,2], B. Weir[2].* 1) Medicine, University of Washington, Seattle, WA; 2) Biostatistics, University of Washington, Seattle, WA.

   Population-specific FST is a parameter relevant to forensics, the detection of natural selection, and the adjustment of population structure in genome-wide association studies. A frequentist method of moments estimator (MME) is commonly used to estimate this parameter, and we prove that it is unbiased assuming an island model. By incorporating a likelihood and priors to generate a posterior distribution, a Bayesian estimator may be preferable. However, existing Bayesian approaches for estimating this parameter generally do not account for dependence between populations, which arises when there is migration between populations. We investigate a new Bayesian estimator of population-specific FST that accounts for dependence between populations. With the software GENOME, we simulated genetic data according to an island model with random mating between two populations, mutation, and migration. For each combination of population genetic parameter values, we simulated 500 data sets and calculated the MME, an existing Bayesian estimator, and our proposed Bayesian estimator on each data set. To obtain Bayesian posterior mean estimates, we implemented and extensively checked a Gibbs sampler with accept-reject algorithm to draw from full conditional distributions, and a Metropolis within Gibbs algorithm. We tested for convergence of MCMC chains using the Geweke test. We compared the three estimators in terms of RMSE for a variety of population genetic parameter combinations.

586W

The problem of heritable PCs in multivariate data with related subjects. *M. Abney.* Dept Human Gen, Univ Chicago, Chicago, IL.

   Powerful analysis of multivariate data (e.g. gene expression) is hampered by the presence of hidden confounders and other unknown variates. Approaches for discovering these confounders, such as principal components analysis (PCA) assume samples are independent. This assumption, however, is violated when the sample is from a population with elevated kinship, e.g. isolated and/or founder populations, or one where there is familial relatedness. Applying PCA, or PCA-based methods, in these samples results in the estimated unknown variates to be a mixture of true hidden variates and genetic effects. Here, we discuss the causes and consequences of identifying hidden variates that unintentionally include a genetic component. In particular, we demonstrate this effect by applying PCA to an expression data set from an isolated population and find the first 200 PCs to have substantial heritability. We also show that using these PCs as covariates can substantially reduce the estimate of heritability of the expression traits. That is, genetic signal is being removed from expression. Using simulations we study how the heritable PC covariates affects estimates of eQTL effect size, type 1 error and power. We also propose approaches to solve these problems.

587T

Comparing total and allelic expression for mapping polymorphisms associated with modulation of expression in *cis*. *M.F. Santibanez Koref.* Insitute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom.

Modulation of transcription, in particular through changes in DNA sequence acting in *cis*, plays an important role in modifying disease susceptibility. The effects of genetic variation on expression can be assessed by comparing the total levels of a transcript, without differentiating the contributions of each allele, in individuals with different genotypes at the polymorphism of interest, and treating expression levels as a quantitative trait. Proximity to the target gene is seen as evidence for a *cis* effect. Expression is also affected by influences acting through diffusible factors in *trans*. These include environmental, genetic and technical factors and affect transcripts from both alleles. Alternative approaches for mapping *cis* acting loci on autosomes use the relative contribution of each the two alleles within an individual to total expression of the transcript of interest. This can be quantified as allelic expression ratio. Instead of comparing expression values across individuals, such approaches compares ratios across individuals. Here we use published and simulated data to compare the ability of different approaches to detect genetic variation acting in *cis*. We show that factors such as the extent of transacting variation can increase the relative advantage of using methods that rely on relative allelic expression levels. However, these procedures rely on the heterozygosity transcribed markers to differentiate between the contributions of each allele and could therefore exclude information from a substantial proportion of the material available. This shortcoming can be addressed by integrating both total and allelic data. Such methods can also be used to map *cis* and *trans* influences separately obviating the needs to use arbitrary distance criteria, and they can also help to overcome issues such as biases in the ability to detect transcripts from different alleles.

588F

Polygenic profiles for predicting risk of early menopause. *T. Laisk-Podar[1,2], A. Salumets[1,2], R. Mägi[3].* 1) Women's Clinic, University of Tartu, Tartu, Estonia; 2) Competence Centre on Health Technologies, Tartu, Estonia; 3) Estonian Genome Center, University of Tartu, Estonia.

Introduction. Reproductive aging impacts female fertility and health, and involves a considerable genetic component as evidenced by recent genome-wide association studies (GWAS). Currently, no genetic markers are used for predicting menopausal age. The aim of this study was to evaluate the applicability of polygenic risk scores (PRS) to predict the risk for early menopause (before age 45). **Material and methods.** PRS were generated using the publicly available ReproGen consortium menopausal age GWAS meta-analysis summary statistics including data for 2.4 million markers and involving approximately 70,000 women. Correlation between PRS and menopausal age was tested among 3,072 women (age at least 45) in the Estonian Biobank. Receiver operating characteristic (ROC) curves were generated to evaluate the predictive value of PRS for discriminating women with early menopause. **Results.** Polygenic risk profiles generated using different marker cut-offs were significantly correlated with age at natural menopause (p=1.7 x 10^{-24}). The same profiles enabled to predict early menopause with an AUC (area under curve) of 0.65, outperforming the predictive value of smoking status (AUC=0.54), which is one of the most important lifestyle factors affecting menopausal age. When the extremes of the PRS were compared, a 10-fold difference in risk for early menopause was observed. **Conclusion.** Polygenic risk profiles could be considered for detecting women at risk of early menopause and have the potential to become an additional tool to increase the accuracy of ovarian reserve assessment, leading to more personalized counselling regarding family planning and patient management.

589W

Multi-ethnic polygenic risk scores improve risk prediction in diverse populations. *C. Marquez-Luna[1], P. Loh[2,3], A. Price[1,2,3], The SIGMA Type 2 Diabetes Consortium.* 1) Biostatistics Department, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

Methods for genetic risk prediction have been widely investigated in recent years. However, most available training data involves European samples, and it is currently unclear how to accurately predict disease risk in other populations. Previous studies have used either training data from European samples in large sample size or training data from the target population in small sample size, but not both. Here, we introduce a multi-ethnic polygenic risk score approach, MultiPRS, that combines training data from European samples and training data from the target population. We applied MultiPRS to predict type 2 diabetes in a Latino cohort using both publicly available European summary statistics in large sample size (effective N=40,100) and Latino training data in small sample size (effective N=8,181), and observed a >70% relative improvement in prediction accuracy compared to methods that use only one source of training data, consistent with large relative improvements observed in simulations. Notably, the polygenic risk score derived from European summary statistics was strongly negatively correlated with proportion of European ancestry in the Latino cohort ($R = -0.70$); this systematically lower load of T2D risk alleles in Latino individuals with more European ancestry could be explained by polygenic selection in ancestral European and/or Native American populations. We also applied MultiPRS to predict height in a UK Biobank cohort with samples of UK European (N=113,851), South Asian (N=1,730) and African (N=1,767) ancestry. Despite the much larger European sample size, MultiPRS again significantly improved prediction accuracy in each non-European data set, with 27% to 35% relative improvements compared to single ancestry methods. In addition, the polygenic risk score derived from European summary statistics was positively correlated with proportion of European-related ancestry in South Asian (R=0.14) samples; again, the systematically higher load of alleles associated with increased height in individuals with more European ancestry could be explained by polygenic selection. Our work reduces the gap in risk prediction accuracy between European and non-European target populations.

590T

Modeling allele-specific expression by single-cell RNA-sequencing. *Y. Jiang[1], N. Zhang[1], M. Li[2].* 1) Department of Statistics, University of Pennsylvania, Philadelphia, PA; 2) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA.

The two alleles of a gene can be expressed at different levels, with monoallelic expression, in which one allele is completely silent, being the extreme case. Allele-specific expression (ASE) has been associated with disease, and to understand the mechanisms of such association, it is important to study ASE at the cellular level. ASE is traditionally studied by bulk RNA-seq, in which genes showing imbalanced expression between alleles are declared to have ASE. Although useful, bulk RNA-seq measures the average expression across many cells and is prone to confounding caused by the mixing of disparate cell types. Single-cell RNA sequencing (scRNA-seq) now offers the opportunity to study ASE at single-cell resolution; however, current scRNA-seq protocols introduce technical biases that, if not properly removed, can mislead analyses. Additionally, ASE modeling at the single-cell level needs to account for the stochasticity of gene expression as exhibited by the pervasive phenomenon of transcriptional bursting. Here we propose a framework to study ASE at the single-cell level. First, by an empirical Bayes method we classify genes into *silent*, *monoallelically expressed*, *bursty*, and *constitutively expressed*, and quantify the genomic distribution across categories. Second, we adopt a Poisson-Beta hierarchical model to estimate allele-specific transcription kinetics while accounting for dropout, amplification and sequencing bias. Third, we devise statistical procedures to test for independent firing and differential kinetics between the two alleles. We apply our method to a scRNA-seq dataset profiling mouse cells during preimplantation development. Our results indicate that: 1) single-cell ASE patterns change through embryonic stem cell differentiation, 2) the two alleles from 90% of the bursty genes fire independently and share the same transcription kinetic parameters, 3) there are six times more genes whose differential ASE are modulated by differential burst frequency than by differential burst size, 4) genes with non-independent and repulsed allelic firing are enriched for cell surface and transmembrane proteins, which are associated with cell diversity and identity. In addition to known positive controls such as X-chromosome and imprinted genes, we are currently seeking experimental validation by RNA-FISH. Collectively, our method offers a genome-wide approach to systematically study gene regulation in an allele-specific manner with single-cell resolution.

591F

Joint allele-specific analysis and molecular QTL mapping with large samples. *Y. Lee, X. Wen.* Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

Quantitative trait loci (QTL) of molecular phenotypes (e.g., gene expressions) can be identified by genetic association testing of inter-individual variations (e.g., allele-specific expression) and intra-individual variations (e.g., traditional QTL analysis). To improve the power of QTL discovery, recently several approaches have been proposed to perform joint analysis of aseQTL and expression QTL (eQTL). In eQTL mapping, majority of these approaches directly model the read counts of RNA-seq data and assume genetic effects of a causal variant measured by aseQTL and eQTL are identical. While statistical tests from combined likelihoods have shown the improvement of QTL discovery, we found that count model assumptions on eQTL analysis are inferior to the approach that performs explicit latent factor control and maps QTLs using normalized data. However, there is no principled way to perform joint analysis with aseQTL using the standard QTL mapping approach. We propose a computational approach to perform joint analysis of aseQTL and eQTL without combined likelihoods, which allows to apply standard procedures for eQTL analysis and simultaneously improve the power of QTL discovery by utilizing aseQTL information. In our approach, we analyze aseQTL and eQTL separately by applying standard procedures for eQTL analysis with normalized expression and likelihood tests for aseQTL analysis with read counts. In particular, we do not assume that genetic effect sizes measured by aseQTL and eQTL are at the same scale. However, the causal QTL should exhibit directional consistency in effects despite of effect size heterogeneity. To this end, we propose a Bayesian approach with a novel prior. This approach can be applied to analyze other molecular QTL data such as ATAC-seq. We evaluate our method by applying it to both simulated and real data. Simulation study shows joint analysis has better power on detecting genuine QTL compared to separate aseQTL and eQTL analyses in various effect sizes and minor allele frequencies. In real data analysis, we also find that joint analysis can potentially narrow the confidence set of causal variants.

592W

KCNQ1 **p.L353L affects splicing and modifies the phenotype in a founder population with long QT syndrome type 1.** *L. Arbour[1,2], J. Kapplinger[3,4], A. Erickson[2], S. Asuri[2], S. McIntosh[2], D. Tester[5], C. Kerr[6], J. Morrison[7], A. Tang[8], S. Sanatani[9], M. Ackerman[3,4,5,10].* 1) Department of Medical Genetics, University of British Columbia, Victoria, BC, Canada; 2) Division of Medical Sciences, University of Victoria, Victoria, BC, Canada; 3) Mayo Medical School, Mayo Clinic, Rochester, MN, USA; 4) Department of Molecular Pharmacology & Experimental Therapeutics, Windland Smith Rice Sudden Death Genomics Laboratory, Mayo Clinic, Rochester, MN, USA; 5) Department of Cardiovascular Diseases, Division of Heart Rhythm Services, Mayo Clinic, Rochester, MN, USA; 6) Division of Cardiology, University of British Columbia, Vancouver, BC, Canada; 7) Gitxsan Health Society, Hazelton, BC, Canada; 8) Department of Medicine, University of Western Ontario, London, Ontario, Canada; 9) Department of Pediatrics, Division of Cardiology, University of British Columbia, BC Children's Hospital, Vancouver, British Columbia, Canada; 10) Department of Pediatrics, Division of Pediatric Cardiology, Mayo Clinic, Rochester, Minnesota, USA.

Background: Variable expressivity and incomplete penetrance between individuals with identical long QT syndrome (LQTS)-causative mutations largely remains unexplained. Founder populations provide a unique opportunity to explore the effects of multiple genetic factors. **Objective:** We examined the role of a novel (ExAC MAF 0.000) synonymous *KCNQ1* p.L353L variant on the splicing of exon 8 and on heart rate corrected QT interval (QTc) in a population known to have a pathogenic, LQT1-causative mutation, p.V205M, in *KCNQ1* encoded Kv7.1. **Methods:** Four hundred and twenty adults were genotyped for p.V205M, L353L, and a previously described QTc modifier (*KCNH2*-K897T). Adjusted linear regression determined effect of each variant on the QTc, alone and in combination. In addition, peripheral blood RNA was extracted from 3 controls and 3 L353L-positive individuals. The mutant transcript levels were assessed via quantitative PCR and normalized to overall *KCNQ1* transcript levels to assess the effect on splicing. **Results:** The L353L alone conferred a 9.7 ms increase in QTc for females (p=0.04) and 13.2 ms for males (p=0.06) above baseline. The mechanism of L353L's effect was attributed to a ~3-fold increase in exon 8 exclusion resulting in ~25% mutant transcripts of the total *KCNQ1* transcript levels. Of interest, a significant interaction effect on the QTc was observed when male participants were compound heterozygotes for the L353L and theV205M. In men, the V205M mutation alone increased the QTc by 28.2 ms (p<0.001) above baseline, but the L353L*V205M variant combination increased the QTc by another 32.3 ms (p=.03) above that dictated by each variant alone, with a predicted over-all QTc of about 500ms. This combination effect was not seen in women who were compound heterozygotes for the same variants. **Conclusion:** Our results provide first evidence that synonymous variants outside the canonical splice sites in *KCNQ1* can alter splicing and clinically impact repolarization reserve. Through this mechanism, we identified that L353L can precipitate QT prolongation by itself, and produce a clinically relevant interactive effect in conjunction with other LQTS-causative and modifier variants. Our study suggests the modifying effect may be sex specific, a phenomenon observed in other LQTS founder populations.

593T

Molecular genetic alterations contributing to CCM pathogenesis: Do multiple lesions arise from a shared second-hit somatic mutation? *C. Gallione[1], A. Peiper[1], H. Zeineddine[2], O. Fedrigo[3], I. Awad[2], D. Marchuk[1].* 1) Department of Molecular Genetics & Microbiology, Duke University Medical Center, Durham, NC; 2) Department of Neurosurgery, University of Chicago Medical Center, Chicago, IL; 3) Duke Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC.

Cerebral cavernous malformations (CCMs, [MIM 116860]) are vascular lesions of the central nervous system consisting of clusters of dilated blood vessels which are associated with vascular leak and hemorrhage. These lesions frequently occur sporadically but are also seen as an autosomal dominant trait, the latter involving multiple, distinct lesions in the brain. Mutations in any one of three identified genes, *CCM1/KRIT1, CCM2,* and *CCM3/PDCD10,* cause the inherited form. Our lab has previously found bi-allelic germline/somatic mutations in the endothelial cells lining the vascular caverns of the mature CCM lesion, suggesting a Knudsonian two-hit mutation model for CCM pathogenesis. The question remains as to whether the multiple CCM lesions seen in inherited cases, which can range in number from a few to over one hundred, are seeded by independent somatic mutational events or a single somatic mutation in a precursor cell during development. Due to the paucity of available surgical material and the nearly complete lack of cases where multiple lesions are resected from an individual CCM patient, we are using our mouse models of CCM to investigate this question. Our mouse models recapitulate the human disease in that the mice, heterozygous for either *Ccm1, -2,* or *-3,* exhibit multiple brain lesions with histology, molecular signatures and ultrastructure identical to human lesions. Using DNA extracted from the mouse CCM lesions, we employ targeted deep resequencing of the three CCM genes to identify second-hit somatic mutations. We report here preliminary evidence from our pilot study of a common somatic mutation seeding multiple lesions, specifically, in our *Ccm3* model, a c.316A>T, p.106Lys>Ter variant in two lesions from the same animal. We are currently examining additional lesions using this same approach.

594F

Changes in heterochromatin structure are associated with a new form of cardiomyopathy with extensive fibrosis and juvenile cataract in the Hutterite population. *N. Abdelfatah[1], C. Seifer[5], C. Huculak[2], I. Buffo[3], H. Duff[1], P. Frosk[3,4], B. Gerull[1,2].* 1) Department of Cardiac Sciences and Libin Cardiovascular Institute, University of Calgary; 2) Department of Medical Genetics, Alberta Health Services, University of Calgary; 3) Department of Pediatrics and Child Health, Rady Faculty of Heath Sciences, University of Manitoba; 4) Department of Biochemistry and Medical Genetics, Rady Faculty of Heath Sciences, University of Manitoba; 5) Section of Cardiology, Rady Faculty of Heath Sciences, University of Manitoba.

   Introduction: Dilated cardiomyopathy (DCM) is an inherited cardiac condition characterized by ventricular dilatation and reduced systolic function. Patients including young children suffer from heart failure or sudden death. Nuclear envelope proteins have been shown to play an important role in the pathogenesis of DCM. Here, we present a remarkable cardiac phenotype associated with a novel homozygous *LEMD2* mutation c.38C>T in patients of the Hutterite population with juvenile cataract. *LEMD2* encodes an inner nuclear protein which has an important role in stabilization of the nuclear membrane. **Purpose:** To characterize the cardiac phenotype in homozygous mutation carriers and evaluate structural and functional effects of the mutation in heart tissue and *in vitro* cell culture. **Methods and Results:** Homozygous mutation carriers with a history of bilateral juvenile cataract at different age groups underwent detailed cardiac investigations. 8 out of 15 carriers between age 12 and 43 years showed signs of a cardiomyopathy phenotype with normal or mildly reduced cardiac function, segmental wall motion abnormalities and delayed enhancement of the left ventricle. One subject died suddenly at the age of 43, another one survived a sudden cardiac arrest at the age of 33. Heart tissue of the deceased individual has been examined by electron microscopy and revealed extensive changes of myocyte nuclei which demonstrate severe morphologic changes, including extensive elongation and bizarre shapes with clumping of peripheral heterochromatin. Histology showed myocyte disarray and extensive interstitial fibrosis. However, immunohistochemistry demonstrated regular nuclear membrane localization of LEMD2 in affected cardiac tissue. In addition, recombinant expression of LEMD2 protein in C2C12 cells showed no difference between the mutant and wild-type form as well as proper co-localization with lamin A/C indicating that localization and expression level of LEMD2 are not impaired by the mutation. **Conclusion:** Mutations in *LEMD2* lead to adult onset cardiomyopathy with risk for sudden death. Affected cardiac tissue shows remarkable changes in nuclei with abnormal heterochromatin, whereas the expression and localization of the mutant protein did not differ. Further investigations will focus on the interaction of LEMD2 with BAF, an important factor which is involved in heterochromatin formation. .

595W

Identification of *FLNC* variants in arrhythmogenic right ventricular cardiomyopathy. *R.L. Begay[1], T.J. Rowland[1], D.B. Slavov[1], S.L. Graw[1], G. Sinagra[2], K.L. Jones[3], K. Gowan[3], L. Mestroni[1], M.R.G. Taylor[1].* 1) Department of Medicine, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) Cardiovascular Department, Ospedali Riuniti and University of Trieste, Trieste, Italy; 3) Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Aurora, CO.

   Background – Arrhythmogenic right ventricular cardiomyopathy (ARVC) presents as a fibro-fatty replacement of the right ventricular myocardium leading to arrhythmias and heart failure. ARVC is familial in 30-50% of cases and is transmitted by autosomal dominance with incomplete and age-related penetrance. ARVC is mainly associated with variants in desmosomal genes, however, efforts continue to identify additional ARVC genes. *FLNC* encodes for filamin-C, an actin-binding protein localized at the Z-disc of striated muscle. Variants in *FLNC* cause skeletal myopathies and hypertrophic cardiomyopathy, restrictive cardiomyopathy, and dilated cardiomyopathy in the absence of skeletal muscle pathology. Here we present data that *FLNC* may be a novel disease-causing gene producing an ARVC phenotype in two families. *Methods and Results* – Our group sequenced families with ARVC utilizing Illumina TruSight One Sequencing Panel, which queries 4,813 genes associated with clinical phenotypes. Two FLNC variants (one truncation and one missense) were identified in ARVC probands (probands were screened for variants in 54 other cardiomyopathy-related genes and no pathogenic variants were detected). Both variants were confirmed by Sanger sequencing with a c.G6565T truncation variant present in family TSRVD029 and a c.G6988A missense variant in family TSRVD021. In addition, both variants are absent from the 1000 Genomes Project and ClinVar, and the truncation variant is additionally absent from the Exome Aggregation Consortium (ExAC) Browser database, while the missense variant is present at a frequency of 0.0007887 in the latter database. Both variants had a GERP score >4.2. Our next step is to expand the family pedigrees to enroll other affected family members for segregation analysis. *Conclusion* – *FLNC* variants have been inadequately understood with respect to cardiac muscle diseases. This study provides additional findings that *FLNC* truncation and even missense variants may lead to ARVC.

596T

A novel variant in the *ANK2* membrane-binding domain is associated with ankyrin-B syndrome and structural heart disease in a First Nations community of Northern British Columbia. *L. Chen[1,2], N. Murphy[3], S. McIntosh[4], S. Asuri[4], C. Kerr[5], S. Sanatani[6], E. Sherwin[6], P. Mohler[3], L. Swayne[1,2], L. Arbour[2,4].* 1) Division of Medical Sciences, University of Victoria, Victoria, BC, Canada; 2) Island Medical Program, University of British Columbia, Victoria, BC, Canada; 3) Dorothy M. Davis Heart and Lung Research Institute, Division of Cardiovascular Medicine, Department of Internal Medicine, and Department of Physiology and Cell Biology, The Ohio State University Wexner Medical Center, Columbus, OH, USA; 4) Department of Medical Genetics, University of British Columbia, Victoria, BC, Canada; 5) Division of Cardiology, University of British Columbia, Vancouver, BC, Canada; 6) Department of Pediatrics, Division of Cardiology, University of British Columbia, BC Children's Hospital, Vancouver, BC, Canada.

   Background: Long QT syndrome (LQTS) confers susceptibility to ventricular tachyarrhythmia, predisposing to syncope, seizures, and sudden death. LQTS type 4, also called 'ankyrin-B syndrome' due to the prevalence of other cardiac phenotypes, is caused by loss-of-function variants in *ANK2* (encodes ankyrin-B). Ankyrin-B syndrome-causing variants have been identified in the spectrin-binding and C-terminal regulatory domains of ankyrin-B. Here, we report the clinical, molecular, and cellular phenotypes associated with the first disease-causing variant identified in the membrane-binding domain (*ANK2* c.1937C>T p.Ser646Phe).   **Methods:** Two families, not known to be related, but from the same community were identified with LQTS. Sequencing of the 12 most common genes causing LQTS was carried out in the index case of each family. Genotyping and clinical information was collected on family members. A combination of in silico, biochemical, and primary cardiomyocyte cellular techniques were used to assess the functional significance.   **Results**: The *ANK2* p.S646F variant was detected in each index case. We identified 16 additional p.S646F carriers. Average QTc was 476.2 ms (± 33 ms) with a range of 430-560 ms in carriers. Two carriers presented with structural heart disease, one (age 36) with cardiomyopathy resulting in sudden death, another with congenital heart disease (TAPVR). This variant allele was not reported in the ExAC, NHLBI and 1000 Genomes (global MAF of 0.000). Ankyrin-B p.S646F resides on the αB of the 19th *ANK* repeat on the outer surface of the *ANK* repeat solenoid. Recombinant purified membrane-binding domain containing the p.S646F mutation displayed normal folding properties. Ankyrin-B p.S646F displayed reduced stability in myoblasts and expression in primary ankyrin-B+/- cardiomyocytes revealed aberrant targeting of the ankyrin-B p.Ser646Phe polypeptide and inability to rescue abnormal targeting of Na/Ca exchanger.   **Conclusion:** We identify the first disease-causing *ANK2* variant localized to the membrane-binding domain resulting in reduced expression and abnormal localization. Further study is warranted on the potential association of this variant with structural heart disease given the role of *ANK2* in targeting and stabilization of key structural and signaling molecules in cardiac cells.

597F

Differences in vascular disease presentation of patients with null and missense *SMAD3* mutations. *E.M. Hostetler[1], E.S. Regalado[1], P. Arnaud[2], N. Hanna[2], M. Aubart[2], M. Langeois[2], S.E. Wallace[1], D. Guo[1], G. Jondeau[2], C. Boileau[2], D.M. Milewicz[1].* 1) Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, TX; 2) Département de Génétique, Hôpital Bichat, Paris, France.

   SMAD3 mutations cause heritable thoracic aortic aneurysms and dissections, along with aneurysms of other arterial beds. Both null (nonsense and frameshift mutations) and missense mutations that disrupt the MH2 domain of the SMAD3 protein have been identified to cause disease. We sought to determine whether there are differences in clinical presentations and underlying disease mechanisms between null and missense *SMAD3* mutations. We analyzed a cohort of 156 individuals from 42 families with 35 different mutations, categorized as null (n=85) and missense (n=71). This cohort had a mean age of 39 years (SD 9.0) at enrollment and was 88% Caucasians and 56% males. Thirty-five percent (n=54) had an aortic event, defined as aortic dissection (n=38) or aneurysm requiring repair (n=16). There was no significant difference in frequency of aortic events in the missense and null mutation groups; however, the mean age of aortic event was significantly younger in individuals with missense mutations (mean 42 years, 95% CI 37.7, 45.5) than those with null mutations (mean 48 years, 95% CI 44.8, 51.7). Overall cumulative freedom from aortic event for *SMAD3* mutations was 21% (95% CI 6.2, 41.8) at the age of 76 years, but the survivor functions differed significantly between missense and null mutations (log rank test p value =0.03). Hazard rate of aortic event was 1.8 (95% CI 1.2, 2.7) times higher in patients with missense mutations as compared to null mutations after adjusting for intra-familial correlation, sex and site of recruitment. Interestingly, more individuals with null mutations presented with distal aortic disease compared to those with missense mutations, i.e. descending thoracic dissections (15% vs 5%) and abdominal aortic aneurysms (10% vs 0%). Also, more individuals with missense mutations than null presented with intracranial aneurysms (12% vs 3%) and other arterial aneurysms (9% vs 6%). These findings show that individuals with missense *SMAD3* mutations have earlier onset and higher risk of vascular events compared to null mutations. Assessment of additional individuals through the multi-site Montalcino Aortic Consortium will help elucidate these differences.

598W

Multiple rare variants in dilated cardiomyopathy are more frequent than expected. *A. Morales[1], D.D. Kinnamon[1], A. Whiting[1], M. Bamshad[2], D. Nickerson[2], R.E. Hershberger[1].* 1) Ohio State Univ. Human Genetics, Columbus, OH; 2) University of Washington Center for Mendelian Genomics.

   INTRODUCTION: Dilated cardiomyopathy (DCM) is primarily an autosomal dominant condition in which plausible genetic cause can be found in approximately 40% of cases. Previous reports in hypertrophic cardiomyopathy (HCM) have shown that multiple rare variants are uncommon, occurring in 3-5% of individuals. METHODS: To examine the frequency of multiple rare variants in DCM, we analyzed exome sequence, conducted at the University of Washington Center for Mendelian Genomics, of 405 individuals affected with DCM from 279 pedigrees, 103 who reported a negative family history of DCM and 176 with familial DCM (FDC). Using DCM variant adjudication rules developed by our group, variants from 40 previously reported DCM genes were analyzed and adjudicated as pathogenic (P), likely pathogenic (LP), or variant of uncertain significance (VUS). Multiple variants were analyzed if segregation occurred among all available affected DNAs in a family. RESULTS: We identified 62 (22%) pedigrees with more than one rare variant, for a total of 147 unique variants in 33 genes (9 P, 34 LP, 104 VUS). Among these probands, 34 were male; 28 were female. Fifty two were White (3 Hispanic), 9 were Black, 1 was Asian, and 1 reported more than one race. The median age of diagnosis was 42 years. Of these, 25 (41%) were diagnosed at <40 years and 4 (6%) were diagnosed before age 18. The median left ventricular (LV) size was 64 mm and the median ejection fraction was 27%. The multiple variant pedigrees involved 4 (n=2), 3 (n=15), 2 (n=42), or 1 (n=3) genes and had 4 (n=4), 3 (n=15), or 2 (n=43) variants. Variant combinations involving confirmed relevant DCM genes (*LMNA*, n=5; *MYH7*, n=5; *TNNT2*, n=3) were identified. Truncating variants (tv) in *TTN* were most frequently found (21/147=14%), mainly along with variants in *MYBPC3* (n=4), *BAG3* (n=3), *ACTN2* (n=3), *NEBL* (n=3), *MURC* (n=3), and *MYPN* (n=2). Thirty four multiple variant pedigrees had FDC confirmed by medical records. Of those, 13 had relatives available for segregation analysis. Of the 13 pedigrees with multiple variants and available family members' DNA, 5 (38%) had more than one multiple variant classified as P or LP. In these families, P and LP variants in *BAG3* and *TTN*tv were most frequent (4 times each), followed by *RBM20, ANKRD1*, and *MYH6*. CONCLUSION: Multiple variants classified as LP or P are present in 38% of evaluable pedigrees. Our findings suggest that a significant proportion of DCM cases carry relevant variants in multiple genes.

599T

Genetic aortopathy due to gene mutations related and not-related to TGF-β signal pathway in Japan. *H. Morisaki[1,2,3], T. Watanabe[3], A. Yoshida[3], H. Sasaki[3], K. Minatoya[3], T. Morisaki[2,3].* 1) Dept Medical Genetics, Sakakibara Heart Institiute, Fuchu, Tokyo, Japan; 2) Tokyo University of Technology School of Health Sciences, Ota, Tokyo, Japan; 3) National Cerebral and Cardiovascular Center, Suita, Osaka, Japan.

   Recent progress of molecular study revealed similar to Marfan syndrome (MFS) but distinct disorders caused by gene mutations in TGF-β signal pathway and contractile molecule. During more than 10 years, we performed mutation analysis in more than 1000 probands with young-onset or hereditary aortopathy in National Cerebral and Cardiovascular Center. Among these patients, we found pathogenic *FBN1* mutations in 496 patients. In addition, we found 119 patients in 91 families to have pathogenic gene mutations in *TGFBR1* (45 patients), *TGFBR2* (45 patients), *SMAD3* (18 patients), *TGFB2* (3 patients) or *TGFB3* (8 patients). Also, we found 91 patients in 61 families to have pathogenic gene mutations in *ACTA2* (38 patients), *COL3A1* (52 patients) and *MYH11* (1 patient). Ten probands with pathogenic *COL3A1* mutations were surgically treated with little complication even with noticeable tissue fragility and bruising tendency. Patients with *TGFBR1* or *TGFBR2* mutation frequently showed enlargement of Valsalva sinus, sometimes even in young childhood. Skeletal features including scoliosis and/or funnel chest and the history of pneumothorax were also frequently seen in these patients. Bifid uvula, hypertelorism, or arterial tortuosity, characteristic for Loeys-Dietz syndrome, was also observed in some MFS patients but less common. Some patients with young-onset or hereditary aortopathy showed cerebral arterial aneurysm and/or abnormality as reported before. These results indicate that patients with young or familial aortopathy patients need to be diagnosed properly not only by the clinical features but by the genetic study since it may not be easy to distinguish those patients from those with MFS.

600F

A platform for large-scale functional annotation of cardiomyopathy gene variants. *L. Qiao[1], W. Lv[2], K. Musunuru[2].* 1) Department of Stem Cell & Regenerative Biology, Harvard University, Cambridge, MA; 2) Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

   TNNT2 [MIM 191045], which encodes the sarcomere protein troponin T, has been linked to familial hypertrophic cardiomyopathy (HCM [MIM 115195]) and dilated cardiomyopathy (DCM [MIM 601494]), as well as other myocardial disorders such as restrictive cardiomyopathy [MIM 612422] and left ventricular noncompaction [MIM 601494]. The introduction of next-generation DNA sequencing into the clinic is dramatically increasing the discovery of variants of unknown significance in *TNNT2* and other genes linked to cardiomyopathies. Roughly 120 *TNNT2* variants have been catalogued in ClinVar; while some are reported as pathogenic for cardiomyopathies, many have conflicting evidence or are of unknown significance. We are developing a platform for large-scale functional annotation of these 120 variants, using a novel genome-editing approach that combines clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) and dual integrase cassette exchange (DICE) to rapidly introduce single copies of the *TNNT2* variants into a wild-type induced pluripotent stem cell (iPSC) line with a pooled approach requiring only a single targeting procedure. The allelic series of isogenic cell lines can then be differentiated into cardiomyocytes, followed by gene expression profiling and phenotyping to ascertain whether each mutation is pathogenic and, if so, to which type of cardiomyopathy it predisposes. In a proof-of-principle application of the platform, we generated 14 unique mutant cell lines and found that differentiated cardiomyocytes with a variant previously linked to DCM (p.R173W) displayed a substantial increase in *PDE2A* gene expression compared to control cardiomyocytes, consistent with previously published observations. We envision that the platform could ultimately be used in a clinical context to rapidly annotate novel patient-specific variants in cardiomyopathy genes.

601W

Phenotype variability in twins and sibling pairs with genetic predisposition to thoracic aortic aneurysms and dissections. *E. Regalado, E. Hostetler, S. Wallace, L. Mellor-Crummey, D. Guo, D. Milewicz.* Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX.

   Heritable thoracic aortic aneurysms and dissections are commonly passed down in families as a dominant disorder with reduced penetrance and widely variable age of onset, even within the same family. An unanswered question is whether this variability is due to genetic or modifiable environmental factors. We aimed to quantify the variability of age of onset (AO) of aortic event (aortic dissection or aneurysm repair) in monozygotic (MZ), dizygotic (DZ) twins and sibling pairs, and examine the concordance of aortic event and AO. We identified twins and sibling pairs from a large cohort of patients with familial thoracic aortic disease. Zygosity was confirmed by analysis of 15 microsatellite markers, when DNA was available, or by self-report. All possible sibling pair combinations whose mutation status or aortic disease could be ascertained were included. There were 15 pairs of MZ twins, including twins with known mutations of *ACTA2*, *FBN1*, *LOX*, *SMAD3* and *TGFBR2*, only two pairs of DZ twins, and 276 full sibling pairs from 60 families with mutations of *ACTA2* (135), *TGFBR1* (32), *TGFBR2* (71), and *SMAD3* (38). Concordance for aortic event was 80% for MZ twins and 56% for *ACTA2*, 47% for *TGFBR1*, 55% for *TGFBR2* and 58% for *SMAD3* sibling pairs. AO was not statistically different between MZ twins (paired t test p=0.08; 6.6 years, 95% CI -3.5, 16.7), but correlation of AO was low (0.4983) and was influenced by a twin pair in which one twin dissected and the second twin was medically managed for 20 years without an event. In comparison, mean difference in AO between sibling pairs was statistically significant (p <0.01): 12 years (95% CI 9.4, 14.2) for *ACTA2*, 7.2 (95% CI 3.3, 11) for *TGFBR1*, 10.9 (95% CI 6.2, 15.5) for *TGFBR2*, and 5.1 (95% CI 2.6, 7.5) for *SMAD3*, and intra-pair correlations were 0.72, 0.79, 0.88, and 0.42, respectively. These findings suggest that genetic factors play a role in modifying the effect of genes for aortic disease as seen in the higher concordance for aortic event among MZ twins compared to sibling pairs. Incomplete concordance among MZ twins also indicates a role for the unique environment in altering age of onset. These findings have clinical significance in predicting disease risk in twins and sibling pairs and understanding the sources of variability in age of onset of aortic disease.

602T

Obscurin variants in patients with left ventricular noncompaction. *T.J. Rowland[1], S.L. Graw[1], M.E. Sweet[1], M. Gigli[2], M.R.G. Taylor[1], L. Mestroni[1].* 1) Cardiovascular Institute and Adult Medical Genetics Program, University of Colorado Denver Anschutz Medical Campus, Aurora, CO; 2) Cardiovascular Department, "Ospedali Riuniti," University of Trieste, Trieste, Italy.

Left ventricular noncompaction (LVNC) is a rare type of cardiomyopathy. In our patient population, we identified a possible association between LVNC and variants in the obscurin gene (*OBSCN*). Obscurins are giant proteins (≈700 to 900 kDa) involved in sarcomere structure and play key roles in myofibrillo-genesis and cytoskeletal arrangement through interaction with several other binding partners. While disrupting these interactions can have severe consequences for normal striated muscle function, obscurin's pathogenic involvement in cardiomyopathies is unclear. The association of *OBSCN* missense variants with more common cardiomyopathies in two previous clinical studies, combined with suggestions of a functional cardiac role in limited animal studies, made obscurin an intriguing target for exploration in our cardiomyopathy cohort. Using the Illumina TruSight One Sequence Panel, which queries 4,813 genes associated with clinical phenotypes, we identified and confirmed four probands with *OBSCN* frameshift or splicing variants (four variants total) in our population of 335 cardiomyopathy patients. Three of these probands have *OBSCN* frameshift variants, and the remaining one has a splicing variant. Interestingly, while the majority of patients analyzed had dilated cardiomyopathy (DCM) (325 out of 335 [97.0%]), only one of the four *OBSCN* probands had DCM. The other three probands had LVNC, which was rare in our population (10 out of 335 [3.0%]), giving an odds ratio of 139 [CI 13-1506] for the association between an *OBSCN* variant and LVNC. All four *OBSCN* variants identified localize to the C-terminal end of obscurin-B-like isoform and occur upstream of the fibronectin type-III 4 and protein kinase 2 domains, which the frameshift variants are predicted to eliminate. The earliest variant occurs upstream of the protein kinase 1 domain. We screened all four probands for variants in 54 other known cardiomyopathy-related genes and no pathogenic variants were detected. Affected family members were unavailable for segregation analysis of the *OBSCN* probands. Our findings suggest a strong correlation between *OBSCN* frameshift and splicing variants, all clustering to the C-terminal end of the same isoform group, with the occurrence of the rare LVNC phenotype.

603F

Reassessment of cardiomyopathy gene and variant pathogenicity using 7,855 clinical cases and 60,706 reference samples. *K.L. Thomson[1,2], R. Walsh[3,4], J.S. Ware[3,4], B.H. Funke[5,6], J. Woodley[2], K.J. McGuire[2], F. Mazzarotto[3,4], E. Blair[7], A. Seller[2], J.C. Taylor[8,9], E.V. Minikel[10,11], D.G. MacArthur[10,11,12], M. Farrall[1,8], S.A. Cook[3,13,14], H. Watkins[1,8,9], Exome Aggregation Consortium.* 1) Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 2) Oxford Medical Genetics Laboratory, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; 3) National Heart and Lung Institute, Imperial College, London, United Kingdom; 4) NIHR Royal Brompton Cardiovascular Biomedical Research Unit, Royal Brompton Hospital, London, United Kingdom; 5) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Massachusetts, 02139, USA; 6) Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 7) Department of Clinical Genetics, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; 8) The Wellcome Trust Centre for Human Genetics,Oxford,United Kingdom; 9) NIHR Biomedical Research Centre, Oxford, United Kingdom; 10) Analytical and Translational Genetics Unit, Massachusetts General Hosptial, Boston, MA 02114, USA; 11) Broad Institute of MIT and Harvard, MA 02142, USA; 12) Department of Medicine, Harvard Medical School, Boston,USA; 13) National Heart Centre Singapore, Singapore 169609, Singapore; 14) Duke-National University of Singapore, Singapore 169857.

Introduction: Recent population studies have cast doubt on the pathogenicity of variants previously reported in Mendelian disease genes and the validity of some gene:disease relationships. In the clinical laboratory, this has led to more variants being classified as variants of uncertain significance (VUS); these inconclusive results reduce the clinical utility of testing and increase the risk of false positive diagnosis. The challenges of variant interpretation are particularly well illustrated by inherited cardiomyopathies (CMs), which are among the most common serious Mendelian disorders. Using data from CM cohorts and the Exome Aggregation Consortium (ExAC), we show how new opportunities for large–scale comparison of rare variation can enable us to verify the genes and variant types that can be reliably interpreted in a clinical setting. **Methods:** Data from 7855 cases with Hypertrophic Cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM) or Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) were obtained from two clinical laboratories. 20 HCM, 46 DCM and 8 ARVC genes were evaluated. Rare variant frequency (minor allele frequency [MAF] $<1\times10^{-4}$ in ExAC) was assessed in clinical cases and 60,706 ExAC samples. Statistical analyses were performed to compare rare variant frequency between cohorts. **Results:** In most well-established CM genes a significant excess of rare variation was found in cases. In addition, our results suggest that in these key genes, a significant proportion of variants currently classified by the clinical laboratories as VUS, are likely to be disease causing. However, in many other genes, including some reported as important causes of a particular CM, no significant difference in rare variant frequency was found. This indicates that novel rare variants in these genes cannot be reliably interpreted. Also, our findings suggest that many previously reported variants in these genes are false positives and that some published gene:disease pairs may not be valid. **Conclusions:** Robust assessment of rare variant frequency is essential to evaluate the genes and variants implicated in rare diseases. Large collaborative datasets of cases, and importantly reference samples, provide unprecedented power for such assessment. These approaches, applicable across a range of Mendelian diseases, will improve selection of genes for clinical test panels, reduce uncertainty in interpretation and increase the clinical utility of testing.

604W

Titin-truncating variants, a common incidental finding, are penetrant in the general population and may predispose the heart to fail. *J.S. Ware[1,2,3], A. de Marvao[2], S. Schafer[4], R. Walsh[1,3], T. Dawes[2], P.J.R. Barton[1,3], N. Hubner[5], D. O'Regan[2], S.A. Cook[1,2,3,4,6].* 1) National Heart & Lung Institute, Imperial College London, UK; 2) MRC Clinical Sciences Centre, Imperial College London, UK; 3) NIHR Royal Brompton Cardiovascular Biomedical Research Unit, Royal Brompton Hospital & Imperial College London, UK; 4) National Heart Centre Singapore, Singapore; 5) Cardiovascular and Metabolic Sciences, Max-Delbrück-Center for Molecular Medicine, Berlin, Germany; 6) Duke-National University of Singapore, Singapore.

Dilated cardiomyopathy (DCM) is a frequent cause of heart failure and a common indication for heart transplantation. DCM has a strong genetic basis, with variants that truncate the sarcomeric giant protein titin the commonest genetic cause (TTN-truncating variants, TTNtv, in up to 1 in 4 cases). However, rare variants in titin are cumulatively common: rare missense variants have a prevalence >50% and truncating variants 1-2% in the general population, and it is not known whether these are have physiological or disease relevance. First, we re-examined *TTN* transcripts in the human heart, and identified a subset of variants that do not impact the predominant cardiac isoforms. These variants are not associated with DCM. Stringent assessment of allele frequencies further identified a tranche of variants that, though rare (<1:1000), remained implausibly common in reference samples to cause disease. These were typically splice-disrupting, and we speculate may be rescued by alternative splice sites or in-frame exon-skipping. After excluding such variants, the prevalence of rare TTNtv predicted to influence the heart is ~0.4% in reference samples.   We characterised the heart of healthy volunteers in great detail and determined TTNtv status. Volunteers with TTNtv had significantly increased cardiac chamber dimensions and reduced contractility on conventional cardiac MRI, while ultra-high resolution 3D MRI coupled with atlasing techniques had increased power to demonstrate regional differences in structure and function. Effect sizes of TTNtv on the hearts of healthy volunteers were large with regression Beta values ranging from 4-15% and of a magnitude that may be expected to affect clinical outcomes.   We generated rat models carrying TTNtv, to dissect disease mechanisms. Young heterozygous mutant animals had normal cardiac function, but a distinct metabolic phenotype (mTORC activation and altered fatty acid metabolism) suggestive of an adaptive compensated state at baseline. Examining cardiac function *ex vivo*, mutant hearts had reduced reserve and failed under mechanical stress. We have previously demonstrated that TTNtv in humans are associated with peri-partum cardiomyopathy, similarly suggesting that TTNtv predispose to failure under haemodynamic stress.   Approximately 30,000,000 people world-wide carry TTNtv that are associated with cardiac remodelling and DCM, and TTNtv in the general population may predispose otherwise healthy individuals to heart failure.

605T

Identification of a second gene mutated in capillary malformation–arteriovenous malformation (CM-AVM2). *M. Amyere[1], N. Revencu[2], E. Pairet[1,2], R. Helaers[1], M. Baselga[3], M. Cordisco[4], W. Chung[5], J. Dubois[6], J.P. Lacour[7], L. Martorell[8], J. Mazereeuw-Hautier[9], R. Pyeritz[10], J.B. Mulliken[11], L. Boon[1,12], M. Vikkula[1].* 1) Human Molecular Genetics, de Duve Institute, Brussels, Belgium; 2) Center for Human Genetics, Cliniques universitaires St Luc, Université catholique de Louvain, Brussels, Belgium; 3) Department of Dermatology, Hospital de la Santa Creu I Sant Pau, Barcelona, Spain; 4) Department of Dermatology, Hospital Garrahan, Buenos Aires, Argentina *Current affiliation: Strong Hospital, University of Rochester School of Medicine and Dentistry, Rochester, U.S.A; 5) Departments of Pediatrics and Medicine, Columbia University, New York, New York, USA; 6) Department of Medical Imaging, Sainte-Justine Mother-Child University Hospital, Montreal, Canada; 7) Service de Dermatologie, Centre Hospitalo-Universitaire de Nice, Nice, France; 8) Genética Molecular, Hospital Sant Joan de Déu, Barcelona, Spain; 9) Service de Dermatologie, Centre de Référence des Maladies rares de la peau, Hôpital Larrey, Toulouse, France; 10) Departments of Medicine and Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA; 11) Vascular Anomalies Center, Boston Children's Hospital and Harvard Medical School, Boston, U.S.A; 12) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques universitaires Saint-Luc and Université catholique de Louvain, Brussels, Belgium.

Capillary Malformation–Arteriovenous Malformation (CM-AVM) is an autosomal dominant disorder manifesting multifocal CMs together with high risk for fast-flow vascular malformations, especially in the head and neck. We have detected *RASA1* heterozygous loss-of-function mutations in about 50% of screened families. This suggests genetic locus heterogeneity. We performed a genome-wide linkage study in a large family with autosomal dominantly inherited CMs without a *RASA1* mutation, and identified a locus on Chr7. Whole exome sequencing was performed including an additional 9 blood samples of patients from 7 unrelated families. Subsequently, a candidate-gene was screened by targeted massively parallel sequencing for mutations in 365 patients with CMs associated or not with fast-flow vascular malformations. A damaging mutation was identified in 5 out 8 families. Moreover, targeted massive parallel sequencing unravelled mutations in an additional 49 families with CMs with/without fast-flow vascular malformations. In total, 54 mutations were identified: 50% were non-sense, frame-shift, or splice site mutations, and 50% were substitutions predicted to strongly impact protein function. Expression studies of selected variants demonstrated loss of protein function. Mutations were identified altogether in 103 individuals: > 95% had capillary malformations, usually multifocal, and 20% an associated fast-flow vascular malformation. The cutaneous capillary malformations in these patients are more telangiectatic than in CM-AVM, and thus a clinical distinction can be possible. We suggest to call this new entity CM-AVM2. As the fast-flow lesions can cause severe morbidity and even mortality, it is important to recognize for entities.

606F

Origin of the *KRIT1* common Hispanic founder mutation causing cerebral cavernous malformation type 1. *L. Pawlikowska[1,2], H. Choquet[1], C.E. McCulloch[3], A. Akers[4], L. Morrison[5], H. Kim[1,2,3].* 1) Center for Cerebrovascular Research, Dept of Anesthesia, Univ California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, Univ California, San Francisco, CA; 3) Dept of Epidemiology and Biostatistics, Univ California, San Francisco, CA; 4) Angioma Alliance, Durham, NC; 5) Depts of Neurology and Pediatrics, Univ New Mexico, Albuquerque, NM.

   Objective: Familial cerebral cavernous malformation type 1 (CCM1, MIM116860) is an autosomal dominant disease caused by mutations in *KRIT1* (krev-interaction trapped 1, *CCM1*), and characterized by multiple brain lesions that can lead to intracerebral hemorrhage, seizures, and neurological deficits. The high prevalence of the disease in families of Mexican-American heritage is due to a founder effect caused by the *KRIT1 c.1363C>T,* p.Q455X "common Hispanic mutation" (CHM) inherited from a shared ancestor. Genealogic studies in large CCM1-CHM families have traced this mutation to a likely shared ancestor from Spain. We investigated the geographic origin of the *KRIT1* CHM founder mutation using haplotype analysis. **Methods:** We focused on a 7q region of 0.66Mb flanking the *KRIT1* gene, including 31 single nucleotide polymorphisms (SNPs). Genotypes were extracted from the Affymetrix Axiom Genome-Wide LAT 1 Human Array for CCM1-CHM subjects and Illumina Human1M or Affymetrix SNP 6.0 data for HapMap controls. Haplotype estimation and association testing were performed in Haploview v4.2 in 189 Hispanic CCM1-CHM patients enrolled in the Brain Vascular Malformation Consortium CCM study, and 474 healthy controls (165 European, 86 Mexican, 137 Chinese and 87 African ancestry subjects) from the International HapMap Project. Linkage disequilibrium (LD) structure using pairwise correlation (r^2) was examined in populations and used to determine haplotype boundaries. **Results:** The LD structure of the region surrounding *KRIT1* differed among different control groups. As expected, the CCM1-CHM cohort shared much longer haplotypes around the CHM locus than did controls, consistent with their shared ancestry and familial relationships. We identified a shared haplotype spanning 525kb around the *KRIT1* gene consisting of 16 genetic markers (CHM underlined, TGACGCATCATTAATA). The corresponding ancestral haplotype, TGACGCATCATCAATA was found in 22% of Mexican, 26% of European, 19% of Chinese and 7% of African control chromosomes. **Conclusions**: The CCM1 common Hispanic founder mutation (*KRIT1 c.1363C>T,* p.Q455X) arose on a haplotype that is common among both European and Mexican controls. We are performing further analysis utilizing longer haplotypes, Native American controls and local genetic ancestry prediction algorithms to identify the ancestry of the CHM haplotype.

607W

Genetic testing and diagnoses within a cardiovascular connective tissue clinic. *G. MacCarrick[1], J.P. Habashi[1], H.C. Dietz III[1,2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 2) Howard Hughes Institute, Chevy Chase, Maryland, USA.

   Many genetics clinics find themselves evolving to specialty clinics, utilizing the expertise of their attending physicians; this is especially true in cardiac genetics. We sought to review the demographic, diagnosis and genetic testing data from a busy cardiovascular connective tissue clinic to examine the utilization of genetic testing in this setting. Data from 530 unique patients seen in one year by two cardiologists with genetics specialty and one genetic counselor was reviewed. The average patient age was 25.1 years, with 295/235 M:F ratio. 61% of patients were from out of state, with 88.5% Caucasian ethnicity. The most common diagnoses seen were Marfan syndrome (28%), aortic aneurysm/dissection (18%), Loeys-Dietz syndrome (13%) and ascending aortic aneurysm +/- bicuspid aortic valve (7%). Out of 55 patients referred for possible Marfan or Loeys-Dietz syndromes based on skeletal findings only, 1 had aneurysmal disease leading to a Loeys-Dietz diagnosis. Nine patients came with variants of unknown significance (VUS) previously detected, seeking greater expertise in interpretation of results. Of 42 patients with aortic aneurysm or dissection on which aortopathy panel testing was pursued, 9 (21%) had positive results (6 *FBN1*, 4 *TGFBR1/2*, 2 *ACTA2*, 1 *TGFB2*) and 12 had one or more VUS. There were no positive panel results in 6 patients with isolated arterial dissections including coronary artery dissections. A total of 62 aortopathy panels were sent out through clinic, with additional panel testing including: 3 Ehlers-Danlos, 1 Noonan, 1 Stickler. Single gene testing included: 8 *FBN1*, 2 *COL3a1*, and 6 familial mutations. SNP arrays were pursued in only 3 patients presenting with intellectual disability, and 6 clinical whole exomes were pursued with no positive results. 6 patients with ectopia lentis (ECL) without aneurysm were evaluated; consideration of exome slice for ECL genes was considered, but found to be cost-prohibitive. In patients with Marfan syndrome, 60% had molecular diagnoses; 40/95 of these positive tests were sent out through their care at our institution. In summary, in the cardiovascular connective tissue clinic, aortopathy panels continue to offer great appeal for a streamlined diagnostic approach with seemingly low burden of VUS contrasted with significant positive impact in individualized medical care based on diagnosis.

608T

Is the common *Cerebral Cavernous Malformation 2* **exon 2-10 deletion due to a founder in the US population?** *M. Detter[1,2], A. Akers[3], D. Marchuk[1].* 1) Molecular Genetics and Microbiology, Duke University, Durham, NC; 2) Medical Scientist Training Program, Duke University, Durham, NC; 3) Angioma Alliance, Norfolk, VA.

Cerebral cavernous malformations (CCMs) are clusters of dilated blood vessels that occur predominately within the central nervous system. These ectatic vessels are prone to hemorrhage and patients present clinically with headache, focal neurologic deficits, seizure, stroke, and death. The prevalence of CCMs is estimated to be 0.5% of the general population. Mutations within three genes, *KRIT1* (*CCM1*), *MGC4607* (*CCM2*), and *PDCD10* (*CCM3*), have been identified as contributing to the pathogenesis of CCMs in both the sporadic and heritable forms of disease. Previous work in our lab identified a 77.6-kb deletion of *CCM2* exon 2-10 shared among 8 probands who were negative for mutations detected through sequence analysis. In order to distinguish between this shared deletion occurring multiple time within the population or as the result of a founder effect, we determined microsatellite and SNP haplotypes within 650-kbs proximal and 1410-kbs distal to the deletion. Haplotype analysis revealed a variation at the microsatellite 140kb distal to the deletion site. The results of this study suggest either multiple independent deletion events or a founder effect with subsequent recombination within the 140kb region between the deletion and varying microsatellite. To further investigate a potential founder effect, our current study has recruited a new cohort of 12 unrelated American families known to harbor the *CCM2* exons 2-10 deletion. By DNA sequence analysis of the 5-kb regions proximal and distal to the deletion point, we are performing a full SNP haplotype analysis of the region flanking the deletion. Complete sequencing of the 10-kb of DNA immediately adjacent to the deletion will provide more statistical power to determine the presence of a founder within the US population of CCM patients. While the identification of a founder effect within this population would be an interesting result, different SNP haplotypes would also be provocative. This deletion is known to occur through recombination of an *AluSx* sequence in intron 1 and an *AluSg* sequence distal to exon 10. Different SNP haplotypes would highlight the stochastic nature of this recombination event that results in an identical 77.6-kb deletion shared among unrelated families.

609F

Haplotype analyses of CYP2C19*2 and CYP2C19*17 genetic polymorphisms in clopidogrel non-responsiveness after percutaneous coronary intervention with stent implantation. *F. Saydam[1], I. Degirmenci[2], A. Birdane[3], M. Ozdemir[4], C. Ozbayer[5], N. Ata[3], H. Gunes[2].* 1) Department of Medical Biology and Genetics, Recep Tayyip Erdogan University Faculty of Medicine, Rize, Turkey; 2) Department of Medical Biology, Eskişehir Osmangazi University Faculty of Medicine, Eskişehir, Turkey; 3) Department of Cardiology, Eskişehir Osmangazi University Faculty of Medicine, Eskişehir, Turkey; 4) Department of Pharmacology, Eskişehir Osmangazi University Faculty of Medicine, Eskişehir, Turkey; 5) Department of Midwifery, Dumlupınar University Kütahya School of Health, Kütahya, Turkey.

Dual antiplatelet medication with aspirin and clopidogrel is currently recommended strategy for patients undergoing percutaneous coronary intervention (PCI) with stent implantation. However, emerging non-responsiveness to clopidogrel often results in fatal stent thrombosis. Previous findings suggest that the CYP2C19*2 polymorphism is associated with non-responsiveness to clopidogrel and the CYP2C19*17 polymorphism enhances antiplatelet activity of clopidogrel. We therefore aimed to analyze the haplotypes of these polymorphisms in clopidogrel non-responsiveness. Antiplatelet activity of clopidogrel activity was measured by the VerifyNow P2Y$_{12}$ assay in blood samples collected from patients that took a standard dose of clopidogrel (75 mg/day) for at least 7 days. 243 responder and 104 non-responder patients underwent PCI with stent implantation are included in our study. The CYP2C19*2 (c.681G>A) and CYP2C19*17 (c.-806C>T) polymorphisms are genotyped using the Sequenom MassARRAY system. The genotype frequencies for each polymorphism were in good agreement with the predicted Hardy–Weinberg equilibrium values. Haplotypes with a frequency of <0.03 were excluded. Only frequencies of the A-T haplotype were <0.03 in our study group. The comparison of the G-T haplotype frequencies between the responder and non-responder group did not reveal any significant difference. The A-C haplotype was significantly associated with non-responsiveness to clopidogrel (OR=2.855, 95%CI: 1.868–4.364, *P*<0.001). Conversely, the G-C haplotype was associated with responsiveness to clopidogrel (*P*=0.002). Our findings suggest that the A-C haplotype has a 2.8-fold increased risk and the G-C haplotype enhances antiplatelet activity of clopidogrel. Checking the A-C haplotype as a pharmacogenetics test for risk patients will make treatment effective.

610W

Mechanism of priapism in a mouse model of Fabry disease. *J. Shen[1], E. Arning[1], X. Meng[1], M. Wight-Carter[2], T. Day[1], T. Bottiglieri[1], R. Schiffmann[1].* 1) Institute of Metabolic Disease, Baylor Research Institute, Dallas, TX; 2) Animal Resources Center, UT Southwestern Medical Center, Dallas, TX.

Fabry disease is caused by deficient activity of lysosomal enzyme α-galactosidase A. Fabry disease exhibits broad range of clinical manifestations including cardiovascular diseases and renal insufficiency. Nitric oxide (NO) pathway dysregulation was thought to contribute to vasculopathy in this disease. Fabry disease is one of a few diseases that are associated with priapism, an abnormal prolonged painful erection of the penis. The pathophysiology of priapism in Fabry disease is not well understood. Our goal was to investigate underlying mechanism of priapism in Fabry disease using a mouse model. We found that Fabry mice develop priapism in an age-dependent manner with onset of symptoms at approximately 15 months of age. Wild type (WT) controls with the same genetic background had very low frequency of priapism (0.5% (1/207)). Compared to normal controls, expression of plasminogen activator inhibitor-1, which plays important roles in progression of fibrosis, was significantly upregulated in Fabry mouse penile tissues, suggesting that Fabry disease promotes penile fibrosis. Plasminogen abnormality has been found in children with Fabry disease (2007; 104: 2873). Expression of neuronal nitric oxide synthase (nNOS), but not endothelial and inducible NOS, was significantly increased in Fabry mouse penis. The nNOS upregulation was controlled at transcription level. It has been suggested that excess adenosine and reduced gene expression and activity of phosphodiesterase-5A (PDE5A), the cGMP-specific esterase, contribute to priapism in mouse models of sickle cell disease and NOS-deficient conditions. There was no significant change in tissue adenosine level and PDE5A expression level in Fabry mouse penis compared to WT controls. In conclusion, our findings suggest that increased nNOS expression and NO production may contribute to priapism in Fabry disease. Furthermore, nNOS was also upregulated in Fabry mouse heart, suggesting that nNOS-mediated NO dysregulation is a general pathogenic mechanism in Fabry disease.

611T

Loss-of-function mutations in *NONO* can cause an X-linked form of syndromic left ventricular non-compaction. *D.A. Scott[1,2], A. Hernandez-Garcia[1], M. Azamian[1], V.K. Jordan[2], B.J. Kim[1], M. Starkovich[1], J. Zhang[1], L.-J. Wong[1], S.A. Darilek[1], A.M. Breman[1], Y. Yang[1], J.R. Lupski[1,3,4,5], A.K. Jiwani[6], B. Das[7], S.R. Lalani[1], A.D. Iglesias[8], J.A. Rosenfeld[1], F. Xia[1].* 1) Department of Molecular & Human Genetics, Baylor College Medicine, Houston, TX 77030; 2) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 4) Department of Pediatrics, Baylor College Medicine, Houston, TX 77030; 5) Texas Children's Hospital, Houston, TX 77030; 6) Department of Pediatrics, University of Texas Medical Branch, Galveston, TX 77555; 7) Department of Pediatrics, Children's Medical Center, UT Southwestern Medical Center, Dallas, TX 75235; 8) Department of Pediatrics, Columbia University, New York, NY 10027.

Left ventricular non-compaction (LVNC) is a distinct form of cardiomyopathy characterized by deep intertrabecular recesses in hypertrophied segments of left ventricular myocardium. LVNC can present as an isolated finding, in association with congenital heart defects, or as part of a genetic syndrome. The non-POU domain containing, octamer-binding gene (*NONO*) is located on chromosome Xq13.1 and encodes a member of a small family of RNA and DNA binding proteins that includes SFPQ and PSPC1. Members of this family form homo- and hetero-dimeric complexes that perform a variety of tasks involved in RNA synthesis, transcriptional regulation and DNA repair. Loss-of-function mutations in *NONO* have been described as a cause of intellectual disability in males but not as a cause of LVNC or congenital heart defects. By searching a clinical database of over 6,000 individuals referred for exome sequencing and over 60,000 individuals referred for copy number variant (CNV) analysis, we identified two Hispanic males with LVNC, atrial and ventricular septal defects, developmental delay and intellectual disability, who harbored *de novo*, loss-of-function mutations in *NONO*. We also identified a Hispanic male infant with developmental delay, congenital brain anomalies and severe LVNC requiring cardiac transplantation, who inherited a single-gene deletion of *NONO* from his asymptomatic mother. Features commonly observed in males with *NONO* loss-of-function alleles included abnormalities of the corpus callosum and relative macrocephaly. We conclude that loss-of-function mutations in *NONO* can cause an X-linked form of syndromic LVNC associated with congenital heart defects, global developmental delay and intellectual disability.

612F

Pseudoxanthoma elasticum (PXE): Dysregulation of local ATP metabolism and treatment with a tissue non-specific alkaline phosphatase (TNAP) inhibitor. *S.G. Ziegler[1], C.R. Ferreira[2,3], E. Gallo MacFarlane[1], R.C. Riddle[4], R. Tomlinson[4], C. Ma[5], E. Sergienko[5], A.B. Pinkerton[5], J.L. Millan[5], W.A. Gahl[2], H.C. Dietz[1,6].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Division of Genetics and Metabolism, Children's National Health System, Washington, DC; 4) Department of Orthopaedic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA; 6) Howard Hughes Medical Institute, Chevy Chase, MD.

   Biallelic mutations in *ABCC6* cause PXE, characterized by calcification in the eyes, skin, and vessels. While the function of ABCC6 in PXE remains elusive, the mechanisms of related ectopic calcification disorders are better understood. Generalized arterial calcification of infancy is caused by biallelic mutations in *ENPP1*, which encodes an enzyme that converts ATP to AMP and pyrophosphate (PPi), a major inhibitor of tissue calcification. AMP is further degraded to adenosine and inorganic phosphate by CD73, encoded by *NT5E*; biallelic mutations in *NT5E* cause calcification of joints and arteries. We reasoned that a test for genetic interaction in mouse models of ectopic calcification might inform the pathogenesis of PXE. *Enpp1*[-/-] and *Abcc6*[-/-] mice showed fibrous capsule calcification of the vibrissae (an early marker of ectopic calcification) at 15 weeks of age, while *Nt5e*[-/-] mice calcified after one year. *Abcc6*[-/-] mice with one mutated *Enpp1* or two defective *Nt5e* alleles showed accentuated calcification with strong statistical evidence for synergy. Additionally, ABCC6, ENPP1, and CD73 exhibited regulatory interactions; human fibroblasts with biallelic *ABCC6* mutations had increased ENPP1 and decreased CD73 activity. Taken together, these data suggest that ABCC6 participates with ENPP1 and CD73 in ATP metabolism. *ABCC6* mutant cells exhibited a calcification phenotype in culture, suggesting a cell-autonomous defect and opposing the prevailing hypothesis that PXE results from failed liver secretion of an endocrine inhibitor of peripheral calcification. Moreover, a conditional *Abcc6*-targeted mouse revealed that liver-specific deletion failed to recapitulate the calcification phenotype produced by ubiquitous *Abcc6* deletion. Given their ability to recapitulate pathogenic events, PXE patient fibroblasts emerged as a viable model for investigating therapies. Under osteogenic conditions, *ABCC6* mutant cells had increased expression and activity of TNAP, an enzyme that degrades PPi. A novel, selective, and orally bioavailable TNAP inhibitor robustly attenuated calcification in PXE patient cells and in *Abcc6*[-/-] mice. A parsimonious model suggests that ENPP1 serves to generate PPi from ATP, while ABCC6 and CD73 cooperate in the further catabolism of AMP to adenosine, which normally represses TNAP expression. Perturbation of this pathway limits the bioavailability of PPi and hence results in calcification inferring broad therapeutic relevance for TNAP inhibitors.

613W

Case report: Two likely pathogenic mutations one in *MYH7* and one in *TTN* in a 12 year old girl with hypertrophic cardiomyopathy with restrictive phenotype. *L.M. Sigsworth[1], M.J. Bock[1], J.A. Gold[1,2].* 1) Loma Linda University , Loma Linda , CA; 2) University of California, Irvine, CA.

   Introduction: Hypertrophic cardiomyopathy (HCM) is a common form of cardiomyopathy, which may present at any age. Most cases of HCM are thought to occur due to single gene mutations. This case study explores the rare occurrence of a patient with early onset severe HCM who possesses mutations in two different genes, leading to HCM. **Case:** The patient is a 12-year-old Hispanic female who initially presented with dyspnea and changes in vision when running the mile during physical education class. Physical exam: grade 3/6 mid-systolic ejection murmur. Her cardiac silhouette was unremarkable. Echocardiogram showed severe circumferential left ventricular hypertrophy with a mid-cavitary gradient and an ejection fraction of 60-65%. She has a pulmonary capillary wedge pressure of 24 mmHg by catheterization. A family history was noncontributory. **Management and Outcome:** The patient was then transferred to Loma Linda Children's. She was clinically diagnosed with HCM with a restrictive phenotype and was seen by pediatric genetics. A SNP chromosomal microarray showed a 234 kB duplication at 10q24.32. The duplicated interval involves six known genes (*BTRC*, *DPCD*, *POLL*, *MIR3158*-2, *MIR3158*-1, and *FBXW4*). *BTRC* and *FBXW4* are partially duplicated and not associated with HCM. The cardiomyopathy panel revealed mutations in MYH7 on chromosome 14 (OMIM #192600) (cDNA change:c.5258A>G) (amino acid change :pGlul753Gly), which is is likely pathogenic. MYH7 encodes for the beta-myosin gene in the heavy chain of the sarcomere. The in-silico analysis suggests this may affect protein function. The TTN mutation on chromosome 2 (OMIM #613765) (cDNA change:c.46759G>A) (amino acid change: p.Ala15587Thr). This is a novel mutation, predicted to affect protein function.
Discussion: MYH7 is the most common mutation that causes hypertrophic cardiomyopathy (Kaski 2009). TTN is more commonly reported to cause dilated cardiomyopathy (Itoh-Satoh 2002, Neiva-Sousa 2015). TTN has been reported to cause HCM (Satoh 1999). A unique aspect of this patient's presentation is the early onset of late stage HCM. Researchers have reported a "dose effect" for double mutation carriers of cardiomyopathy mutation variants (Bales 2016). This is another case that shows a possible "double-dose effect" leading to a more severe phenotype. Whether or not her duplication is significant will depend on the results of her parents testing. Other family members could develop late onset HCM with one gene mutation.

614T

A genomic exploration to explain Glatiramer Acetate's adverse cardiovascular effects. *I. Braenne[1,2,3], C. Willenborg[1], L. Zeng[4,5], J. Erdmann[1,3], H. Schunkert[4,5], CARDIoGRAM Consortium, CARDIoGRAMplusC4D Consortium.* 1) IIEG, University of Luebeck, Luebeck, Germany; 2) Public Health Genomics, University of Virginia, Charlottesville, VA 22908; 3) DZHK (German Research Center for Cardiovascular Research), Partner site Hamburg/Luebeck/Kiel, Germany; 4) Deutsches Herzzentrum Muenchen, Technische Universitaet Muenchen, Muenchen, Germany; 5) DZHK (German Research Center for Cardiovascular Research), Partner site Muenchen, Germany.

   Aims: Glatiramer acetate (GA), which is used in the treatment of multiple sclerosis, is reported to show multiple coronary artery related adverse effects including increased blood pressure, hypercholesterolemia and coronary artery disease. Here, we study the genes affected under GA treatment to identify the disturbed genetic mechanisms underlying the adverse effect and at the same time identify new CAD risk genes. **Methods:** We identified genes and gene products reported to interact with GA using a drug gene interaction database. We traced the association signal in 200-kb region around the genomic position of these genes in around 60,000 cases and 123,000 controls and validated our findings in a second meta-analysis consisting of additional ~20,000 cases and ~ 80,000 controls. **Results:** Four genes were reported to interact with GA, but only TGFB1 shows association with CAD in our analysis. We identified presumably five independent new CAD loci within the TGFB1 region. The lead SNP, rs12459996, was genome-wide significantly associated with CAD in our extended meta-analysis (p=1.58×10^{-12}). The other four assumed sub loci showed p-values ranging from 1.63×10^{-7} to 6.89×10^{-7}. The five loci show significant additive effect. **Conclusion:** Studying genes reported to interact with Glatiramer acetate, we identified the interaction with TGFB1 as the one that presumably increases the risk of CAD under treatment. TGFB1 is also a new CAD risk gene, not reported for CAD previously. Hence, indirectly, we identify a new CAD gene in this study.

615F

The genomics of ELITE endurance athletes. *D. Waggott[1], M. Mattsson[1,2], M. Wheeler[1], E. Ashley[1].* 1) Stanford University, Stanford, CA., Select a Country; 2) Karolinska Institute.

   Health exists as a spectrum from disease to some outlier physiological optimum. To date most molecular genetic research has focused on disease states and less on extreme health populations. We hypothesize that interrogating outlier elite endurance athletes, with strict physiological eligibility criteria, will inform cardiovascular research through the identification of complementary pathways and therapeutic targets. Eligibility criteria for the ELITE study (Exercise at the Limit - Inherited Traits of of Endurance http://med.stanford.edu/elite.html) required a lifetime VO2max, which measures maximal oxygen consumption during peak aerobic exercise, at a threshold estimated to be attainable in less than 1 in 50,000 people (men 80ml/kg/min; women 65ml/kg/min). VO2max is reported to have substantial genetic influence (h2~0.5) and is correlated with endurance sport performance along with work efficiency. Several well documented cases of athletic outliers have been tied to rare genetic variants including the Finnish cross country skier Mäntyranta (*EPOR*) and Priscilla Lopes-Schliep (*LMNA*). In the later, the same domain of the LMNA gene is related to rare forms of muscular dystrophy. Additionally, adaptive hypoxia variations have been identified in high altitude populations in Tibet (*EPAS1*), Andes and Ethiopia. To date we have sequenced 268 ELITE participants using clinically enhanced exomes and run 550 samples on high density multi-ethnic SNP chips. Preliminary analysis has focused on a combination of rare variant curation and common variation association. Rare variation curation included prioritization of LOF variants within candidate genes related to oxygen transport, muscle physiology and metabolism (i.e. *PPARA, PPARGC1A, RYR2, ACTN3, MSTN*). Global gene screening was accomplished using in silico functionally weighted burden testing. Common variant association (the largest GWAS of its kind) has been used to support rare variant findings and identify non-coding and structural variant association signals. We believe that our methodology of combining rare LOF variants with common variation association in a population with extreme endurance physiology will systematically identify pleiotropic genes with both protective and pathogenic features similar to role of *PCSK9* and LDL levels. .

616W

Whole exome sequencing of a restrictive cardiomyopathy family: Rare variant vs polymorphisms. *M. Kapoor[1], S. Das[1], A. Biswas[1], B. Murry[1], S. Seth[2], B. Bhargava[2], V. Scaria[3], S. Sivasubbu[3], V. Rao[1].* 1) Anthropology, Delhi University, Delhi, India; 2) Department of Cardiology, AIIMS, New Delhi; 3) CSIR-IGIB, New Delhi.

Cardiomyopathies are a disease of complex nature with genetic heterogeneity. Restrictive cardiomyopathy (RCM) is complex genetic disorder, in which incidence is still unknown. Genetic heterogeneity is a marked characteristic of cardiomyopathies which cannot be defined with single driver mutation to cause diverse phenotypes. Polymorphism could be considered in sarcomere genes as a cumulative effect to cause diverse phenotypes with same driver mutations. With emergence and continued upgrading in next generations sequencing technologies provide a cost effective path for geneticist. Here, we tried to explain one rare mutation and two candidate gene polymorphism as a cause of restrictive cardiomyopathy with genetic heterogeneity and phenotypic plasticity. The study was ethically approved and written consent was taken from patient and family members. On the basis of clinical severity, early age of onset and worst prognosis, patient with parents (trio) was selected for whole exome sequencing. Good quality DNA is processed for library preparation, capture and sequencing on illumina HiSeq 2500. Segregation of rare variant F764Y in MYH7 and two polymorphisms R25529H and R23473H in TTN is found. Rare mutation is present on highly conserved region of the genes, susceptible to cause protein damage known to cause cardiomyopathies. This 764 residue in MYH7 resides in converter region, acts as a flexible joint in myosin which during contractile cycle undergo conformational changes. So, substitution of a tyrosin at this position of phenylalanine might cause the change in polarity causing disturbance in the transmitted movement causing diminished efficiency of contractile function. Rare variant is present in neck region of MYH7 and two polymorphisms are present on Fibronectin type III region of TTN protein. Protein-protein interaction of TTN and MYH7 in the contraction and relaxation mechanism of heart muscle could explain the phenotypic plasticity leading to stiffness of heart muscle as the main characteristics of restrictive cardiomyopathy rather than dilated cardiomyopathy as earlier reported. These kinds of protein-protein interaction of polymorphism and driver mutation in different gene should be consider and validated through in-vitro or in-vivo techniques to elucidate final pathway causing phenotypic plasticity and genetic heterogeneity among cardiomyopathy cases.

617T

Associations between cysteine-rich protein 1 (CRIP1) expression and blood pressure. *C. Müller[1,2], A. Jagodzinski[1,2], K.J. Lackner[3,6], T. Münzel[7,6], I. Hermanns[5], P.S. Wild[4,5,6], S. Blankenberg[1,2], T. Zeller[1,2].* 1) Department of General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany; 2) DZHK (Deutsches Zentrum für Herz-Kreislauf-Forschung e.V.), partner site Hamburg/Kiel/Luebeck, Germany; 3) Department of Clinical Chemistry and Laboratory Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; 4) Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; 5) Center for Thrombosis and Hemostasis, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; 6) DZHK (German Center for Cardiovascular Research), partner site RhineMain, Mainz, Germany; 7) Center for Cardiology I, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany.

Gene expression of cysteine-rich protein 1 (intestinal) (*CRIP1*) is associated with systolic (SBP) and diastolic blood pressure (DBP) in monocytes (Rotival et al. Plos Genet. 2011) and whole blood cells (Huan et al. Plos Genet. 2015) of large population-based cohorts. Until today, the role of *CRIP1* in hypertension is unknown. The aims of our study were to study longitudinal expression changes of CRIP1 protein and mRNA in a BP-lowering clinical trial and to determine cross-sectional associations between *CRIP1* mRNA and subclinical phenotypes of cardiovascular disease in a large population-based study. In the BP-lowering clinical trial TeamSta, 406 hypertensive patients were included and treated for 6 months with antihypertensive treatment. At baseline visit and 6-months follow-up, BP was measured and *CRIP1* mRNA was quantified in peripheral blood mononuclear cells by qPCR. Differential gene expression after treatment was estimated by linear mixed models and adjusted for sex and age. In the population-based Gutenberg Health Study (GHS), gene expression was measured using Illumina HT12 v3 BeadChips (n = 1285). Plasma renin activity was assessed by LIAISON Direct Renin Assay. Interventricular septal thickness (IVSD), left ventricular posterior wall thickness (LVPW), left ventricular hypertrophy (LVH) and left ventricular mass (LVM) were selected as subclinical phenotypes. Associations between *CRIP1* mRNA and plasma renin activity, SBP or subclinical phenotypes were calculated by linear regression and adjusted for sex and age. After hypertensive treatment, the mean SBP was decreased by 11.5 mmHg in TeamSta subjects. *CRIP1* mRNA was significantly reduced in all patients (p = 5.6×10^{-14}). In GHS, *CRIP1* mRNA was significantly elevated with SBP in all individuals (n = 1285, p = 1.0×10^{-19}), and after excluding subjects receiving antihypertensive treatment (n = 941, p = 3.6×10^{-13}). Increased *CRIP1* mRNA associated with IVSDd (p = 1.9×10^{-3}), LVPWd (p = 1×10^{-4}), LVH (p = 5.2×10^{-3}) and LVM (p = 9×10^{-4}). Plasma renin activity negatively associated with *CRIP1* mRNA expression in monocytes (p = 8.3×10^{-4}). *CRIP1* mRNA levels were significantly changed in relation to BP and intervention in a clinical trial. In addition, monocytic *CRIP1* mRNA associated with left ventricular hypertrophy and renin plasma activity in a large population cohort.

618F

Targeted sequencing of genes related to LDL cholesterol in patients with peripheral arterial disease. *M. Safarova[1], E. Austin[1], D. Schaid[2], I. Kullo[1].* 1) Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

   Objectives. Peripheral arterial disease (PAD) is most commonly due to atherosclerosis and is associated with significant morbidity and mortality. Low-density lipoprotein cholesterol (LDL-C) is an independent risk factor for PAD; however, whether genes related to LDL-C are associated with PAD is not known. We, therefore, tested whether: 1) genomic regions known to be associated with LDL-C are associated with PAD, and 2) these associations remain significant after adjustment for LDL-C. **Methods.** PAD cases were defined as a resting/post-exercise ankle-brachial index (ABI) ≤0.9 or ≥1.4 and/or history of lower extremity revascularization, and no history of CHD at the time of recruitment. Controls were subjects with normal ABI or no history of PAD. We performed targeted sequencing of 10 genes (*ABO, APOB, APOC1, APOE, LDLR, LPA, PCSK9, PDGFD, SORT1, TRIB1,* including coding regions, 2 kb upstream and 1kb downstream*)* identified by GWAS to harbor variants associated with LDL-C, using Illumina HiSeq 2000 (read depth ≥30). **Results**. Following quality control of the sequence data, 1470 PAD cases (mean ± SD, 69 ± 11 years, 62% men, LDL-C, 120 ± 43 mg/dL) and 1558 controls (60 ± 11 years, 56% men, LDL-C, 125 ± 36 mg/dL) were available for analysis. In 645 individuals on a statin at the time of recruitment, we imputed LDL-C levels by assuming a 25% reduction in LDL-C on therapy. Using the adaptive sum of powered score (aSPU) global association test and adjusting for age and sex, *APOE*, *APOB* and *APOC1* (all, P ≤ 0.01) were associated with LDL-C but not with PAD, whereas *ABO*, *PCSK9* and *SORT1* (all, P < 0.001) were associated with PAD but not with LDL-C. After additional adjustment for LDL-C these genes remained significantly associated with PAD. **Conclusion**. This study provides novel data on the association of LDL-C-related genes with the presence of PAD. In our study sample, *PCSK9*, *SORT1* and *ABO* were associated with PAD despite a lack of association with LDL-C. Further research is warranted to investigate whether non-lipid-related effects of *ABO, PCSK9* and *SORT1* mediate these associations and, therefore, help identify novel therapeutic targets for PAD.

619W

Integrated whole genome sequencing and transcriptome study identifies rare mutations in atherosclerosis and thrombosis pathways in a case of severe restenosed atherothrombosis. *A.M. Veerappa[1], S.M. Dinesh[2], N.B. Ramachandra[1], D. Sameer[3], J. Shah[4].* 1) Department of Genetics and Genomics, University of Mysore, Mysore, Karnataka, India; 2) NUCSER, KS Hegde Medical Academy, Nitte University, Mangalore-575 018, India; 3) Department of Cardiology, Apollo Hospitals, Ahmedabad-382428, India; 4) Dept of Cardiology, HCG Multi Speciality Hospital, Ahmedabad-380 006, Gujarat, India.

   The globally debated 34-year-old male case with rare severe restenosed atherothrombosis was presented to us with 8 clogged arteries resulting in myocardial infarction, ischaemic stroke, type 2 diabetes and hypertension. This study assessed the possible cause for severe restenosed atherothrombosis in the present case recruited. We performed whole genome sequencing (WGS) and transcriptome profiling and disease pathway analysis. Here we report the consequences of novel and rare damaging mutations in *LRP6, APOBR*, *APOH*, *ACAT1*, *ITGAM* and *STAB2* on the cellular pathways resulting in disease. Overlaying the transcriptome profile on mutant genes resulted in the disruption of several processes such as LDL uptake, clearance, inflammation and thrombosis. Enrichment analysis of the dysregulated genes showed significant elevations in inflammatory processes and cardiovascular diseases. Integrative analysis of WGS and transcriptome profiling identified mutations in rare genes, and these were specifically involved in causal processes of Endothelial Dysfunction, Reverse Cholesterol Transport, Plaque Development, Inflammation and Plaque Rupture, and Thrombosis. The findings indicate Atherosclerosis/Atherothrombosis to be a highly heterogenous disease and the conclusions of this study will help in better understanding of the significant contributions of genetics towards the etiology of this disease.

620T

Homozygous *TPM1* mutation in a familial pediatric dilated cardiomyopathy. *K. Al Harbi[1], K. Thangaraj[2], A.M. Abdallah[3], S. Justin Carlus[1].* 1) Pediatrics Department, College of Medicine, Taibah University, Al-Madinah, Saudi Arabia; 2) Centre for Cellular and Molecular Biology, Hyderabad, India; 3) West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust , Birmingham , UK.

Cardiomyopathies represent the second largest group of patients in sudden cardiac death (SCD) victims, next to coronary artery disease. Cardiomyopathy's causal pathology usually lies in the genes that are responsible for generating the mechanical properties of the heart.   Methods-We sought to identify the genetic basis of severe, non-syndromic Dilated cardiomyopathy (DCM) diagnosed in a Saudi family with 2 out of 3 children affected with DCM and 1 affected with Patent ductus arteriosus (PDA). Their parents displayed no phenotypic evidence of DCM on screening by echocardiography. We applied targeted sequencing and custom data analysis and interpretation pipelines to identify pathogenic base substitutions and insertions and deletions in 181 genes associated with cardiomyopathies. Our sequencing panel covered coding exons, exon-intron boundaries and known variants in intronic regions of target genes.   Results: Genetic analysis revealed a homozygous missense mutation in TMP1 gene (c.7G>A, p.Gly3Arg) in all three affected individuals. This gene is a member of the tropomyosin family of highly conserved, widely distributed actin-binding proteins involved in the contractile system of striated and smooth muscles and the cytoskeleton of non-muscle cells. Mutations in this gene are associated with DCM, hypertrophic cardiomyopathy (HCM) and Left ventricular noncompaction.   Conclusion: Using targeted sequencing strategy we have successfully identified a likely pathogenic variant within the TMP1 gene and provided evidence that TMP1 is a risk gene for DCM.

621F

Factor analysis of clustered cardiovascular risk in adult Punjabi female population. *V. Bains[1], B. Doza[1,2], K. Kaur[1].* 1) Human genetics, guru nanak dev university, Amritsar, India; 2) Director, Aligarh Muslim University off-campus, Murshidabad, West Bengal.

Cardiovascular Disease (CVD) is a abnormal functioning of heart or blood vessels. Hypertension, hyperlipidemia and heart failure are among the most commonly seen cardiovascular disease. Metabolic syndrome including hypertension, elevated triglycerides, increased abdominal fat, low HDL and elevated fasting blood glucose increases the risk of CVD even further. The objective of the present study was to determine significant cardiovascular risk among obese and non-obese adult Punjabi females. A total of 200 females (aged 21-37 years) were recruited from Amritsar region of Punjab with 100 obese and 100 normal females. Principal component factor analysis (PCFA) was applied to extract orthogonal components from anthropometric and physiometric variables. Association between components were explained by factor loadings. PCA was performed with varimax rotation to reduce 14 intercorrelated variables into three and four independent factors for the obese and non obese females respectively. The three factors explained 74.73% of total variance among obese females and four factors explain 73.98% among non-obese females. Factor 1 was loaded (>0.5) with high BMI, WC, WHR, Bicep skinfold and SBP among both the obese and non-obese females. Hip circumference, tricep skinfold and subscapular skinfold for obese females and age, DBP, MBP and pulse rate had maximum loadings on factor 2 for obese females while hip circumference, tricep skinfold and subscapular skinfolds had highest loadings for non obese females. The anthropometric variables such as weight, BMI, WC, WHR and bicep skinfolds were grouped together on factor 1 among both the obese and non-obese adult females. Physiometric variables SBP, DBP, MBP and pulse rate are grouped together on factor 1 for obese females and factor 2 for non-obese females. Hence, this may be interpreted from present analysis that both the anthropometric and physiometric varibles are equally important indicator for the development of CVD among both the obese and non-obese adult females. Further common greater communality estimates (>0.75) found on weight, WC, HC, tricep skinfolds, DBP and MBP among both the obese and non-obese females shows that these parameters may be consider as good predictors of CVD.   .

622W

The impact of genetic variation on doxorubicin-induced cardiotoxicity.
J.D. Blischak[1,2], D.A. Knowles[3,4], C.K. Burrows[1], K.W. Patterson[1], C. Ober[1], J.K. Pritchard[3,5,6], Y. Gilad[1]. 1) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 2) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA; 3) Department of Genetics, Stanford University, Stanford, CA, USA; 4) Department of Radiology, Stanford University, Stanford, CA, USA; 5) Department of Biology, Stanford University, Stanford, CA, USA; 6) Howard Hughes Medical Institute, Stanford University, CA, USA.

Doxorubicin, a commonly prescribed cancer drug, has cardiotoxic side effects in a subset of patients. The ability to predict who will experience these cardiotoxic side effects would enable physicians to prescribe optimal dosage regimens for each patient. To investigate the genetic basis of doxorubicin-induced cardiotoxicity, we developed an *in vitro* cardiomyocyte model. We obtained 46 immortalized lymphoblastoid cell lines (LCLs) derived from blood samples collected from adults unselected with respect to cardiovascular disease. We reprogrammed these LCLs into induced pluripotent stem cells (iPSCs) and differentiated the resulting iPSCs into cardiomyocytes. We then treated the cardiomyocytes with variable doses of doxorubicin and measured the extent of cardiac damage by i) assessing levels of cardiac troponin, a clinical cardiac stress marker, and ii) sequencing RNA to identify gene expression changes. Incorporating previously available genotype data, we identified response expression quantitative trait loci (response eQTLs), i.e. genetic variants associated with gene expression levels only pre- or post-treatment with doxorubicin. Next we trained a classifier using the response eQTLs as the input to predict the extent of cardiac damage in the *in vitro* model as measured via cardiac troponin. Our results provide insight into the mechanisms of doxorubicin-induced cardiotoxicity and provide a framework for the incorporation of patient genotypes in determining personalized doxorubicin dosage.

623T

A role for the non-sarcomeric protein WTIP in hypertrophic cardiomyopathy.
H. De Jong[1], F.E. Dewey[2], F. Lan[2], R. Lui[2], A. Erbilgin[2], M.T. Wheeler[2], S.P. Cordero[3], A. Werdich[4], R.S. Chen[1], S. Pan[2], Y. Kim[2], K. Bommakanti[2], J.M. Bos[6], F. Haddad[2], V. Froelicher[2], M. Ackerman[6], R.S. Vasan[5], C. McCrae[4], J.C. Wu[2], M. Snyder[1], E.A. Ashley[1,2]. 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Medicine, Stanford University, Stanford, CA; 3) Training program in biomedical informatics, Stanford University, Stanford, CA; 4) Department of Medicine, Harvard University, Boston, MA; 5) Department of Medicine, Boston University, Boston, MA; 6) Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN.

Hypertrophic cardiomyopathy (HCM [MIM 192600]), a common cause of sudden death in young people, is an inherited condition characterized by a thickening of the left ventricular wall of the heart. While many cases of HCM are caused by variants in genes that encode the cardiac sarcomere, a large proportion of cases have an unknown mechanism. Using exome sequencing and identity-by-descent analysis, we identified a putatively causal novel variant in Wilms Tumor Interacting Protein (WTIP [MIM 614790]) in a three-generation HCM kindred. Induced pluripotent stem cell-derived cardiomyocytes from family members with this variant recapitulated the hypertrophic phenotype. In neonatal rat ventricular myocytes, shRNA knockdown of WTIP increased phenylephrine-induced hypertrophy. An association between the genomic region surrounding *WTIP* and left ventricular thickness in a large population cohort also indicates a role for WTIP in complex forms of hypertrophy. Together, these results suggest a novel non-sarcomeric mechanism for HCM.

624F

Genetic risk factors of arterosclerosis in Turkish Cypriot population. *M.C. Ergoren[1,4], E. Ozerkman[2], A. Bostanci[2], S.G. Temel[3], G. Mocan[1].* 1) Medical Biology, School of Medicine, Near East University, Nicosia, Cyprus; 2) Medical Genetic Laboratory, Near East University Hospital, Near East University, Nicosia, Cyprus; 3) Histology and Embryology, School of Medicine, Near East University, Nicosia, Cyprus; 4) Experimental Health Sciences Research Centre, Near East University, Nicosia, Cyprus.

Genetic variation is a rich source of knowledge for cardiovascular disease because many, if not all, cardiovascular disorders are highly heritable. Genetic risk scores are a useful tool for examining the cumulative predictive ability of genetic variation on cardiovascular diseases arterosclerosis. Important considerations for creating genetic risk scores include the choice of genetic variants, biochemical parameters, and ethnicities. The questions still remain about the ultimate clinical utility of the genetic risk score, further investigation in high-risk populations and new ways to combine genetic risk scores with traditional risk factors may prove to be fruitful. To investigate the arterosclerosis genetic risk score profile, we compared 200subjects with a cardiac problem and 200 without; we based on HapMap, 1000 genome and dbSNP datas and picked previously identified 36 different SNPs on 24 different genes that are suggested to have association with arterosclerosis for different populations. This study is the first analysis of the highest SNP coverage that shown the association of genetic variants with arterosclerosis in North Cyprus. Our data is the first data shown the association of all 24 gene and 36 polymorphism to arterosclerosis and thus these data are demonstrating the cardio-genetic profile of North Cyprus. North Cyprus has a unique mixture of allele distribution for each SNP to the other close by country neighbors. Thus, SNP-SNP interactions and also their relation with biochemical pathways might play critical role for developing genetic related diseases like CVD, metabolic syndromes etc. To conclude, this study will help for understanding the genetic profile of arterosclerosis in the Island and also will be great source and useful tool for prevention of arterosclerosis.

625W

RNA sequencing of human atrial tissue implicates members of the *CLDN* **gene family in atrial fibrillation.** *A. Isaacs[1,2,3], J. Winters[1,4], A. Witten[5], S. Zeemering[1,4], E. Bidar[6], B. Maesen[6], J. Maessen[1,6], H. Crijns[1,7], M. Stoll[1,2,3,5], U. Schotten[1,4], CATCH-ME Consortium.* 1) CARIM School for Cardiovascular Diseases, Maastricht University, Maastricht, Netherlands; 2) Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, Netherlands; 3) Dept. of Biochemistry, Maastricht University, Maastricht, Netherlands; 4) Dept. of Physiology, Maastricht University, Maastricht, Netherlands; 5) Dept. of Genetic Epidemiology, Institute for Human Genetics, University of Muenster, Germany; 6) Dept. of Cardiothoracic Surgery, Maastricht University, Maastricht, Netherlands; 7) Dept. of Cardiology, Maastricht University, Maastricht, Netherlands.

Atrial fibrillation (AF), a complex, multifactorial disease, is the most common form of arrhythmia, imposing substantial burdens on health care systems, and predisposing patients to adverse outcomes, such as stroke and heart failure. Next-generation RNA sequencing is a valuable tool for elucidating genes, pathways, and variants that either play a role in disease etiology or are sensitive markers of disease and/or progression. In the current pilot of an ongoing study, right atrial tissue biopsies were obtained for 35 patients during open-heart procedures performed at Maastricht University Medical Centre (19 sinus rhythm (SR), 8 paroxysmal AF (ParAF), 8 persistent AF (PersAF)). Messenger RNA was extracted from these samples and directional, paired-end libraries were prepared (NEBNext Ultra Directional RNA Library Prep Kit) and sequenced (Illumina NextSeq 500). The well-known Tuxedo Suite was used to perform sequence alignment and feature identification, quantification, and differential expression. Differential expression analysis was used to compare the three groups (SR, PersAF, and ParAF); results were considered significant if non-zero expression was observed in each group, the effect in fold-change was ≥ 2, and the experiment-wide false discovery rate was ≤ 0.05. 52 genes met these criteria for the SR vs. ParAF comparison, 102 for SR vs. PersAF, and 121 for ParAF vs. PersAF. There were also substantial intersections of genes between comparisons (SR vs. ParAF ∩ SR vs. PersAF = 12, SR vs. ParAF ∩ ParAF vs. PersAF = 18, and SR vs. PersAF ∩ ParAF vs. PersAF = 56). One gene, a member of the claudin family, was differentially expressed in all three pairwise comparisons; of note, this gene was also identified as having differential isoform expression in each of the three analyses. A second member of this gene family was differentially expressed in two comparisons (SR vs. PersAF and ParAF vs. PersAF), while a third was significant in the ParAF vs. PersAF analysis (and nominally significant for the other two). Claudin proteins are crucial for the formation and maintenance of tight junctions. Family members are often co-expressed and create complexes; these play critical roles in permeability to sodium and potassium ions, among others, polarity, and signal transduction. Continuing ascertainment and sequencing of patients will allow for a much more fine-grained and sensitive analysis of the role of atrial tissue expression, including claudins, in AF.

626T

eQTL analysis of megakaryocytes (MKs) derived from induced pluripotent stem cells (iPSCs). *K. Kammers[1], M.A. Taub[1], I. Ruczinski[1], J. Martin[2], L.R. Yanek[2], A. Frazee[1], Y. Gao[3], D. Hoyle[3], N. Faraday[2], D.M. Becker[2], L. Cheng[3], Z.Z. Wang[3], J.T. Leek[1], L.C. Becker[2], R.A. Mathias[2].* 1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) The GeneSTAR Research Program, Johns Hopkins School of Medicine, Baltimore, MD; 3) Division of Hematology and Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore, MD.

Understanding the biology of platelet aggregation is important to prevent inappropriate vascular thrombosis. GWAS studies have identified common variants associated with platelet aggregation, but because they are intronic or intergenic, their biological link to platelet function is unclear. To explore potential function of genetic variants in this context, we have produced pluripotent stem cells (iPSCs) from mono-nuclear cells in subjects from our parent platelet aggregation studies, and then derived megakaryocytes (MKs), the precursor cells for anucleate platelets, from the iPSCs to determine patterns of gene expression in the MKs related to specific genetic variants. For 198 MK cell lines we generated genotype data on the Illumina 1M GWAS array with 1,003,451 SNPs, and RNA-seq data from extracted non-ribosomal RNA. To estimate transcript expression, we aligned and assembled raw RNA-seq reads with the Tuxedo pipeline and conducted statistical downstream analyses in R using the Ballgown and Matrix eQTL packages. In particular, we present necessary statistical considerations that include transformation and filtering of the genomic data sets as well as adjustments for known and unknown batch-effects in the statistical models. We show that failing to make such adjustments can result in spurious apparent associations between genotype and gene expression. Including iPSC-derived MKs on 108 European American (EA) and 90 African American (AA) subjects, we identify a total of N=7,190 and N=525 cis-eQTLs in the two racial groups (q<0.05), respectively. Given unequal capture of common variation on GWAS arrays between African and European ancestry subjects as represented here, we rely on a distance-based replication scheme to look for eQTL replication between the two groups (replication at a distance of 5kb). Of the 7,190 cis-eQTLs discovered in the EAs, 328 were replicated in AAs (q<0.05), and of the 525 cis-eQTLs discovered in the AAs, 261 were replicated in the EAs (q<0.05). A high number of the detected cis-eQTLs (38%) are unique to MKs compared to 44 other tissue types that are reported in the latest version of the GTEx Project. We are currently performing integrative analysis by including the eQTL signatures in our GWAS for platelet aggregation phenotypes to identify potential mechanisms of action for previously identified GWAS loci, and to identify additional GWAS signals that fail stringent Bonferroni thresholds in the absence of prior functional relevance.

627F

Combining conventional risk factors and genetic variants to predict the risk of hypertension in Taiwanese population using Taiwan Biobank. *C.W. Lin[1,2], C.Y. Shen[1,2].* 1) Taiwan Biobank, Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Hypertension is a common complex disorder and results from multiple genetic and environmental risk factors and their complex interactions. Proper lifestyle modification could prevent and delay the onset of hypertension in individuals at high risk, thus it is particularly important to develop risk prediction tools for the use in population-base screening and prevention programs. This study aimed to construct a risk assessment model, combining conventional risk factors and genetic variants, to predict the risk of hypertension in Taiwanese population. We used data from health cohort of Taiwan Biobank, which provide phenotypic information regarding lifestyle patterns, dietary habits, environmental exposure histories, health status, physical examinations, and biochemistry measurements of 53,613 subjects (stage 1) and whole-genome genotypic information of 14,432 unrelated subjects using TWBv1.0 chip (653,291 single nucleotide polymorphisms, SNPs, specifically for the Han Chinese in Taiwan) (stage 2). We built a conventional risk model, including age, sex, alcohol consumption, family history of hypertension, body mass index, fasting plasma glucose, triglyceride and microalbumin in stage 1, and the area under the receiver operating characteristic curve (ROC-ACU) was 0.855. Twenty-six SNPs, each of which reflected a genetically independent locus, were found to be associated with the risk of prevalence hypertension in stage 2. The genetic risk model based on genetic risk score incorporated 26 SNPs was significantly associated with prevalence of hypertension (Odds ratio=1.96, 95% confidence interval=1.80-2.13, P<0.001), and the AUC-ROC was 0.623. Though, the addition genetic risk model into conventional risk model slightly improves the ROC-AUC from 0.855 to 0.873, the magnitude of this increase was significant (P<0.001). A prediction model that includes both conventional risk factors and genetic variants could increase the predictive power and helps to classify individuals into risk subgroups for prevention programs.

628W

Family based whole exome sequencing in left-sided heart defects. *K.L. McBride[1256], D. Corsmeier[13], G.A. Zender[12], S.M. Fitzgerald-Butt[18], L. Kelly[14], B. Kelly[13], J. Fitch[13], C.C. Askwith[57], G.E. Herman[1456], H. El-Hodiri[1456], V. Garg[12568], P. White[1356].* 1) Nationwide Children's Hospital, Columbus, OH; 2) Center for Cardiovascular Research; 3) Center for Microbial Pathogenesis; 4) Center for Molecular and Human Genetics; 5) College of Medicine, Ohio State University; 6) Department of Pediatrics; 7) Department of Neuroscience; 8) The Heart Center.

Congenital heart disease (CHD) involving the left side of the heart (left-sided lesions–LSLs) constitute 10% of all CHD but a disproportionate amount of CHD morbidity and mortality. LSLs (aortic stenosis, aortic coarctation, hypoplastic left heart) demonstrate a strong genetic component, but few specific disease genes have been identified. Specifically, case-control whole exome sequencing (WES) studies of non-syndromic CHDs are able to identify likely causative variants in only a few percent. We hypothesized that a family-based approach may be more successful at finding disease genes than case-control designs. A total of 21 multiplex LSL families (2 – 5 affected individuals) were recruited. Linkage analysis and WES were performed to identify segregating possible pathogenic variants. WES data was filtered by segregating, rare variants (freq<0.001) with *in silico* prediction of pathogenicity. High priority candidate genes based on biological role were functionally tested. Three families harbored predicted pathogenic variants in genes previously implicated in CHD, including *MATR3, MCTP2,* and *NOTCH1*. Two additional families had variants with functional consequences in novel CHD genes. A variant in *NRP2* was studied by morpholino knock-down in Xenopus. The knockdown phenotype of ventricular hypoplasia and dextrocardia was rescued by wild-type human *NRP2*, but not the *NRP2* variant. A variant in *BMP10* demonstrated decreased protein expression and excretion in cell based assays. Several other candidates did not show functional effects. A variant in the potassium channel gene, *KCNJ2*, had normal electrophysiological properties in Xenopus oocyte studies, while a variant in a downstream target of *NOTCH1, HEY1,* had normal transactivation in cell based luciferase assays. Many families had multiple plausible candidates, thus burden studies were performed using Ingenuity Pathway Analysis by comparing these 21 families to 21 control families of similar structure. Rare variants in genes important in heart development were significantly over-represented. The use of multiplex families results in a higher yield for identifying likely pathogenic variants (5/21) than case-control designs, but bottlenecks exist in rapid functional screening of variants to identify true pathogenic changes. While non-coding variants may be causal in families without exonic changes, the genetics of CHD, even in those families with apparent Mendelian inheritance, is likely complex and oligogenic. .

629T

HUNTing for genetic causes of atrial fibrillation. *J.B. Nielsen[1,2], W. Zhou[2,3], L. Fritsche[3,4], E.M. Schmidt[3], M. Lin[1,2], M.B. Elvestad[4], C.M. Brummett[5], O.L. Holmen[4,6], C.R. Abecasis[1,3], K. Hveem[4,7], C.J. Willer[1,2,3].* 1) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan; 2) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan; 3) Department of Bioinformatics and Computational Medicine, University of Michigan, Ann Arbor, Michigan; 4) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway; 5) Department of Anesthesiology, University of Michigan, Ann Arbor, Michigan; 6) St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 7) Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, Levanger, Norway.

Atrial fibrillation (AF) is a common cardiac arrhythmia and a major risk factor of stroke, heart failure, and death. Despite an estimated heritability of ~60%, only a few loci have so far been identified as genetic risk factors of AF in the population. We undertook an array-based genome-wide association study (~600,000 markers with enrichment for coding variants) in individuals from the ethnically homogeneous HUNT study, a community-based health-survey in Norway. The cohort included 5,584 individuals with and 55,426 without AF. We identified one novel susceptibility locus for AF at chromosome 2 (leading SNP rs3731746, p.Thr17571Lys, MAF = 0.18, OR = 1.19, 95% CI 1.12 – 1.26, P-value = 1.9×10^{-9}) that was replicated in an independent hospital-based US cohort of 964 AF cases and 5,487 controls matched on age sex and inferred ancestry (OR = 1.15, P-value = 0.04). The novel locus comprised several highly correlated missense variants situated in the A and M-band of titin (TTN), the largest human protein and responsible for the passive elasticity of heart and skeletal muscle. Protein-truncating variants in TTN have been previously linked to dilated cardiomyopathy, and dilated cardiomyopathy is a risk factor for AF, so we performed a sensitivity analysis by conditioning on cardiomyopathy. This revealed results very similar to the unconditioned analysis, indicating a mechanism that does not act through overt cardiomyopathy. In addition to the novel locus, we confirmed three known loci for AF (PITX2, NEURL, and ZFHX3). Association tests, including burden tests, based on imputation of another ~15 million markers are currently ongoing. In conclusion, we have identified a novel locus for AF at TTN, a strong biological candidate for structural remodeling of the myocardium, that could help improve our understanding of this clinically important cardiac arrhythmia.

630F

Identifying candidate microRNA biomarkers for hemodynamic stability in patients with mild ischemic stroke or transient ischemic attack. *A.P. Pilbrow[1], P. Allan[2,3], J. Faulkner[4], T. O'Donnell[2], J. Lanford[5], B. Woolley[6], D. Lambrick[7], L. Stoner[6], Y-C. Tzeng[2,3].* 1) Christchurch Heart Institute, University of Otago Christchurch, Christchurch, New Zealand; 2) Centre for Translational Physiology, University of Otago Wellington, Wellington, New Zealand; 3) Department of Surgery and Anaesthesia, University of Otago, Wellington, New Zealand; 4) Department of Sport and Exercise, University of Winchester, Winchester, United Kingdom; 5) Department of Neurology, Wellington Hospital, Wellington, New Zealand; 6) School of Sport and Exercise, Massey University, Wellington, New Zealand; 7) Faculty of Health Sciences, University of Southampton, Southampton, United Kingdom.

Background. Biomarker discovery for stroke has proved difficult due to the slow release of proteins across the blood-brain barrier. However, recent studies suggest that microRNAs (small RNAs that regulate gene expression) can move across the blood-brain barrier. Interestingly, both microRNAs and markers of hemodynamic stability (blood pressure variability, BPV; cerebral blood flow variability, CFV; cerebral autoregulation, CA) may have utility as prognostic markers in stroke. This study aimed to test associations between plasma microRNA levels and hemodynamic stability parameters in patients recently diagnosed with mild ischemic stroke or transient ischemic attack. **Methods**. We measured plasma levels of 372 microRNAs in 10 male patients (mean age ± standard deviation, 71.4 ± 8.9 years) with mild ischemic stroke (n = 5) or transient ischemic attack (n = 5) at 1 week (baseline) and 14 weeks after hospital admission, using real-time PCR (Exiqon). Fluctuations in blood pressure and cerebral artery blood flow were measured in the affected brain hemisphere by spectral power analysis. Cerebral autoregulation was determined by wavelet phase synchronization analysis. Associations between plasma microRNA levels and markers of hemodynamic stability at baseline were tested with Pearson correlation. Changes in microRNA levels over 14 weeks were analyzed with paired t-tests. **Results**. Of the microRNAs tested, 82 were detectable in the majority of patients. At baseline, levels of four microRNAs, let-7d-3p, miR-18b-5p, miR-26a-5p and miR-451a, were strongly correlated with CFV (r = 0.93, -0.89, -0.88 and -0.79 respectively, p < 0.05). Levels of miR-126-5p were negatively correlated with worse CA (r = -0.76, p = 0.03). Of these microRNAs, only let-7d-3p differed between baseline and 14 weeks, with median levels 8-fold higher at 14 weeks (p = 0.033). **Conclusions**. Our data suggest that let-7d-3p, miR-18b-5p, miR-26a-5p and miR-451a may be candidate biomarkers for CFV in patients with mild ischemic stroke or transient ischemic attack. Moreover miR-126-5p may be a marker for poor cerebrovascular control in these patients.

631W

Familial case of antithrombin-III deficiency caused by new mutation in the *SERPINC1* **gene.** *M.E. Polyak, E.V. Zaklyazminskaya.* Petrovsky Russian Research Centre of Surgery, Moscow, Abricosovsky lane, 2.

Introduction. Inherited deficiency of antithrombin-III is relatively rare (1:10000) autosomal-dominant disorder leading to a high risk of venous thromboembolism. Plasmatic levels of AT-III are usually less than 50% from normal values in the most of patients. Mutations in the *SERPINC1* gene may lead to Quantitative (type I) and Qualitative (type II) types of AT-III deficiency by different molecular mechanisms. Usually mutations leading to haploinsufficiency lead to AT-III type I. Many of known missense mutations cause AT-III type II, with normal levels of dysfunctional protein. **Materials and methods**. We observed a 29 y.o. female proband with a history of venous thrombosis at the age of 18. Father and grandmother have had thromboembolic events. The level of AT-III was measured with routine methods; mutation screening in *SERPINC1* gene was performed by bi-directional Sanger sequencing. Resources PolyPhen2, SIFT, NetGene2, and Splice Site Predictor were used for *in silico* characterization of mutant protein. **Results**. New VUCS p.W221S in *SERPINC1* gene was detected in proband and her father, and was absent in healthy sister. This variant is absent in "1000 Genomes" and other available databases. Both PolyPhen2 and SIFT resources characterize p.W221S as probably damaging. No changes in splicing were detected using open resources. The residual AT-III levels in our patients were significantly decreased up to 44-48% (AT-III type I). **Conclusion**. We assume that novel p.W221S variant is disease-causing mutation for AT-III type I mutation. Examples of missense mutations causing quantitative are previously described. The possible pathogenic mechanism could affect post-translational processing of AT-III protein (folding, intracellular transport, or excretion). .

632T

Gene expression analysis of megakaryocytes (MKs) derived from induced pluripotent stem cells (iPSCs) shows platelet-related signature of CD41+CD42a+ percentage. *M.A. Taub[1], K. Kammers[1], I. Ruczinski[1], J. Martin[2], L.R. Yanek[2], A. Frazee[1], Y. Gao[3], D. Hoyle[3], N. Faraday[2], D.M. Becker[2], L. Cheng[3], Z.Z. Wang[3], J. Leek[1], L.C. Becker[2], R.A. Mathias[2].* 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) The GeneSTAR Research Program, Johns Hopkins School of Medicine, Baltimore, MD; 3) Division of Hematology and Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore, MD.

Platelet aggregation in blood on ruptured or eroded atherosclerotic plaques may initiate harmful cardiovascular events. As the precursor of anucleate platelets, the megakaryocyte (MK) is an important cell type in understanding platelet aggregation genetics. However, MKs are hard to collect in large numbers due to invasive sampling procedures. Our group has generated induced pluripotent stem cells (iPSCs) from peripheral blood mononuclear cells and differentiated these iPSCs into MKs. As part of a larger study on heart disease genetics, we generated transcript expression data (RNA-seq) on these samples. As previously described by our group, a marker of relevance in the derivation of MKs from iPSCs is percent CD41+CD42a+ megakaryoblasts (CD41/42%) in the differentiated MK cell pellet. We see a wide range in the CD41/42% in the final pellets used in our -omics approaches. To understand the influence of this variability in cell-differentiation stageon gene expression, we performed differential expression (DE) analysis to relate transcript abundance to CD41/42% in 198 iPSC-derived MK cell lines from 108 European American and 90 African American samples in the GeneSTAR Study. RNA-seq was performed on extracted non-ribosomal RNA. We generated gene and transcript expression values with the Tuxedo pipeline, and conducted downstream analyses in R using the Ballgown package. We transformed and filtered the expression data, and adjusted for known and unknown batch-effects in our models. We calculated DE as a function of CD41/42% and did gene ontology (GO) analysis on the results, separately for transcripts with positive and negative associations between expression and CD41/42%. We identified N=13761 (q<0.05) transcripts showing DE by CD41/42%. Limiting our analysis to N=10701 transcripts with positive association between expression and CD41/42%, of which 7889 showed significant DE, significant GO results (q<0.05) include many categories directly related to platelet function, including platelet activation, blood coagulation, platelet degranulation, and platelet aggregation. These results indicate CD41/42% is an important marker of MK cell differentiation in our setting, and is an important co-variable in the downstream analysis of iPSC-derived MK transcriptomic signatures.

633F

Sequencing identifies new variants and new genes for carotid intima-media thickness in Dominican Republicans. *L. Wang[1,2], N. Dueker[1], A. Beecham[1], S.H. Blanton[1,2], T. Rundek[1,3], R.L. Sacco[1,2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) John T. McDonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Neurology, Epidemiology and Public Health, University of Miami, Miami, Fl.

Carotid intima-media thickness (cIMT) is a subclinical marker for cardiovascular diseases (CVD). We have been using cIMT as a precursor phenotype to reduce heterogeneity in genetic studies of CVD. Previously, we reported a quantitative trait locus (QTL) on chromosome 14q for cIMT in Dominican families and identified *PRIMA1, FOXN3,* and *CCDC88C* as candidate genes via genotyping informative common variants (CVs). Herein, we sought to extend our efforts with next-generation sequencing. We sequenced all exons within the QTL, as well as introns and flanking regions of *PRIMA1, FOXN3* and *CCDC88C* in 116 individuals from 7 Dominican families with evidence for linkage at the QTL (LOD>0.1). 561 unrelated Dominicans from Northern Manhattan Study (NOMAS) with exome-chip genotyping and imputation to 1000-genome references were used for validation. For CVs (minor allele frequency [MAF]≥5%), single variant analysis using the Quantitative Transmission-Disequilibrium test (QTDT, for families) and linear regression (for NOMAS) were performed. For rare variants (RVs, MAF<5%), we performed gene-based tests using the Family SNP-set Kernel Association Test (Fam-SKAT, for families) and SKAT-O (for NOMAS). CVs in *PRIMA1* and the *SPATA7-PTPN1-*ZC3H14-*EML5-TTC8* gene cluster displayed the most robust evidence for association (P<0.005 in families and P<0.05 in NOMAS). Consistent with previous findings, the strongest association with CVs was found in *PRIMA1* (P=8x10[-5] in families, P=0.01 in NOMAS). Interestingly, the top marker is not in linkage disequilibrium (r^2<0.01) with the previous top marker (P=3x10[-4] in families, P=0.1 in NOMAS). Gene-based test suggested that RV is not a major mechanism for *PRIMA1* to contribute to cIMT (P=0.7 in both datasets). In contrast, gene-based test on exonic RVs found evidence for association in 4 genes within the 5-gene cluster (P<0.05). To further evaluate exonic RVs in the 5-gene cluster, we combined them in a gene-set-based analysis and used Combined Annotation-Dependent Depletion (CADD) to prioritize RVs. We found that the evidence for association is enriched in RVs with CADD>10 (N=15, P=0.003) compared to RVs with CADD<10 (N=31, P=0.02) in the family dataset. Our data demonstrate a complicated genetic architecture of cIMT that could not be appreciated in our previous studies: multiple independent CVs in *PRIMA1* and exonic RVs in new candidate genes. Future studies are needed to understand the contribution of these genes and variants to cIMT.

634W

Many novel genetic loci associated with blood pressure are identified from a GWAS in UK Biobank. *H.R. Warren[1,2], E. Evangelou[3,4], C.P. Cabrera[1,2], H. Gao[3], I. Tzoulaki[3], M.R. Barnes[1,2], M.J. Caulfield[1,2], P. Elliott[3], UKB-CMC BP working group and ICBP, CHARGE+, T2D-GENES, GoT2DGenes, ExomeBP, and CHD Exome+ Consortia.* 1) Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London, UK; 2) NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 3) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; 4) Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.

High blood pressure (BP) is a major risk factor for cardiovascular disease (CVD). Many BP-associated genetic loci have previously been identified, but together only explain a small percentage of the total trait variance. The UK Biobank (UKB) cohort provides a unique opportunity to perform the largest ever GWAS of BP within a single study. We have analysed systolic BP, diastolic BP & pulse pressure traits from 140,886 unrelated Europeans from UKB across ~73 million genotyped and imputed autosomal variants. Using UKB for discovery, with replication in independent study meta-analyses (N~200k), we have identified and validated ~110 novel genetic loci. Amongst the new variants, we have found rare or low-frequency variants with larger effect sizes on BP than common variants, and some exonic missense variants. Conditional analyses have been used to determine multiple independent signals within novel loci. Many of the novel BP-associated loci are also associated with hypertension (HTN). Lookups in non-European ancestries, despite lower power, suggest some novel loci are also associated in other populations, e.g. Hispanics. Integrative bioinformatics analyses have found eQTL associations and strong evidence of plausible biological function for many novel genes, as well as an enrichment of pathways implicated in CVD, and an over-representation of genes affected by anti-HTN drugs. Integration with NMR metabolomics data has detected significant downstream associations across different plasma spectral regions and for a range of lipoprotein fractions. UKB's deep imputation coverage and large sample size enables the assessment of all previously reported BP loci. As well as confirming prior findings, we have validated additional independent variants at many known BP regions, which remain significant after conditioning on known SNPs. Results for specific mutations within genes published for monogenic HTN disorders show that, despite lower power, some of these rare variants also have nominally significant associations with large effects on BP at a population level. Our large discovery yield illustrates the benefits of UKB data and our functional results show promise for identifying potential causal genes impacting HTN mechanisms, which will increase our understanding of the genetic risk and underlying biology for HTN.

635T

Integration of genetics and transcriptomics to identify missing heritability for platelet aggregation. *L.R. Yanek[1], K. Kammers[2], M.A. Taub[2], J. Martin[1], N. Faraday[1], I. Ruczinski[2], Y. Gao[3], D. Hoyle[3], Z. Wang[3], L. Cheng[3], D.M. Becker[1], J. Leek[2], L.C. Becker[1], R.A. Mathias[1].* 1) GeneSTAR Research Program, School of Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Division of Hematology and Institute for Cell Engineering, School of Medicine, Johns Hopkins University, Baltimore, MD.

Background: Missing heritability in GWAS studies is a common occurrence. Variants with lower effect sizes are routinely missed for failure to cross stringent significance thresholds required for the millions of SNPs tested. The argument for inclusion of functional relevance in the re-evaluation of these SNPs with marginal significance has gained favor in the era of Next Generation Sequencing enabling the rapid integration of genetic and transcriptomic data. Methods: GeneSTAR (Genetic Study of Atherosclerosis Risk) is an ongoing family-based study which includes platelet function assessment before and after 2 weeks of aspirin (ASA). We performed GWAS on platelet function separately in 1094 European American (EA) and 773 African American (AA) participants, adjusting for age, sex, population stratification, and family structure. In a subset (N=90 AAs and N=108 EAs) with induced pluripotent stem cell-derived megakaryoctyes (MKs), we performed RNA sequencing with the Illumina HiSeq 2500 platform. Samples were processed through TopHat for alignment, Cufflinks/Cuffmerge for assembly, and Tablemaker and Ballgown for analysis. Expression quantitative trait loci (eQTL) analyses were run separately in EAs and AAs, adjusting for age, sex, batch, CD41/42 (an index of MK differentiation), and population stratification and expression principal components. Results: In one exemplary phenotype, post-ASA aggregation to ADP 2uM, GWAS identified several loci that did not cross the traditional 5×10^{-8} threshold of significance but were suggestive with p-values between 10^{-5} and 10^{-7}. eQTL analysis in MKs identified 7,190 *cis* eQTLs in EAs and 525 *cis* eQTLs in AAs (q<0.05). By integrating the eQTL analysis with the GWAS results, we found that SNPs in a region of Chr 20 with suggestive evidence regulate the expression of *CPNE1* in MKs. rs761825 is the lead SNP in *CPNE1* with GWAS p-value in EAs $p=6.73 \times 10^{-6}$, in AAs p=0.702; eQTL $p=1.28 \times 10^{-5}$ in EAs, p=0.23 in AAs. Discussion: Leveraging the eQTL analysis with our GWAS data, we now identify a novel locus that may regulate platelet aggregation. Interestingly, *CPNE1* encodes a calcium-dependent membrane-binding protein, and has been reported to interact with mitogen activated protein kinases that are known to be important to platelet activation. We are repeating this analysis with additional platelet aggregation phenotypes and will present results from several integrated analyses.

636F

Mutation spectrum and clinical polymorphism in Russian patients with left ventricular non-compaction. *E.V. Zaklyazminskaya[1,3], M.E. Polyak[1], E.A. Mershina[2], V.E. Sinitsin[2], A.A. Bukaeva[1], O.V. Blagova[3], L.B. Mitrofanova[4], N.P. Kotlukova[5], Ju. Lutokhina[3], A.G. Shestak[1], S.L. Dzemeshkevich[1,3].* 1) Petrovsky Russian Research Centre of Surgery, Moscow, Russian Federation; 2) Research Centre of Nuclear Imaging and Computerized Tomography; 3) Sechenov First Moscow State Medical University; 4) Federal Almazov North-West Medical Research Centre; 5) 7th Moscow City Children Hospital.

   Introduction. Left ventricular non-compaction (LVNC, MIM#300183**)** is a cardiomyopathy characterized by prominent trabeculation in the left ventricular apex, lateral wall and ventricular septum. It is not completely clear if LVNC represents a distinct clinical entity, cardiac embryogenesis disruption or a variant of clinical appearance of other types of cardiomyopathies (dilated and/or hypertrophic). Mutation spectrum, genotype-phenotype correlations, clinical polymorphism and natural course of LVNC are still to be elucidated. **Materials and methods**. Clinical and genetic evaluation of patients was performed in accordance with Helsinki Declaration. Fifty four patients with LVNC were included into the target group. Clinical and instrumental evaluation had included physical examination, familial history taking, resting ECG and 24-hours Holter monitoring ECG, Echo-cardiography, cardiac MRI or computer tomography, and myocardial biopsy (by indication); anti-myocardial antibodies measurement. The genetic study was performed by massive parallel sequencing (IonTorrent PGM) of 28 genes with following Sanger sequencing for all under covered regions. **Results**. Clinical group of LVNC patients included 28 males and 26 females. Five patients were diagnosed before 1y.o.. In the adult group (49 patients) the mean age was 41 y.o.. The follow-up period was from 6 months up to 60 months. Most of LVNC patients had decreased cardiac contractility and heart failure of different degree; 5 probands had undergone orthotopic heart transplantation. The vast majority of patients have had apparent signs of other cardiomyopathies beside LVNC: left ventricular hypertrophy, dilated cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy; 1 patient had neuromuscular disorder. The common feature was inflammation (myocarditis). Rhythm disturbances were found in >50% of patients. The autosomal-dominant mode of inheritance was the most frequent in familial cases, though sporadic cases were also present. We had identified pathogenic/potentially pathogenic genetic variants in 18% of probands tested. Two mutations were detected in two probands who underwent the heart transplantation. Most of mutations were detected in *MYH7* and *MyBPC3* genes. **Conclusion**. Most of LVNC patients have concomitant cardiomyopathy, arrhythmias and/or myocarditis. Isolated LVNC scarcely exists. Mutations in *MYH7* and *MyBPC3* genes are the most frequent finding in Russian patients with LVNC. .

637W

The predominant effect of common variation at the *SCN5A-SCN10A* **locus on susceptibility to Brugada Syndrome.** *R. Redon[1], Y. Mizusawa[2], C. Dina[1], J.B. Gourraud[1], F. Simonet[1], E.R. Behr[3], E. Schultze-Bahr[4], T. Robyns[5], J. Saenen[6], P.D. Lambiase[7], P. Guicheney[8], P. Mabo[9], M. Haissaguerre[10], M.M. Borggrefe[11], C. Antzelevitch[12], S. Kaab[13], J. Tfelt-Hansen[14], D.M. Roden[15], R. Brugada[16], G. Sarquella-Brugada[17], J. Brugada[18], S.G. Priori[19], P.J. Schwartz[20], C. Giustetto[21], J.J. Schott[1], V. Probst[1], A.A. Wilde[2], J. Barc[1], C.R. Bezzina[2].* 1) L'institut du thorax, Inserm UMR 1087 / CNRS UMR 6291 / University of Nantes, Nantes, France; 2) Department of Clinical and Experimental Cardiology, AMC, Amsterdam, Netherlands; 3) Cardiovascular Sciences Research Centre, St George's Hospital, London, United Kingdom; 4) Institute for Genetics of Heart Diseases (IfGH), Munster, Germany; 5) Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium; 6) Department of Cardiology, Antwerp University Hospital, Antwerp, Belgium; 7) The Heart Hospital, UCL Cardiovascular Institute, London, United Kingdom; 8) Inserm UMR 1166, Faculté de médecine Pierre et Marie Curie, Paris, France; 9) CHU Rennes, Service of Cardiology, Rennes, France; 10) CHU Bordeaux, IHU LIRYC, Bordeaux, France; 11) Department of Cardiology, University Medical Centre Mannheim, Mannheim, Germany; 12) Lankenau Institute for Medical Research, Lankenau Heart Institute, Wynnewood, PA, United States; 13) Department of Medicine I, University Hospital Munich, Munich, Germany; 14) Laboratory of Molecular Cardiology, Copenhagen, Denmark; 15) Department of Medicine and Pharmacology, Vanderbilt University, Nashville, TN, United States; 16) University of Girona, Hospital Trueta, Girona, Spain; 17) Hospital Sant Joan de Deu, Barcelona, Spain; 18) University of Barcelona, Hospital Clinic Barcelona, Barcelona, Spain; 19) Fondazione Salvatore Maugeri University of Pavia, Pavia, Italy; 20) IRCCS Istituto Auxologico Italiano, Center for Cardiac Arrhythmias of Genetic Origin, Milan, Italy,; 21) University of Turin, Division of Cardiology, Città della Salute e della Scienza, Turin, Italy.

   Introduction: The Brugada syndrome (BrS) is an inherited cardiac disorder associated with ST-segment elevation in the right precordial leads and a high risk for sudden cardiac death (SCD). Although rare variants in *SCN5A* (encoding the major cardiac Na-channel isoform) are found in ~20% of cases, the genetic basis of the disorder remains largely unresolved. In a previous genome-wide association study (GWAS) on 312 patients with BrS, we had identified 3 common haplotypes predisposing to the disorder: two at the *SCN5A-SCN10A* locus and one near the *HEY2* gene. Here we aimed to identify additional common SNPs modulating the susceptibility to BrS and/or SCD. Results: By extending the GWAS to a set of 877 index cases of North-Western European ancestry versus 2,833 ancestry-matched controls, we confirmed the 3 previously reported association signals at the *SCN10A*, *HEY2* and *SCN5A* loci. Conditional analysis on SNPs at the *SCN5A-SCN10A* locus revealed 2 additional independent association signals located between the *SCN5A* and *SCN10A* genes. The cumulative effect of the 5 loci on disease susceptibility reached an odds ratio >50 in the presence of ≥ 8 risk alleles versus ≤ 2. Conclusions: Our novel findings further reinforce the predominant involvement of the *SCN5A-SCN10A* locus in the disease and prompt a reevaluation of the genetic architecture of the BrS for refinement of patient care. We are currently investigating whether additional SNPs are associated with the BrS phenotype or with arrhythmia symptoms by typing additional patients. Acknowledgments: GoNL Consortium; KORA Consortium; UK10K Consortium; D.E.S.I.R. Consortium; RHYTHMOGENE Network in France.

638T

Targeted gene panel for cardiac ion channelopathies reveals candidate pathogenic mutations in Long QT syndrome. *B. Turkgenc[1,2], S.G. Temel[3,4], O. Karadag[5], H.H. Aykan[6], F. Uysal[7], I. Yildirim Bastuhan[8], A. Sulu[9], S. Ugan Atik[10], B. Cinar[10], R. Dedeoglu[10], E. Gunay[11], M. Ramoglu[12], E. Cilsal[13], M. Sahin[14], T. Mese[15], O. Ciftci[16], F. Oztunc[10], T. Karagoz[6], O. Baspinar[9], O.M. Bostan[7], F. Akalin[11], M. Kervanoglu[9], C. Ayabakan[17], Y. Alanay[18], A. Celiker[19,8], S.A. Ozer[1], M.C. Yakicier[20].* 1) Department of Medical Biology and Genetics, Marmara University, Istanbul, Turkey; 2) Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 3) Uludag University, Faculty of Medicine, Department of Histology and Embryology, Bursa, Turkey; 4) Near East University, Faculty of Medicine, Department of Histology and Embryology, Lefkosa, North Cyprus; 5) Hacettepe University, Department of Statistics, Ankara, Turkey; 6) Hacettepe University, Faculty of Medicine, Department of Pediatric Cardiology, Ankara, Turkey; 7) Uludag University, Faculty of Medicine, Department of Pediatric Cardiology, Bursa, Turkey; 8) Koç University, Faculty of Medicine, Department of Pediatric Cardiology, Istanbul, Turkey; 9) Gaziantep University, Faculty of Medicine, Department of Pediatric Cardiology, Gaziantep, Turkey; 10) Cerrahpasa Medical Faculty, Department of Pediatric Cardiology, Gaziantep, Turkey; 11) Marmara University, Faculty of Medicine, Department of Pediatric Cardiology, Istanbul, Turkey; 12) Ankara University, Faculty of Medicine, Department of Pediatric Cardiology, Ankara, Turkey; 13) Adana Numune Hospital, Pediatric Cardiology, Adana, Turkey; 14) Acibadem International Hospital, Pediatric Cardiology, Istanbul, Turkey; 15) Izmir Dr. Behcet Uz Pediatrics Hospital, Pediatric Cardiology, Izmir, Turkey; 16) Kartal Kosuyolu Training and Research Hospital, Istanbul, Turkey; 17) Baskent University Istanbul Hospital, Department of Pediatric Cardiology, Istanbul, Turkey; 18) Acibadem Maslak Hospital, Department of Pediatric Genetic, Istanbul, Turkey; 19) VKV American Hospital, Department of Pediatric Cardiology, Istanbul, Turkey; 20) Acibadem University, Faculty of Science, Molecular Biology and Genetic, Istanbul, Turkey.

Long QT syndrome is characterized by prolongation of the QT interval on electrocardiogram and associated with a high risk of sudden death. To date, mutations in 15 LQTS-susceptibility genes have been implicated. However, the genetic cause for approximately 20-25 % of LQTS patients remains elusive. We sought to add new channelopathy genes data to the existing knowledge of genetic mutations underlying LQTS to both expand our understanding of its genetic basis and assess the value of genetic testing in clinical decision-making. We screened for 68 cardiac arrhythmia genes among 92 unrelated LQT cases (9,66±5,602, 53 male and 8,69±5,676, 39 female) from Turkey. Cardiac symptoms were present in 52% of LQTS patients. Mean age of onset was 7,368±4,982. Mean QTc interval was 480,14±57,335. Targeted sequencing on ion PGM using Custom Panel revealed that known 15 LQT gene variants (79) acounted for 40,5% of the cases, while the remaining gene variants (116) accounted for 59,5% of the cases. 168 missense, 11 splice site, 4 frameshift, 6 inframe deletions, 2 nonsense and 4 UTR variants have been identified in the study. 39 (20%) of these variants were novel, 26 (%13,35) of these variants were reported as pathogenic/probably pathogenic, 40 (%20,5) of these had conflicting interpretations of pathogenicity, 64 (%32,8) of these were classified as unknown significance/nonidentified/untested and the remaining 26 (%13,35) were reported as benign/likely benign in public databases. Family segregations investigated 7 de-*novo* missense variants in *KCNQ1, CACNA1C, RYR1, CALM2* and *SCN5A* genes. While 7 patients were carrying only one pathogenic variant (3 *KCNQ1*, 1 *SCN5A*, 1 *KCNH2*, 1 *GDP1L* and 1 *CALM2*), the other pathogenic mutations (in *KCNQ1, SCN5A, KCNE1, CACNA1C, RYR2, JUP, LMNA, GPD1L, CASQ2* genes) were together with at least one other channelopathy variants. Any suspicious variant has not be detected in the remaining 10 (10,9%) patients. Age of onset was higher in those noncarriers (10,3±4,11 years) when compared to ≥ 1 variant carriers (6,90±5,00 years) (t test, p>0,05). It must be emphasized that the genetic complexity of cLQTS, currently makes it necessary to screen the other cardiac genes besides 15 LQT genes before the molecular-genetic workup is finished. We believe our findings will be an anchor point for finding novel pathogenesis of this disorder. Acknowledgements: Inherited arrhythmia panel project is supported by grant from SANTEZ (0253. STZ.2013-2), Turkey.

639F

Genetic determinants of thrombin generation and clot lysis time: Results of a targeted exome sequencing study. *H.G. de Haan[1], C. Kabrhel[2], T.P. Baglin[3], F.R. Rosendaal[1,4,5], A. van Hylckama Vlieg[1].* 1) Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands; 2) Emergency Medicine, Massachusetts General Hospital, Channing Network Medicine, Harvard Medical School, Boston, USA; 3) Cambridge Haemophilia and Thrombophilia Centre, Addenbrooke's Hospital, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, United Kingdom; 4) Einthoven Laboratory of Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands; 5) Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, The Netherlands.

Background: The thrombin generation (TG) potential and clot lysis time (CLT), which represent the ability to form and dissolve a clot, respectively, are well-established risk factors for venous thrombosis. Despite the heritability of these traits, few genetic determinants have been identified. Exploring the role of protein coding variants could provide insight into the genetic architecture and biology underlying these traits. **Aims**: We aimed to identify novel genetic determinants of TG and CLT using targeted exome sequencing. **Methods**: We included 296 controls (without venous thrombosis) from two population-based, case-control studies on venous thrombosis (MEGA and THE-VTE). Individuals were selected using the following criteria: European ancestry, no surgery or malignancy, no natural anticoagulant deficiency, and no factor V Leiden or prothrombin 20210A carriership. Targeted sequencing was performed of the coding regionsand boundaries of 740 genes involved in thrombosis and hemostasis or closely related pathways. We investigated the association of 3,002 low-frequency, common (minor allele frequency (MAF) >1%) variants with CLT and four parameters of TG: endogenous thrombin potential, peak height, time-to-peak, and lag time. Linear regression models were adjusted for age, sex, country of the study, and additionally for oral contraceptive use in the TG analyses. To account for multiple testing, we calculated false discovery rates (FDR). **Results:** We identified a common synonymous variant in *COL4A1* which was associated with time-to-peak (c.1815T>C, MAF=6.7%, β=0.65 per z-score increase, P=$1.92*10^{-5}$, FDR=6.2%). *COL4A1* encodes a major component of the type IV collagen of the basement membrane, and it is a highly replicated locus for cerebro- and cardiovascular disease risk. CLT and three of the TG parameters showed no association with any of the genetic variants observed, with the lowest FDR (22%) observed for an association between CLT and a missense variant in *TLR5*. **Conclusions:** We identified a variant in *COL4A1* as a novel determinant of time-to-thrombin peak. Replication and functional studies are warranted to further unravel the implication of *COL4A1* in TG and, potentially, venous thrombosis risk. We did not identify any variants associated with either CLT or three of the four TG parameters, which is likely attributable to our small sample size.

640W

Exome sequencing in venous thromboembolic disease identifies mutation burden in *PROS1*, *STAB2* and *SERPINC1*. *K. Desch[1], M. Halvorsen[2], A.B. Ozel[3], M. Germain[4], D. Tregouet[4], P.H. Reitsma[5], J. Li[3], D. Goldstein[2], D. Ginsburg[3].* 1) Pediatrics and Communicable Disease, University of Michigan, Ann Arbor, MI; 2) Institute for Genomic Medicine, Columbia University, New York, NY; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) INSERM, UMR_S 937; Institute of Cardiometabolism And Nutrition (ICAN), Université Pierre et Marie Curie Paris 6, Paris F-75013, France; 5) Leiden University Medical Center, Leiden, Netherlands.

　　Deep vein thrombosis and pulmonary embolism, collectively referred to as venous thromboembolism (VTE [MIM 188050]), are the third leading cause of cardiovascular death in the United States. Genetic factors account for 50-60% of VTE risk and a recent meta-analysis of genome-wide association studies confirmed that common variants in *F5* (MIM 612309), *ABO* (MIM 616093), and seven other loci are associated with VTE. Rare mutations in the anticoagulant genes *PROC* (MIM 176860), *PROS1* (MIM 612336) and *SERPINC1* (MIM 107300) have been linked to VTE in family studies. In order to identify new genetic variants altering the risk for VTE, we performed exome sequencing in 373 unrelated individuals of European ancestry with unprovoked VTE and compared results to a previously sequenced control cohort of 5784 unrelated Europeans. To avoid variant calling bias, only SNVs from exons with less than 5% difference in the average percentage of bases covered at least 10x between cases and controls were included, removing 11,813 of 188,689 intervals. To find variants associated with risk of VTE, we used an emerging framework for a "collapsing" analysis on genes. We defined qualifying variants on the basis of annotation and minor allele frequency, and in the dominant model, cases and controls were scored on the basis of having one or more qualifying variant in each gene. Tests under the dominant model were performed via a Fisher's exact test for a total of 17,395 CCDS genes. Strikingly, ranked by p-value, the top three genes were *PROS1* (P= 3.34E-08), *STAB2* (P=1.65E-05), and *SERPINC1* (P=3.21E-05). We detected 15 qualifying variants in 373 cases and 56 qualifying variants in 5784 controls in *STAB2*. This gene encodes Stabilin-2, which is a transmembrane glycoprotein scavenger receptor. Common variants at *ABO* are associated with VTE and are also known to regulate von Willebrand Factor (VWF) and coagulation Factor VIII (F8) plasma levels. Common variants in *STAB2* are also associated with VWF/F8 levels in a large GWAS and with VTE risk in a smaller candidate gene study, suggesting that haploinsufficiency for Stabilin-2 may increase VTE risk through elevated levels of VWF/F8. Although replication and functional testing of these findings is warranted, this study demonstrates the utility of collapsing analyses using WES data to identify multiple loci harboring an excess of rare variants in individuals with a common complex disease trait.

641T

Common and rare variants in previously identified linkage region on chromosome 12p associated with left ventricular mass in Dominican families. *N.D. Dueker[1], S. Guo[1], A. Beecham[1], L. Wang[2], S.H. Blanton[2], C. Dong[3], D. Cabral[3], T. Rundek[3], R.L. Sacco[2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami - Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Neurology, Epidemiology and Public Health, Miller School of Medicine, University of Miami, Miami, FL.

　　Increased left ventricular mass (LVM), an intermediate phenotype for cardiovascular disease (CVD), is predictive of CVD and stroke. Using families from the Dominican Republic, we have previously shown LVM to be heritable and found evidence for linkage to 12p. For our current study, we aimed to further characterize the QTL by sequencing the 1 LOD unit down region on 12p in ten extended families from the Dominican Republic with evidence for linkage to LVM (family specific LOD score > 0.1). Using these data, we tested 5,477 CVs (minor allele frequency [MAF] ≥5%) using the QTDT test. Gene-based analyses were performed to test rare variants (RVs; MAF<5%) in 181 genes using the family-based sequence kernel association test. A sample of 618 unrelated Dominicans from the Northern Manhattan Study (NOMAS) with Illumina Exome Array data was used for replication analyses. In our analysis of CVs, no significant associations were observed. However, among the top CVs, rs12299972 (p=7.0x10^-4), an intronic variant in *SLC38A1*, was associated with increased LVM and showed evidence for replication in NOMAS (p=0.02). Exonic RVs in *SLC38A1* also showed moderate association in the families (p=0.07) and NOMAS (p=0.06) suggesting that both common and rare variants in *SLC38A1* may be involved in LVM. Exonic RV analyses identified an additional candidate gene, *PKP2*, which was our most strongly associated gene (p=0.005) with suggestive evidence for replication in NOMAS (p=0.10). This association remained in the families and was even stronger in NOMAS sample, when analyses were restricted to non-synonymous variants (p<0.05 in both samples). *PKP2* encodes plakophilin 2, a protein found in desmosomes, which are protein structures in cell membranes that maintain adhesion between neighboring cells. Desmosomes are found in tissues that experience mechanical stress such as the myocardium. Desmosomal genes, including *PKP2*, have been implicated in arrythmogenic right ventricular cardiomyopathy and recently, arrhythmogenic left ventricular cardiomyopathy as well, making *PKP2* an excellent candidate gene for LVM. In conclusion, re-sequencing of our previously identified QTL identified common and rare variants within two candidate genes, *SLC38A1* and *PKP2*, to be associated with LVM. Future studies are necessary to elucidate the role these variants play in the pathogenesis of increased LVM.

642F

Genome-wide association study and linkage study for histidine-rich gly-coprotein levels in a healthy young cohort. *A. Ozel[1], K. Desch[2], D. Sieme-niak[3], D. Ginsburg[1,3], J.Z. Li[1].* 1) Department of Human Genetics, University of Michigan- Ann Arbor; 2) Department of Pediatrics, University of Michigan- Ann Arbor; 3) HHMI, Department of Human Genetics, University of Michigan- Ann Arbor.

Through its inhibition of Factor XIIa, histidine-rich glycoprotein (HRG) regulates the intrinsic pathway of blood coagulation. HRG has a domain structure that facilitates interactions with multiple ligands, potentially contributing to pleotropic roles in coagulation, immunity and angiogenesis. We performed genome-wide association studies in a healthy sibling cohort of 1,152 subjects (in 489 sibships), focusing on the European subset (n = 940), and a second healthy cohort of 2,304 individuals (in 72 sibships) of Irish descent (Desch et al., 2012). HRG plasma levels were determined by AlphaLISA using a polyclonal HRG antibody. Heritability (h^2) was estimated as 69% (using *Merlin-Regress*), consistent with results from intra-class correlation (66%) and *GCTA* (60%). Common variants at *HRG* on chromosome 3 showed significant association with HRG plasma levels (P < 5.0E-8). The top candidate SNP was rs9898(T) (+14% /minor allele, MAF = 0.33, P = 6.6E-115), a non-synonymous variant located at the exon 5 of the *HRG* gene. Meta-analysis of the two cohorts confirmed the signals on chromosome 3, with best p-value of 6.2E-348 with the same effect direction. These SNPs collectively explained 47% of the variation in the HRG levels. Rs9898(T) was also detected in a previously published genome-wide study of activated partial thromboplastin time (aPTT) suggesting that altered HRG levels are a major influence in the variance of the aPTT coagulation test. Linkage analysis using the sibling subset of the two cohorts identified significant signals at three intervals: a ~ 0.6 cM interval on 3q28 (LOD=11.1, permutation P < 0.05), which is near the *HRG* gene; a ~ 0.7 cM interval on 2q37.1-2q37.2 (LOD=4.2, permutation P < 0.05); and a ~ 2.7 cM interval on 18q21.1 (LOD=2.6, permutation P < 0.05). The three peaks in the linkage analysis explained 2.5 % of the variation in the HRG levels. To our knowledge, this is the first genome-wide study investigating histidine-rich glycoprotein level variation in a healthy population. These results provide new insight into the genetic control of circulating HRG levels, and may aid the clinical interpretation of aPTT levels. Furthermore, linkage analysis revealed loci undetected in association studies suggesting allelic heterogeneity at specific genes in these three linkage intervals.

643W

Copy number variants implicate cardiac function and development pathways in earthquake-induced stress cardiomyopathy. *M.A. Kennedy[1], C.J. Lacey[2], K. Doudney[3], P.G. Bridgman[4], P.M. George[3], R.T. Mulder[2], J.J. Zarifeh[5], B. Kimber[2], M. Cadzow[6], M.A. Black[6], T.R. Merriman[6], K. Lehnert[7], V. Bickley[3], J.F. Pearson[8], V. Cameron[9].* 1) Dept Pathology, University of Otago, Christchurch, Christchurch, New Zealand; 2) Department of Psychological Medicine, University of Otago, Christchurch, New Zealand; 3) Molecular Pathology Laboratory, Canterbury Health Laboratories, Canterbury District Health Board, Christchurch, New Zealand; 4) Department of Cardiology, Christchurch Hospital, Christchurch, New Zealand; 5) Psychiatric Consultation Service, Christchurch Hospital, Canterbury District Health Board, Christchurch, New Zealand; 6) Department of Biochemistry, University of Otago, Dunedin, New Zealand; 7) School of Biological Sciences, University of Auckland, Auckland, New Zealand; 8) Biostatistics and Computational Biology Unit University of Otago, Christchurch; 9) Department of Medicine, University of Otago, Dunedin, New Zealand.

The pathophysiology of stress cardiomyopathy (SCM), also known as Takotsubo cardiomyopathy, is poorly understood. SCM can occur sporadically, often in association with a stressful event, but case clusters are also well known to occur after major natural disasters. There is some evidence for familiality in this condition, and we have examined three possible models for an underlying genetic predisposition to SCM. Our primary study cohort consists of 28 women who suffered SCM as a result of two devastating earthquakes that struck the city of Christchurch, New Zealand, in 2010 and 2011. To seek possible underlying genetic factors we carried out exome analysis, Cardio-MetaboChip analysis and array comparative genomic hybridization on these subjects. The most striking finding from these analyses was the observation of an extraordinarily high rate of rare, heterogeneous copy number variants (CNV) of uncertain clinical significance (in 13/28 subjects). Several of these CNVs clearly impacted on genes of cardiac relevance including *RBFOX1*, *GPC5*, *KCNRG*, *CHODL*, and *GPBP1L1*. There is no physical overlap between the CNVs, and the genes they impact do not fall into a clear pathophysiological pathway. However, the recognition that SCM cases display a high rate of unusual CNV, and that SCM predisposition may therefore be associated with these CNVs, offers a novel perspective and a new approach by which to understand this problematic and enigmatic condition.

644T

A novel family with an unreported p.Glu768Lys *TNNI3K* mutation and atrial or junctional tachycardia. *J. Delanne[1], G. Laurent[2], J. Thevenon[1,4], C. Bonnet[3], S. Falcon-Eicher[3], C. Thauvin-Robinet[1,4], Y. Duffourd[4], T. Jouan[4], A. Baurand[1], C. Jacquot[1], N. Jean[1], L. Faivre[1].* 1) Centre hospitalier universitaire de Dijon, centre de génétique, 14 rue Paul Gaffarel BP 77908 21079 DIJON Cedex, Dijon, France; 2) Centre hospitalier universitaire de Dijon, service de cardiologie, 14 rue Paul Gaffarel BP 77908 21079 DIJON Cedex, Dijon, France; 3) Centre hospitalier universitaire de Dijon, service de cardiologie pédiatrique, 14 rue Paul Gaffarel BP 77908 21079 DIJON Cedex, Dijon, France; 4) Centre hospitalier universitaire de Dijon, laboratoire de génétique chromosomique et moléculaire, 14 rue Paul Gaffarel BP 77908 21079 DIJON Cedex, Dijon, France.

    Mutations within the Troponin I-interacting kinase (TNNI3K) gene have been reported in two families only. The mutations are associated with familial dilated cardiomyopathy and abnormal cardiac conduction disease comprising variable atrial arrhythmias. Using Whole Exome Sequencing (WES) in the two most distant relatives of a 5-patient, 3-generation family with an unrecognized abnormal cardiac rhythm phenotype, we identified a novel heterozygous TNNI3K mutation. The five patients presented with palpitations that revealed supraventricular tachycardia (SVT) qualified as congenital Junctional Ectopic Tachycardia (JET) for two of them, and undetermined SVT with or without multifocal zones in the other three. To date, none of the patients has developed dilated cardiomyopathy, but one patient has a moderately dilated aorta and another one has mitral incompetence. The mutation (chr1: g.74957901G>A; c.2302G>A; p.Glu768Lys) cosegregated with the affected family members, and was reported as probably pathogenic according to the prediction scores (SIFT: 0.002; Polyphen2: 0.983). Reverse phenotyping supported the implication of this gene in this familial phenotype. This is another example of the power of next generation sequencing and reverse phenotying in identifying the cause of unrecognized phenotypes. The description of further cases will be helpful for a better delineation of the phenotype associated with TNNI3K mutations.

645F

Whole exome sequencing in the Framingham Heart Study identifies rare variation in *HYAL2* that influences platelet aggregation. *J.D. Eicher[1,2], M.H. Chen[1,2], A. Pitsillides[2], H. Lin[3], N. Veeraraghavan[4], J.A. Brody[5], G.A. Metcalf[6], D.M. Muzny[6], R.A. Gibbs[6], E. Boerwinkle[6,7], L.A. Cupples[2,8], A.D. Johnson[1,2].* 1) Population Sciences Branch, Framingham Heart Study, NHLBI, NIH, Framingham, MA; 2) The Framingham Heart Study, Framingham, MA; 3) Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA; 4) Rady Children's Institute for Genomic Medicine, San Diego, CA; 5) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 7) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 8) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

    Inhibition of platelet reactivity is a common therapeutic strategy in the secondary prevention of cardiovascular disease. Genetic and environmental factors influence inter-individual variation in platelet reactivity. Identifying genes that contribute to platelet reactivity can reveal new biological mechanisms and possible therapeutic targets. Unfortunately, platelet reactivity data are not widely available in large cohorts. The Framingham Heart Study represents one of the largest population-based cohort studies with collected platelet reactivity assays against a wide range of agonists and doses. In this investigation, we performed whole exome sequencing to identify rare coding variation associated with platelet reactivity in the European ancestry Framingham Heart Study. We conducted both single variant and gene-based association tests against measures of platelet reactivity to collagen, ADP, and epinephrine agonists in up to 1,211 individuals. Single variant tests revealed no significant associations ($p<1.44\times10^{-7}$), though we observed a suggestive association with variants in *MRVI1*, which was previously associated with platelet reactivity (rs11042902, $p=1.95\times10^{-7}$). Using gene-based association tests of rare and low-frequency variants, we found significant associations of *HYAL2* with increased ADP-induced aggregation ($p=1.07\times10^{-7}$) and *GSTZ1* with increased epinephrine-induced aggregation ($p=1.62\times10^{-6}$). *HYAL2* also showed suggestive associations with increased epinephrine-induced aggregation ($p=2.64\times10^{-5}$). The rare variants in the *HYAL2* gene-based association included a missense variant (N357S) at a known N-glycosylation site and a nonsense variant (Q406*) that removes a glycophosphatidylinositol (GPI) anchor from the resulting protein. These variants suggest that improper membrane trafficking of HYAL2 influences platelet reactivity. We also observed suggestive associations of *AR* ($p=7.39\times10^{-6}$) and *MAPRE1* ($p=7.26\times10^{-5}$) with ADP-induced reactivity. Expression of these associated genes was confirmed in an independent platelet RNA-seq sample. Our study demonstrates that gene-based tests and other grouping strategies of rare variants are powerful approaches to detect associations in population-based analyses of complex traits not detected by single variant tests, and several new genetic factors that influence platelet reactivity.

646W

Increase burden of rare variants in known genes to cause familial thoracic aortic aneurysms and dissections in patients with early age onset of aortic dissections. *D. Guo[1], E. Regalado[1], S. LeMaire[2], S. Prakash[1], S. Wallace[1], E. Hostetler[1], R. Zhao[1], J. Smith[3], M. Bamshad[3], D. Nickerson[3], X. Liu[4], D. Milewicz[1].* 1) Department of Internal Med, Univ Texas/Houston Med Sch, Houston, TX; 2) Cardiovascular Surgery, Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX 77030, USA; 3) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 4) Department of Epidemiology & Disease Control, Univ Texas/Houston School of Public Health, Houston, TX.

Thoracic aortic dissections (TAD) are associated with a high degree of morbidity and medical expenditure in survivors. Clinical management can be initiated to prevent deaths due to aortic dissection if genetic and clinical predictors can be used to justify the risk of surgical repair of an aneurysm. Up to 20% of thoracic aortic aneurysm and dissection patients have a family history of disease (FTAAD) and fourteen FTAAD genes have been identified for this disorder. To identify genetic factors that predispose to patients with early age onset of sporadic TAD (ESTAD), we performed exome assay on DNA samples from 374 patients with ESTAD. Exclusion criteria are: age onset of dissection more than 55 years old and individuals with a known genetic syndrome (such as Marfan syndrome) or FTAAD. Exome results showed that burden of mutations (n = 32) and rare missense variants with unknown significant function (VUS) (n = 75) in known FTAAD genes was significantly increased in ESTAD patients compared to those of the ESP European American (ESP-EA) cohort (http://evs.gs.washington.edu/EVS/) with p < 0.0001, respectively. In additional, genomic copy number variants of known FTAAD genes were identified in four of the ESTAD patients. In clinical practice, it is important to evaluate the functional input of rare missense VUS in known FTAAD genes to the pathogenesis of TAAD, which are routinely identified by genetic diagnosis lab. To investigate these variants, we performed bioinformatics analyses to evaluate the input of these variants to protein function. The rare missense VUS that were identified in proteins (n = 6), such as ACTA2 and LOX, which barely have missense variants in the general population or that were located at the critical domains (n = 6), such as kinase domain of TGFBR1 or TGFBR2 kinase or calmodulin-binding domain of MYLK, which were the locations of majority known mutations for FTAAD, have higher damaging or conservation scores than rare missense variants (n = 351) identified in ESP-EA cohort and have similar scores compared to missense mutations in known FTAAD genes reported (n = 1155). In summary, increased burden of mutations or rare missense VUS in known FTAAD was increased in ESTAD cohort than that of ESP-EA cohort. Combination of bioinformatic analyses of rare missense VUS and their domain locations is potentially to be used to predict functional input of these variants to the pathogenesis of TAAD.

647T

Whole exome sequencing in 161 sudden infant death syndrome (SIDS) cases with a focus on cardiovascular and metabolic genetic diseases. *J. Neubauer[1], M.R. Lecca[2], G. Russo[2], C. Bartsch[3], A. Medeiros-Domingo[4], W. Berger[5,6,7], C. Haas[1].* 1) Institute of Forensic Medicine, University of Zurich, Zürich, Switzerland; 2) Functional Genomics Center Zurich (FGCZ), University of Zurich / ETH Zurich, Zurich, Switzerland; 3) University of Zurich, Zurich, Switzerland; 4) Department of Cardiology, Inselspital, University Hospital Bern, Switzerland; 5) Institute of Medical Molecular Genetics, University of Zurich, Schlieren, Switzerland; 6) Center for Integrative Human Physiology (ZIHP), University of Zurich, Zurich, Switzerland; 7) Neuroscience Center (ZNZ), University and ETH Zurich, Zurich, Switzerland.

Sudden infant death syndrome (SIDS) is described as the sudden and unexplained death of an apparently healthy infant younger than one year of age. The cause of death remains unexplained after a complete autopsy, review of the circumstances of death and the clinical history. Although the incidence rate of SIDS drastically decreased in the last years, SIDS is still one of the leading causes of postneonatal infant death in developed countries. The occurrence of SIDS is described by a triple risk model involving (1) a critical developmental period, (2) exogenous stress factors, and (3) a vulnerable infant. Genetic studies in SIDS cohorts indicate that 10-15% of SIDS cases might be explained by genetic cardiac diseases, not detectable during conventional forensic autopsy examination. Post-mortem genetic testing by using next generation sequencing (NGS) approaches represents an efficient and rapid technology to further investigate unexplained death cases and might help to elucidate pathogenic mechanisms in cases without a conclusive cause of death. Here, we report the results of a whole exome-sequencing (WES) study in 161 European SIDS infants with a primary focus on 193 genes associated with cardiovascular or metabolic diseases. SureSelect (Human All Exon V5 + UTRs, Agilent) exome capture was followed by sequencing on the Illumina HiSeq2500 platform. The average on-target coverage was 90.2% at ≥ 20x. Pathogenicity of variants was assigned according to a scoring scheme based on the type of variant, *in silico* protein predictions, and minor allele frequencies in European control populations. Potentially causative variants were detected in 26% of the SIDS cases. The majority of infants had variants with likely functional effects in genes associated with channelopathies (11%), followed by cardiomyopathies (9%), and metabolic diseases (1%). Although lethal arrhythmia represents the most plausible and likely cause of death, the majority of SIDS cases still remains elusive and might be explained by a multifactorial event triggered by a combination of different genetic and environmental risk factors. To our knowledge, this is the first WES study in a large SIDS cohort. Since WES is not substantially more expensive than a targeted NGS approach, it represents an unbiased screening of the exome, which would help to investigate different pathogenic mechanisms within the genetically heterogeneous SIDS cohort and to identify new candidate genes in sudden infant death.

648F

Familiality of statin intolerance and characteristics of members of high risk statin intolerance pedigrees. *S. Knight[1,2], H.T. May[1], J. Carlquist[1,2], J.B. Muhlestein[1,2], K.U. Knowlton[1], B.D. Horne[1,2].* 1) Intermountain Medical Center, Murray, UT; 2) University of Utah, Salt Lake City, UT.

Background: With the introduction of new PCSK-9 inhibitor drugs, a new option is available for patients with statin intolerance (SI) to lower their low-density lipoprotein cholesterol (LDL-C). However, PCSK-9 drug costs require careful targeting of those who will benefit the most. Objectives: To determine the familiality of SI, identify pedigrees with a high rate of SI, and determine differences in clinical and demographic characteristics for SI patients with and without a familial component. Methods: Intermountain Healthcare patients with SI (determined using medication allergy/intolerance electronic report), and pedigree records in the Intermountain Genealogy Registry (IGR) were studied (n=2556). Familiality of SI was tested using the genealogy index of familiality (GIF). GIF is the average pairwise kinship coefficient (multiplied by 100,000). Significance of GIF was assessed using empirical distributions from 1000 age and sex matched IGR control pts without SI. To be high risk for SI, pedigrees had to have an SI rate >5 times all IGR pedigrees. Sex, age, race, LDL-C, and comorbidities were compared between high- and low-SI pedigrees. Results: The GIF for SI was 0.71 (controls GIF: 0.60; GIF ratio=1.17, p=0.065). A total of 23 high-risk pedigrees were identified; the average rate of SI was 2.4% (IQR: 2.2% to 3.3%) with all having 4-6 SI cases. Median familial relative risk of SI was 6.2 (IQR: 5.8, 7.8). Patient characteristics are not significant different after correction for multiple comparisons (Table). Conclusions: A trend for evidence of a familial component to SI was found and several high risk pedigrees were identified. Clinical and demographic characteristics were not different. While these factors would not be effective to use to identify SI pts with a familial etiology, further identification of SI indicators is needed.

	SI pts in high risk pedigree (n=96)		SI pts not in high risk pedigrees (n=2460)		p-value
Male	56	58.30%	1207	49.10%	0.08
Age median (IQR)	73.5	(64.5, 81)	73	(65, 81)	0.96
Caucasian	87	90.60%	2333	94.80%	0.07
Average LDL-C, median (IQR)	113.8	(93, 143.6)	118	(97.8, 140.2)	0.33
COPD	45	46.90%	1213	49.30%	0.64
Connective Tissue Disorder	10	10.40%	260	10.60%	0.96
Myocardial Infarction	24	25.00%	521	21.20%	0.37
Stroke	18	18.80%	716	29.10%	0.028
Heart Failure	24	25.00%	690	28.00%	0.51
Dementia	2	2.10%	107	4.30%	0.28
Liver Disease	15	15.60%	514	20.90%	0.21
Peripheral artery disease	21	21.90%	667	27.10%	0.26
Renal Disease	23	24.00%	599	24.30%	0.93
Diabetes	45	46.90%	1120	45.50%	0.8
Cancer	23	24.00%	544	22.10%	0.67
Hypothyroidism	23	24.00%	749	30.40%	0.17

649W

Genome-wide gene-environment interaction analysis identifies genetic signals associated with cardiometabolic phenotypes and physical activity in Hispanic Americans: The Insulin Resistance Atherosclerosis Family Study (IRASFS). *C. Gao[1,2,3], J. Keaton[1,2,4], F. Hsu[5], H. Okut[2], J. Rotter[6], Y. Chen[6], K. Taylor[6], C. Langefeld[3,5], L. Wagenknecht[7], D. Bowden[2,4,8], N. Palmer[2,4,8].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 4) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Biostatistical Sciences; Wake Forest School of Medicine, Winston-Salem, NC; 6) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA; 7) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 8) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC.

Despite the recent success of genome-wide association studies (GWAS) in identifying genetic variants regulating metabolic phenotypes, a substantial proportion of the heritability remains unexplained. Gene-environment interactions are a potential tool to explain the missing heritability. In this study, up to 822 Hispanic Americans from IRASFS were chosen for a genome-wide interaction analysis with 12 cardiometabolic phenotypes using physical activity score acquired by questionnaire. Interaction and joint (main+interaction effects) analyses were performed using SOLAR. The top signal was rs183130 (MAF=28%), located upstream of the cholesteryl ester transfer protein gene (*CETP*), with high density lipoprotein (HDL). This association was characteristic of GWAS results and driven by a main effect (P_{Main}=3.04E-10) with no interaction (P_{Int}=0.16). Among novel signals, SNP rs3735602 (MAF=45%) located downstream of the dihydrolipoamide dehydrogenase gene (*DLD*) was strongly associated with triglyceride levels (TG) and driven by an interaction effect (P_{Int}=1.08E-07) while no main effect was observed (P_{Main}=0.41). This suggests a complex function of *DLD*: the minor allele (C) has a protective effect in carriers with low physical activity with the opposite effect in the high physical activity group, i.e. compared to non-carriers, the homozygous C allele carriers have 35% less and 45% more TG in low and high physical activity groups, respectively. Interestingly, DLD performs mechanistically distinct functions, i.e. the homodimeric form functions as a dehydrogenase regulating energy metabolism and the monomer as a protease. These distinct functions may be responsible for its complex effects. Additionally, intronic SNP rs57556257 (MAF=9%) in cysteine rich transmembrane BMP regulator 1 gene (*CRIM1*) was strongly associated with disposition index (DI), a measure of crosstalk modulating glucose homeostasis pathways, in the joint analysis (P_{Joint}=5.95E-08). The signal exhibited main and interaction effects (P_{Main}=2.44E-07, P_{Int}=1.01E-02), suggesting the genetic effect is partially dependent on physical activity. *CRIM1* encodes a transmembrane protein containing an insulin-like growth factor-binding domain. In conclusion, genome-wide interaction analysis of physical activity indicates that the inclusion of environmental factors allows for the identification of additional genetic signals contributing to variation in cardiometabolic phenotypes.

650T

A three-way interaction among maternal and fetal variants contributing to congenital heart defects. *M. Li[1], J. Li[2], C. Wei[3], Q. Lu[4], X. Tang[2], S. Macleod[2], C. Hobbs[2].* 1) Department of Epidemiology and Biostatistics, School of Public Health, Indiana University at Bloomington, Bloomington, IN; 2) Department of Pediatrics, University of Arkansas for Medical Sciences; 3) Department of Epidemiology and Biostatistics, University of North Texas Health Science Center; 4) Department of Epidemiology and Biostatistics, Michigan State University.

Congenital heart defects (CHDs) develop through a complex interplay between genetic variants, epigenetic modifications, and maternal environmental exposures. Genetic studies of CHDs have commonly tested single genetic variants for association with CHDs. Less attention has been given to complex gene-by-gene and gene-by-environment interactions. In this study, we applied a recently developed likelihood-ratio Mann-Whitney (LRMW) method to detect joint actions among maternal variants, fetal variants, and maternal environmental exposures, allowing for high-order statistical interactions. All subjects are participants from the National Birth Defect Prevention Study, including 623 mother-offspring pairs with CHD affected pregnancies and 875 mother-offspring pairs with unaffected pregnancies. Each individual has 872 single nucleotide polymorphisms encoding for critical enzymes in the homocysteine, folate, and trans sulfuration pathways. By using the LRMW method, three variants (fetal rs625879, maternal rs2169650, and maternal rs8177441) were identified with a joint association to CHD risk (nominal P-value = 1.13e-07). These three variants are located within genes BHMT2, GSTP1, and GPX3, respectively. Further examination indicated that maternal SNP rs2169650 may interact with both fetal SNP rs625879 and maternal SNP rs8177441. Our findings suggest that the risk of CHD may be influenced by both the intragenerational interaction within the maternal genome and the intergenerational interaction between maternal and fetal genomes.

651F

Cardiovascular comorbidity influence genetic association studies results. *O.A. Makeeva[1,2], A.A. Sleptsov[1], E.V. Kulish[1], V.V. Markova[1], O.L. Barbarash[2], A.M. Mazur[3], E.B. Prokhorchuk[3], N.N. Chekanov[3], V.A. Stepanov[1], V.P. Puzyrev[1,4].* 1) Institute of Medical Genetics, Tomsk, Russian Federation; 2) Institute of Complex Issues of Cardiovascular Problems, Kemerovo, Russia; 3) Genoanalitika, Moscow, Russia; 4) Siberian State Medical University, Tomsk, Russia.

Comorbidity or a combination of several diseases in one individual is a common phenomenon. However, the genetic background for the non–random disease combinations is not fully understood. More data is needed to investigate the genetic profile of patients burdened with several diseases (polypathia, disease conglomerates) and its comparison with the profile of patients with single diseases. Our association study featured three groups of patients with various combinations of cardiovascular disorders and a control group of relatively healthy individuals. Patients were selected as follows: presence of only one disease, ischemic heart disease (IHD); a combination of two diseases, IHD and arterial hypertension (AH); and a combination of several diseases, including IHD, AH, type 2 diabetes mellitus (T2DM), and hypercholesterolemia. Genotyping was performed using the "My Gene" genomic service (www.i–gene.ru). About 1,400 polymorphic genetic variants and their associations with the studied phenotypes were analyzed. A total of 14 polymorphic variants were associated with the phenotype "IHD only," including those in the genes *APOB, CD226, NKX2–5, TLR2, DPP6, KLRB1, VDR, SCARB1, NEDD4L,* and *SREBF2*, as well as several intragenic variants. A total of 13 markers were associated with the "IHD and AH" phenotype, including variants in the *BTNL2, EGFR, CNTNAP2, SCARB1,* and *HNF1A* genes, and 6 intragenic polymorphisms. 14 variants were associated with a combination of several cardiovascular diseases (CVD), including in the *TAS2R38, SEZ6L, APOA2, KLF7, CETP, ITGA4, RAD54B, LDLR,* and *MTAP* genes, along with three intragenic polymorphisms. The *SCARB1* rs4765623 was common marker for the "IHD only" and "IHD and AH" phenotypes; rs663048 in SEZ6L and intragenic rs6501455 were common for the "IHD and AH" phenotype and a combination of several diseases (syntropy); there were no common genetic markers for the "syntropy" and "IHD only" phenotypes. Classificatory analysis of the relationships between the associated genes and metabolic pathways revealed that lipid–metabolizing genes are involved in the development of all three studied CVD phenotypes whereas immunity response genes are specific to the "IHD only". The study demonstrates that comorbidity presents an additional challenge for association studies of common diseases: the genetic profile of combined forms of pathology can be different from those for isolated forms of a disease.

652W

Utilizing the strength of pedigree-based study design to identify novel tissue plasminogen activator loci. *N. Zwingerman[1], V. Truong[1], M. Wilson[2], P. Wells[3], P.E. Morange[4], D.A. Trégouët[5], F. Gagnon[1]. 1) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 2) Genetics and Genome Biology Program, Peter Gilgan Centre for Research and Learning, Toronto, Canada; 3) Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 4) INSERM, UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, Marseille, France; 5) INSERM, UMR_S 1166, Team Genomics and Pathophysiology of Cardiovascular Diseases, Paris, France.*

Tissue Plasminogen Activator (tPA) is a serine protease that mediates the conversion of plasminogen to plasmin, the major enzyme responsible for endogenous fibrinolysis. In some populations, elevated tPA levels are associated with thrombotic outcomes. tPA plasma levels are highly heritable, with estimates as high as 67%. A mega-GWAS implicated 3 genes involved in tPA regulation that collectively explain only ~1% of tPA variance. Rare genetic variants with large effect size could underlie the unaccounted tPA variation. A pedigree-based study in a founder population is an efficient design to test for co-segregation of putative rare variants. To identify rare genetic variants associated with tPA plasma levels. The study sample consists of 4 extended French-Canadian pedigrees ascertained on a single proband with venous thromboembolism, with complete data of 178 individuals. We used an iterative Bayesian oligogenic joint linkage and segregation analysis to conduct a genome-wide linkage scan of microsatellite markers with tPA levels, followed by a variance-component approach to validate findings in a hypothesis-testing framework. Then an association analysis, using a linear mixed regression model adjusted for relatedness and covariates was used, to test for association between single nucleotide polymorphisms (SNPs) and tPA plasma levels in the identified linkage region. The significance threshold was determined by estimating the effective number of tests, accounting for the LD structure in the region. Both linkage analysis approaches identified a genome-wide significant signal on chromosome 11q14 region (LOD score=3.51; $\log_{10}$(BF)=2.56, p=0.001) and 15q25 region (LOD score=3.16; $\log_{10}$(BF)= 1.61, p=0.0012). There were 3,286 and 2,170 genotyped SNPs in the linkage regions common to both approaches on chromosome 11 and 15 respectively, corresponding to a significance threshold of $2.89e^{-5}$ and $4.17e^{-5}$. Nine and two genotyped SNPs met the significance thresholds respectively. Imputed SNP data in the regions were analyzed for refinement, which identified the strongest associations with rare variants in this region (top imputed SNP: $p=3.48e^{-7}$, MAF=0.11 vs to MAF=0.039 in population). Linkage analysis followed by genetic association testing in pedigrees identified a novel regions of interest for tPA level variation, providing new insights on tPA determinants.

653T

Protein-altering and regulatory genetic variants near *GATA4* implicated in bicuspid aortic valve. *C.J. Willer[1], W. Zhou[1], J. Jiao[1], J.B. Nielsen[1], M. Mathis[1], M. Heydarpour[2], G. Lettre[3,4], L. Folkersen[5,6], S. Prakash[7], C. Schurmann[8], M. Lin[1], M. Othman[1], J. Kitzman[1], A. Boyle[1], S. Ganesh[1], G. Abecasis[1], K. Eagle[1], R.J.F. Loos[8], P. Eriksson[5], J-C. Tardif[3,4], C.M. Brummett[1], D. Milewicz[7], S. Body[2], B. Yang[1], BAVCon Consortium. 1) University of Michigan, Ann Arbor, MI; 2) Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Montreal Heart Institute, Montreal, Quebec, Canada; 4) Université de Montréal, Montreal, Quebec, Canada; 5) Karolinska University hospital Solna, Karolinska Institutet, Stockholm, Sweden; 6) Technical University of Denmark, Copenhagen, Denmark; 7) University of Texas Health Science Center at Houston, Houston, TX; 8) 12.The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY.*

Individuals with bicuspid aortic valve (BAV), a heritable congenital heart defect, have substantially increased risk of complications of the valve and aorta. To identify the underlying causes of BAV, we performed a genome-wide association scan of 466 non-syndromic BAV cases and 4,660 age, sex, and ethnicity-matched controls, with replication in an additional 1,360 cases and 8,205 controls. We genotyped samples with the Illumina Human CoreExome array with 498,075 markers. We identified association with a non-coding variant (rs6601627, OR = 1.90, 95% CI 1.62 – 2.23, $p=3x10^{-15}$) 151kb from the cardiac-specific transcription factor, *GATA4*, and near-significance for protein-altering p.Ser377Gly in *GATA4* (rs3729856, OR = 1.31, 95% CI 1.19 – 1.45, $p=1x10^{-7}$). These variants were determined to be independent by conditional analysis. This locus allowed investigation of both protein-coding and non-coding regulation of an important cardiac transcription factor and its involvement in a congenital heart defect. We identified chromatin interaction loops in K562, using ChIA-PET data and Hi-C data, between associated variants and a region in the large 40kb intron of GATA4. To investigate the role of GATA4 in aortic valve development, *GATA4* was interrupted by CRISPR Cas9 and single guide RNA (sgRNA) in induced pluripotent stem cells (iPSCs) from donors with normal tri-leaflet aortic valve. The disruption of *GATA4* significantly impaired the transition from endothelial cells into mesenchymal cells, a critical step in heart valve development. In summary, we report the first genome-wide significant association with BAV which identified non-coding variants near GATA4, and a missense variant within *GATA4*. We demonstrate that disruption of *GATA4* in iPSC impairs the differentiation from endothelial to mesenchymal cells, confirming that at this locus *GATA4* is likely the functional gene increasing risk of BAV.

654F

Sequence data processing and analysis of 18,000 human genomes in the NHLBI TOPMed sequencing program. *T. Blackwell[1], G. Abecasis[1], K. Barnes[2], J. Blangero[3], E. Burchard[4], A. Correa[5], L.A. Cupples[16], J.E. Curran[3], P.T. Ellinor[6], S. Gabriel[7], S. Germer[8], R. Hernandez[4], C. Jaquish[9], H.M. Kang[1], S. Kathiresan[7], C.C. Laurie[10], X. Lin[11], R.A. Mathias[12], S.T. McGarvey[13], B.D. Mitchell[14], D. Nickerson[15], J.E. O'Connell[14], G. Papanicolaou[9], V. Ramachandran[16], S. Redline[17], K. Rice[10], E. Silverman[17], D.E. Weeks[18], S.T. Weiss[17], J.G. Wilson[5] on behalf of the NHLBI TOPMed Program.* 1) TOPMed Informatics Research Center, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) University of Colorado, Denver; 3) University of Texas Rio Grande Valley School of Medicine; 4) University of California, San Francisco; 5) Jackson Heart Study, University of Mississippi Medical Center; 6) Massachusetts General Hospital; 7) Broad Institute; 8) New York Genome Center; 9) National Heart, Lung and Blood Institute, NIH; 10) Department of Biostatistics, University of Washington, Seattle; 11) Department of Biostatistics, Harvard School of Public Health; 12) Johns Hopkins University Medical School; 13) Brown University, Providence RI; 14) University of Maryland Medical School; 15) Northwest Genomics Center, University of Washington, Seattle; 16) Framingham Heart Study, Boston University; 17) Brigham and Women's Hospital, Boston; 18) University of Pittsburgh.

The National Heart Lung and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) initiative aims to stimulate discovery of the fundamental mechanisms that underlie heart, lung, blood and sleep disorders by combining whole genome sequence and other omics data with existing high quality phenotype information. The first two years of the project will generate deep whole genome sequence data for >60,000 diverse samples (including a majority of samples with non-European ancestries that are traditionally under-represented in genomic research) and represent a collaboration between diverse groups of scientists, 6 genome sequencing centers, a project data coordination center, an informatics resource center and staff at the National Institutes of Health and the National Center for Biotechnology Information (NCBI). Here we describe efforts to integrate sequence data across participating studies and sequencing centers. This is being carried out at the project's Informatics Research Center (IRC). Currently >18,500 samples from 10 year 1 studies have been sequenced to an average sequencing depth of 37.8x (after duplicate removal), resulting in 98.9% of the genome covered at depth 10 or greater. An initial data freeze with 10,597 sequenced individuals yields 142 million SNPs and 10.2 million short indels, of which 44.3% and 38.7% are singletons, respectively. Each individual carries a non-reference allele at between 3.5 and 4.5 million sites, depending on ethnicity. Deep sequencing, PCR-free sample prep, and harmonized data processing allow us to achieve <0.2% difference in SNP variant calls when the same sample is sequenced at different sites. Although more differences are observed for indel polymorphisms, these are concentrated in low complexity regions of sequence, such as short tandem repeats. Compact file formats and efficient processing allow analysis at this scale. A primary goal of the project is to allow investigators exploring diverse study designs and phenotypes to benefit from access to high-quality sequence data and analysis. There are also many opportunities to be had from combined analysis of the sequence data to be generated by the project, including a diverse catalog of genetic variation, a potential resource of controls for genetic association studies, and an imputation reference panel. We discuss some of these opportunities and associated logistical and technical challenges, as well as opportunities for any scientist to access project results.

655W

GWAS to 30X genomes: Evolution of sequencing in the ARIC cohort to reveal the genetic architecture of complex traits. *G.A. Metcalf[1], A. Morrison[2], B. Yu[2], X. Liu[2], S. White[1], H. Doddapaneni[1], J. Hu[1], W. Salerno[1], D.M. Muzny[1], R.A. Gibbs[1], E. Boerwinkle[1,2].* 1) Baylor College of Medicine, Houston, TX; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

The Atherosclerosis Risk in the Community (ARIC) study is a flagship project seeking to identify susceptibility genes underlying well-replicated GWAS findings for heart, lung, and blood diseases and their risk factors in combined analyses with the larger Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Early efforts involved DNA sequencing via a multiplexed DNA capture approach targeting a 2.2 Mb region of the genome. While informative, analyses were limited to the loci selected from the GWAS findings and thus biased towards common variants. As sequencing platforms improved, more comprehensive applications of whole exome sequencing (WES) and low pass (6X) genomes became affordable. To date, the HGSC has accumulated sequences from 15K WES 5K low pass whole genome samples spanning 5 longitudinal cohort studies with extensive phenotypic and clinical data. To promote novel discoveries for rare coding variants, we jointly analyzed the ARIC and CHARGE data for multiple biomedical traits. Single variant tests of common variants and gene-based tests of rare variants were evaluated for each trait and ancestry separately, followed by multiethnic meta-analyses. Functional annotation of SNVs, genes and regulatory elements was integrated to increase power to detect true phenotype-genotype association using our custom annotation pipeline, WGSA. These data have fueled considerable analytical work in collaboration with the CHARGE consortium, forming the foundation for discovery of novel genes/loci influencing a number of chronic disease risk factors. Recently we conducted a fixed effects meta-analysis of WES data in 14,770 participants for 13 hematology traits in European Americans (EA) and African Americans (AA). Discovery analyses revealed 5 novel, significant gene-based test results and 3 novel, significant single variant associations. For example, a low frequency (MAF=0.009) novel single synonymous variant in GFI1B (rs150813342) was associated with platelet count in EAs (P=4.72x10-8). Generation and analysis of 30X WGS data is underway using the Illumina HiSeq X in an effort to identify coding and non-coding regions in relation to complex quantitative traits, specifically those regions influencing the human serum metabolome and its association with CVD. By focusing on proximal measures of physiologic processes we will optimize the size of a gene's effect relative to corresponding risk factor level or disease endpoint.

656T

Common variants of *ZFPM2* are associated with non-syndromic bicuspid aortic valve. *S. Prakash[1], M. Heydarpour[2], D. Guo[1], D. Milewicz[1], S. Body[2], BAVCon Investigators.* 1) Internal Medicine/Medical Genetics, University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Background: Bicuspid aortic valve (BAV) is the most common congenital heart defect, but the genetic causes are largely unknown. Approximately one-third of BAV patients develop ascending aortic aneurysms and 1-2% develop acute aortic dissections (TAAD). We hypothesized that common autosomal variants predispose to BAV and TAAD and performed a GWAS of BAV cases who were referred for surgical repair to Brigham and Women's Hospital (average age 55 years, 26% female, 44% TAAD, 46% aortic stenosis). Methods: 480 patients of European ancestry with BAV were genotyped using Illumina Omni-2.5 arrays and compared separately with two sets of unselected dbGAP genotypes: the Health and Retirement Study (HRS, phs000428, n=12,845) and Smoking Cessation (SmokCes, phs000404, n=1,525) cohorts. Association tests were carried out using logistic regression, adjusting for gender and principal components. Candidate regions were imputed against 1000 Genomes Phase 3 using shapeit and impute2 and meta-analyzed using PLINK. Results: After removal of cryptically related individuals and population outliers, a total of 447 BWH cases, 9,428 HRS controls and 935 SmokCes controls were available for analysis. The maximum genomic inflation factor was 1.03. The strongest association signals for both the BWH-HRS and the BWH-SmokCes comparisons were observed on chromosomes 7 and 8 and did not overlap with previously reported loci for BAV. A total of 12 genotyped SNPs in the *ZFPM2* gene in 8q23.1 were positively associated with BAV (OR=1.46), but not TAAD, with a minimum genotyped *P*-value of 1×10^{-7} (rs16873287) and minimum imputed *P*-value of 9×10^{-8} (rs12676726). The same haplotype block was associated with BAV in a meta-analysis combining two other cohorts with BAV and TAAD or Turner syndrome ($P=8 \times 10^{-5}$, OR=1.38). *ZFPM2* encodes a transcription factor that interacts with Gata4, was implicated in cardiac outflow tract development and is mutated or affected by rare CNVs in patients with tetralogy of Fallot. In a separate cohort of 765 elderly patients who were ascertained for sporadic TAAD (average age 63, 15% BAV), neither top SNP was significantly associated with BAV (OR=1.02), but one rare deletion of *ZFPM2* was identified in a subject with BAV. Additional replication studies with cohorts that were primarily ascertained for BAV are ongoing. Conclusion: These observations support previous data that common variants affecting cardiac developmental genes may also contribute to BAV. .

657F

A genome-wide association study of >27,000 cases and controls for coronary artery disease in a multiethnic cohort. *M. Verbanck[1], V.W. Setiawan[2], N. Zubair[3], A.C. Morrison[4], P.S. de Vries[4], L. Hindorff[5], C. Haiman[2], R. James[5], L. Le Marchand[6], U. Peters[3], C.L. Wassel[7], K.E. North[8], R.J.F. Loos[1], T. Matise[9], C. Avery[8], C. Kooperberg[3], R. Do[1] on behalf of the PAGE Study.* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) University of Texas Health Science Center, Houston, TX; 5) National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD; 6) University of Hawaii Cancer Center, Honolulu, HI; 7) University of Vermont College of Medicine, Burlington, VT; 8) University of North Carolina, Chapel Hill, NC; 9) Rutgers University, Piscataway, NJ.

Coronary artery disease (CAD) is one of the leading causes of death worldwide. Large-scale genetic studies of CAD have successfully identified common genetic variants associated with CAD. However, these genetic studies have been performed mostly in European ancestry populations. To test whether genetic variants are associated with complex traits and disease in non-European populations, the National Institute of Health's Population Architecture using Genomics and Epidemiology (PAGE) consortium has generated a meta-cohort comprising approximately 50,000 U.S. participants of African, Hispanic, Asian, Native Hawaiian and Native American ancestry. These individuals were genotyped using the 1.7M SNP Multi-Ethnic Genotyping Array (MEGA), a custom-built array designed specifically to include genetic variation from diverse ancestry populations. From the meta-cohort, we identified 5,432 CAD cases and 21,911 controls. Cases were defined by myocardial infarction (MI) and/or revascularization, in addition to CAD-related death. Controls were defined as not having any evidence of CAD/MI. We performed an initial genome-wide association analysis, using logistic regression under an additive model, after adjustment for age, sex and the first 10 principal components of genetic ancestry, on autosomal variants with a minor allele frequency ≥1%. We identified two sub genome-wide significant SNPs in two loci (CYP46A1; CAMK1D; $P<6 \times 10^{-7}$). Both SNPs are rare in European populations (MAF=0.002 and MAF=0.01 respectively) and are in regions that do not have genome-wide significant SNPs in the CARDIOGRAM+C4D study. Future plans include replication of the top associated SNPs in additional samples, refining the association analysis by examining ancestry-specific effects and fine-mapping using imputation data from the Haplotype Reference Consortium panel. We will present findings of these plans, and describe advantages and challenges with analyzing genetic data in multi-ethnic populations.

658W

Pleiotropic effects of Coxib, a phenotype rescue study *in-silico* **and** *in-vivo***.** *K. Vishnolia[1,2], I. Braenne[1,2], H. Schunkert[3,4], J. Erdmann[1,2].* 1) IIEG, University of Luebeck, Luebeck, Germany; 2) DZHK (German Research Center for Cardiovascular Research), Partner site Hamburg/Luebeck/Kiel, Germany; 3) Deutsches Herzzentrum Muenchen, Technische Universitaet Muenchen, Muenchen, Germany; 4) DZHK (German Research Center for Cardiovascular Research), Partnersite Muenchen, Germany.

   Aim: Cyclooxygenase 2 inhibitors (Coxibs) are used in the treatment of chronic pain relief and are known to cause severe side-effects such as Myocardial infarction (MI) and Hypertension. Due to the severe side-effects, coxibs are under scrutiny and several Coxibs have been withdrawn from the market. Coxibs are reported to interact with several gene and gene products either down-stream the prostaglandin synthesis pathway or through other pathways. It is likely, that some of these drug genes interactions influence the increased risk of MI or hypertension. **Results:** In an *in-silico* analysis, we studied the genes representing molecular targets of coxibs to search for genomic signals for coronary artery disease (CAD) risk. We screened for association with CAD in 84.813 cases and 202.543 controls. With this approach, we identified four candidate genes, VEGFA, MMP9, CACNA1E and BCAR1. To examine the role of these genes with regard to the side effects that occur under coxib treatment, we next performed *in-vivo* studies in Zebrafish larvae. The four human genes, VEGFA, MMP9, CACNA1E and BCAR1 are conserved genes and found as homologue genes in Zebrafish genome. Four days post fertilized Zebrafish larvae when treated with coxibs, we identified significantly increased blood pressure compared to control treatment. In parallel, the mRNA expression levels of all four genes was observed to be significantly lower after coxib treatment, validating our *in-silico* prediction. To verify whether the -due to Coxibs- altered gene expression influenced the blood pressure we next repeated the experiment but this time supplementing coxibs with a second drug. **Conclusion:** The hypothesis is that a second drug with an opposite effect on one or all four genes will restore the normal levels, and that we can rescue the phenotype accordingly. This study gives new insights in the mechanisms underlying the severe side effects of coxibs and how we can explore the potential clinical relevance of off-target drug effects.

659T

Antihypertensive drug response in *APOL1* **risk allele carriers and non-carriers in African-Americans using electronic medical record data.** *G. Galarneau[1], G.N. Nadkarni[1], E.P. Bottinger[1,2], R.J.F. Loos[1], R. Do[1].* 1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Berlin Institute of Health, Berlin, Germany.

   Genetic variants forming the *APOL1* risk allele have been associated with kidney disease in individuals of African ancestry (Genovese, G. *et al*. *Science*, 2010). We recently showed that the *APOL1* risk allele was also associated with increased systolic blood pressure (SBP) in young African-Americans (Galarneau et al. Abstract #514 ASHG 2015). Since studies have shown that the response to blood pressure-lowering therapies depends on genotype, drug target and mechanism of action, we hypothesized that *APOL1* carrier status might exhibit differences in the SBP response to antihypertensive classes. We extracted data from the electronic medical record (EMR) of 5,204 African-Americans from the Mount Sinai Bio*Me* Biobank and identified a total of 3,035 events of first prescription of an antihypertensive class of which we could assess class-specific drug response (no more than one class being newly prescribed, but concurrent usage of other antihypertensive classes was permitted to increase power). We used for replication an independent set of 1,623 African-Americans from the Bio*Me* Biobank. We defined the pre-prescription SBP as the mean SBP during the three months prior to prescription and the post-treatment SBP as the mean SBP during the three months after the prescription. The outcome variable was defined as the decrease in outpatient SBP following the prescription of a specific antihypertensive class. Overall, this decrease ranged from 1.4-10.2 mmHg in the different antihypertensive classes. We tested the association of the *APOL1* risk allele with SBP antihypertensive response by performing a linear regression with sex, age, body-mass index and total number of antihypertensive classes prescribed as covariates. Three of the ten antihypertensive classes studied were statistically significant (p<0.05) (aldosterone antagonists, beta blockers and loop diuretics). Among these three, the class of aldosterone antagonists replicated in the replication cohort (p=0.009). Carriers of the *APOL1* risk allele had a significantly better response (12.5 vs 0.9 mmHg decrease in SBP) to aldosterone antagonists compared to non-carriers. These results can potentially lead to new paradigms of blood pressure-lowering treatment in *APOL1* carriers.

660F

A phenome-wide scan detects pleotropic effect of LY96 on new-onset atrial fibrillation and diabetes after coronary artery bypass grafting surgery. *Y.J. Li[1,2], R. Huang[2], W. Qi[1,2], M.V. Podgoreanu[3], M.D. Kertai[3].* 1) Department of Biostatistics and Bioinformatics, Duke Univ Med Ctr, Durham, NC; 2) Duke Molecular Physiology Institute, Duke Univ Med Ctr, Durham, NC; 3) Department of Anesthesiology, Duke Univ Med Ctr, Durham, NC.

Using a genome-wide approach, we previously identified genetic variants in the Lymphocyte antigen 96 (*LY96*) gene that were associated with a reduced risk for postoperative atrial fibrillation (POAF) in patients who underwent coronary artery bypass graft (CABG) surgery. The encoded protein for the *LY96* gene is the toll-like receptor 4 coreceptor LY96. Functional activation of toll-like receptor 4 has been shown to be associated with increased inflammation and oxidative stress. As POAF is one of the most common complications after CABG surgery, occurring 25% to 40%, we sought to further understand the pleiotropic effects of LY96 gene related to health condition of CABG patients in the Pegassus cohort by taking the advantage of the rich electronic medical record data available in these patients. A total of 863 patients with available postoperative ICD-9 codes and LY96 genotype data were investigated. We mapped the ICD-9 codes to PheWAS codes for the disease classification (https://phewas.mc.vanderbilt.edu/). To ensure reasonable sample sizes of cases and controls, we limited our analyses on PheWAS codes with incidence rates between 18% and 80%. Single nucleotide polymorphisms (SNPs) in LY96 analyzed were from Illumina Human610-Quad BeadChip. Logistic regression model was used to test the association between SNP and each phenotype with age and gender adjusted. The phenotypes meeting $p < 0.05$ were included in the multivariate-genotype analysis using TATES to determine the pleotropic effect of the LY96 variants across the targeted phenotypes. A total of 2621 postoperative ICD-9 codes available after the surgery date were surveyed, which mapped to 1259 PheWAS codes. The incidence rates of PheWAS codes range from 0.00113 to 0.9887, where 18 PheWAS codes met our filtering criteria for further analyses. Five SNPs in LY96 presented consistent association with type 2 diabetes mellitus (T2DM). The incidence rate of T2DM was 40.6% in our dataset, but only 12% of patients have both POAF and T2DM. Particularly, our previous reported POAF SNP rs10504554 (OR=0.48, $p=2.97 \times 10^{-5}$) is associated with T2DM (OR=1.36, p=0.024) and tobacco use disorder (OR=1.37, p=0.029). Multivariate-genotype association analysis across these three phenotypes remains significant ($p=8.91 \times 10^{-5}$). By adopting the PheWAS approach, we provide new evidence of the role of LY96 variants on other comorbidities in CABG patients.

661W

Distribution of *CAPN10* gene haplotypes according to clinical and biochemical parameters related to hypertension. *A.G. Castro-Martínez[1], J. Sánchez-Corona[1], A.G. García-Zapién[1], G. Ramírez-López[2], S. Fonseca-Reyes[3], M.C. Morán-Moguel[1], S.E. Flores-Martínez[1].* 1) División de Medicina Molecular, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 2) Unidad de Investigación Epidemiológica y en Servicios de Salud del Adolescente, Instituto Mexicano del Seguro Social, Tonalá, Jalisco, México; 3) Instituto de Investigación Cardiovascular, Nuevo Hospital Civil de Guadalajara, Guadalajara, Jalisco, México.

The present study was aimed to analyze the distribution of haplotypes derived from four *CAPN10* gene variants (SNP-44, SNP-43, indel-19 and SNP-63) according to clinical and biochemical parameters related to hypertension, because it has been described that calpain 10, which is encoded by the *CAPN10* gene, is involved in the secretion and insulin action and has also been reported its association with metabolic disorders related to hypertension. **Methods.** The study included 189 patients with hypertension and 97 normotensive subjects. Measurements of anthropometric parameters and biochemical estimations were performed. SNP-44 (rs2975760, IVS3 g.4841 T>C), SNP-43 (rs3792267, IVS3 g.4852 G>A) and SNP-63 (rs5030952, IVS13 g.16378 C>T) were identified by PCR-RFLP and the indel-19 variant (rs3842570, IVS6 g.7920 ins/del 32-bp) by gel based detection of PCR products-length differences. Chi-squared tests, with Fisher's correction when necessary, were applied to compare haplotype frequencies between patients with hypertension and normotensive subjects. In the hypertensive patients, haplotype distribution was compared according to clinical and biochemical parameters related to hypertension using the non-parametric Mann–Whitney U test, and the contribution of each haplotype to the risk of developing hypertension was determined by multivariate logistic regression analysis. **Results.** Haplotype frequencies did not differ significantly between patients with hypertension and normotensive subjects. However, in hypertensive patients, when analyzing the distribution of haplotypes according to systolic pressure, diastolic pressure, BMI, glucose, insulin, HOMA, and lipid profile (cholesterol, triglycerides, LDL, HDL), statistically significant difference was observed for the diplotype 1121/2111 regarding the systolic pressure (OR=0.186, CI=0.040-0.864; p=0.032) and the diplotype 1221/2111 with LDL levels (OR=0.117, CI=0.015-0.934; p=0.043). **Conclusion.** The results suggest that the diplotype 1121/2111 is associated with decrease of systolic pressure, and the diplotype 1221/2111 with a decrease in LDL levels in our study group.

662T

Genetic analysis of 1968 WHO Group 1 Pulmonary Arterial Hypertension (PAH) patients enrolled in the National Biological Sample and Data Repository for PAH (PAH Biobank). *M. Pauciulo, K. Lutz, C. Winslow, A. Walsworth, A. Gygi, M. Barve, A. Reponen, L. Martin, H. He, W. Nichols.* Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH.

The PAH Biobank was established to build a resource of biological specimens, clinical data and genetic data available to the PAH research community to further our understanding of the pathogenesis of this potentially fatal disease. Information on the resource and access to the samples and data can be found at www.pahbiobank.org. To date, 2,308 patients have been enrolled from 32 Enrolling Centers across the U.S. Blood samples collected at the Enrolling Centers are shipped overnight to the PAH Biobank. Blood is processed to store plasma and serum, extract DNA and RNA, and prepare immortalized cell lines. DNA sequencing is performed using the Illumina TruSeq Custom Amplicon system to interrogate the coding sequences/intron-exon junctions of *BMPR2*, *ALK1*, *ENG*, *CAV1*, *KCNK3*, *EIF2AK4*, and *SMAD9*. Genotyping of >4.3 million SNPs is done using the Illumina HumanOmni5 Beadchip. MLPA is performed to identify exonic deletions/duplications of *BMPR2*, *ALK1*, and *ENG*. Genetic analysis has been completed in 1,968 patients including 951 IPAH, 889 APAH, 67 FPAH, and 47 Drugs and Toxins PAH. 445 distinct variants occurred 11,639 times. 176 distinct synonymous and intronic variants occurred 9,438 times. The remaining 269 coding variants were observed 2201 times. This included 26 nonsenses in 42 patients, 51 ins/dels in 366 patients, and 11 splice site variants in 15 patients. 181 missenses occurred 1178 times. 74 missenses are predicted to be pathogenic/suspected pathogenic with 47 classified as variants of unknown significance. Of the 368 distinct coding/splice variants, 123 were in *BMPR2* with 98 predicted to be pathogenic or suspected pathogenic. 63 of the remaining 245 coding variants are predicted to be pathogenic or suspected pathogenic and spread across *ALK1* (15), *ENG* (11), *CAV1* (6), *SMAD9* (11), *KCNK3* (4), and *EIF2AK4* (15). MLPA identified 21 exonic insertion/deletions in *BMPR2* (19), *ALK1* (1) and *ENG* (1). The PAH Biobank is a powerful resource available to the PAH research community. This is the largest consecutively enrolled PAH cohort for which genetic analysis has been completed to date. While mutations have been identified in several genes, *BMPR2* remains the gene with the most pathogenic variants. Stratification of the cohort based on these genetic analyses could enable the identification of additional, novel genes and/or modifier genes affecting disease severity/penetrance in those harboring pathogenic or suspected pathogenic mutations in previously identified genes.

663F

Metabolic and transcriptomic fingerprints of menopausal hormone therapy in Finnish women. *A. Joensuu[1,2], J. Kettunen[3,4,2], M. Jauhiainen[2], A.J. Kangas[3], P. Soininen[3,4], A. Jula[5], V. Salomaa[2], J. Eriksson[6], M. Ala-Korpela[3,4,7], M. Perola[1,2,8], K. Auro[1,2,9].* 1) National Institute for Health and Welfare, Helsinki, Finland; 2) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 3) Computational Medicine, Faculty of Medicine, University of Oulu & Biocenter Oulu, Oulu, Finland; 4) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 5) National Institute for Health and Welfare, Turku, Finland; 6) Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 7) Computational Medicine, School of Social and Community Medicine & Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 8) University of Tartu, Tartu, Estonia; 9) Helsinki University Central Hospital, Department of Gynecology and Obstetrics, Helsinki, Finland.

Background: Menopause influences the health of all middle-aged women. Hormone therapy (HT) is commonly used to treat menopausal symptoms, but the effect of HT on cardiovascular disease (CVD) risk is under strong debate. The analysis of metabolic and transcriptomic profiles of HT users and non-users introduces novel information on factors affecting disease risk. **Methods:** We studied 4146 non-pregnant, non-lipid medicated 40-70-year-old women from four Finnish population cohorts, FINRISK 1997, DILGOM, HBCS and Health2000. Levels of 74 serum metabolites were quantified using an NMR platform. Metabolites were individually treated as dependent variables in linear regression corrected for age and fasting. HT users (N=1017) were studied against a) premenopausal (pre-MP, N=1362) and b) postmenopausal (post-MP, N=1759) women independently in each cohort and results were combined with random effects meta-analysis. The transcriptomic effects of HT use was studied against postmenopausal age-matched women in 74 DILGOM participants. **Results:** The metabolic profiles of pre-MP women and HT users are very similar: $p<0.05$ Bonferroni corrected differences were seen in only three metabolites; monounsaturated fatty acids, the ratio of monounsaturated fatty acids to total fatty acids and histidine. On the contrary, a clear atherogenic shift was demonstrated when comparing HT users and post-MP women: levels of 42 of the 74 studied metabolites differed. Levels of VLDL, IDL and LDL subclasses were higher and HDL subclasses lower in post-MP women. Also levels of serum cholesterol, triglycerides, apoB-100, omega-3 and 6 fatty acids, the amino acids glutamine, glycine, isoleucine and tyrosine as well as glucose, albumin and the degree of glycoprotein acetylation were clearly higher in post-MP women. The transcriptomic results support the metabolomics findings; levels of 194 transcripts differed ($p<0.01$) between HT and post-MP women, and DAVID annotation analysis reveals genes involved in the GO Biological processes Generation of precursor metabolites and energy and Regulation of gene expression, epigenetic as well as Steroid biosynthetic process and Steroid metabolic process enriched in these transcripts. **Conclusions:** Our results provide the first metabolome-wide population-based knowledge on using HT. The data demonstrate that HT usage is associated with a metabolic profile similar to that of premenopausal women, and compatible with a cardio-protective effect.

664W

Chamber-specific miRNA expression profiles of the human heart. *Y. Kakimoto[1], M. Tanaka[2], H. Kamiguchi[2], H. Hayashi[2], E. Ochiai[1], M. Osawa[1].* 1) Department of Forensic Medicine, Tokai University School of Medicine, Isehara, Japan; 2) Support Center for Medical Research and Education, Tokai University, Isehara, Japan.

The human heart contains two characteristic muscles: the ventricular myocardium and the atrial myocardium. The ventricular myocardium plays a central role in contraction to pump blood out from the heart to the entire body, whereas the atrial myocardium mainly serves as the source and target of neurohumoral signals. Gene expression analyses have shown that some mRNAs are specifically expressed in certain human cardiac chambers; however, chamber-specific miRNA expression signatures remain unclear. We performed deep sequencing of miRNA in 4 pairs of human left atria (LA) and left ventricles (LV) under normal physiological conditions. Of the total 19,830,155 reads generated from 8 libraries, 17,939,353 (90.5%) could be aligned to the human genome sequence, of which 6,975,989 reads (35.2%) were mapped to known miRNA sequences. In total, 1,134 mature miRNAs were identified and miR-1 was found to be the most abundant miRNA in both chambers, representing 21% of the miRNAs in the LA and 26% of the miRNAs in the LV. Twenty-five miRNAs were differentially expressed between the LA and the LV; 14 miRNAs were upregulated in the LA and 11 were upregulated in the LV. Subsequent real-time PCR of the miRNAs of 9 pairs of LA and LV validated the chamber-specific miRNA expression patterns. Notably, the miR-208 family members showed prominent chamber specificity in their expression; miR-208a-3p and miR-208a-5p were abundant in the LA, whereas miR-208b-3p and miR-208b-5p were preferentially expressed in the LV. miR-208a is encoded within an intron of the α-cardiac myosin heavy chain (α-MHC) gene (*MYH6*) and miR-208b is encoded within an intron of the β-cardiac myosin heavy chain (β-MHC) gene (*MYH7*), both of which are located on human chromosome 14. In the human heart, α-MHC is mainly expressed in the atrium, whereas β-MHC is extensively expressed in the fetal ventricle and moderately expressed in the adult ventricle. It is known that genetic deletion of miR-208a induces atrial fibrillation and the expression of ventricular miR-208b increases under cardiac stress, which is caused by conditions such as hypertrophy and ischemia. The anatomical expression profiles of the miR-208 family members support their chamber-specific functions in the human heart. The findings of this study might enhance our understanding of cardiac miRNAs and various heart diseases.

665T

An essential splice site mutation in *flt1* protects against early-stage atherosclerosis in zebrafish larvae. *M. den Hoed[1], M.K. Bandaru[1], L. Conrad[1], A. Emmanouilidou[1], P. Ranefall[2], C. Wählby[2], A. Larsson[3], E. Ingelsson[1,4].* 1) Department of Medical Sciences, Molecular Epidemiology, SciLifeLab, Uppsala University, Sweden; 2) Department of Information Technology, Division of Visual Information and Interaction and SciLifeLab, Uppsala University, Sweden; 3) Department of Medical Sciences, Biochemical structure and function, Uppsala University, Sweden; 4) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.

Objective: Genome-wide association studies (GWAS) identified 56 loci that are associated with the risk of coronary artery disease (CAD). For most loci the causal genes are currently unknown. We previously showed that: 1) CAD-associated loci are also associated with early-stage atherosclerosis; 2) zebrafish larvae are a promising model system for high-throughput, largely image-based screens of early-stage atherosclerosis. We now aim to identify causal genes in GWAS-identified loci for CAD using zebrafish model systems. **Methods**: Bioinformatics analyses identified *CXCL12*, *COL4A1* and *FLT1* as positional candidate genes in CAD-associated loci. Zebrafish orthologs were identified and heterozygous carriers of nonsense (sa1516 in cxcl12b; fh336 in col4a1) or essential splice site (sa1504 in flt1) mutations were crossed into transgenic backgrounds with fluorescent markers on macrophages (Tg:mpeg1-mCherry) and neutrophils (Tg:mpo-EGFP). After an in-cross of transgene positive heterozygous carriers, offspring were (over)fed on a normal or cholesterol-supplemented diet from 5 to 10 days post-fertilization. At 10 days, larvae were soaked in a lipid-staining dye, followed by imaging the vasculature with an automated positioning system, fluorescence microscope and CCD camera. Atherogenic traits were subsequently quantified using a custom-written image analysis pipeline in ilastik, CellProfiler and ImageJ. After imaging, whole-body total cholesterol and triglyceride levels were assessed enzymatically and larvae were genotyped using KASP technology. Multilevel mixed models were used to assess the effect of each mutation (additive model) on atherogenic traits, while adjusting for diet, time-of-imaging and batch (random factor). **Results**: In data from up to 258 larvae, each additional copy of sa1504 in flt1 resulted in 0.3±0.1 SD less vascular lipid deposition (P=6E-5) and co-localization of lipids and macrophages (P=1E-2), without effects on total cholesterol or triglyceride levels (P>0.1). No effects were identified for mutations in col4a1 and cxcl12b (N up to 344 and 291, respectively). **Discussion**: Our results represent the first mutant model demonstrating that flt1 plays a role in early-stage atherosclerosis, implying that *FLT1* is the culprit for CAD in the 13q12 locus. *COL4A1* and *CXCL12* remain the most promising candidates in their respective loci. The lack of effect for these genes in our screen may reflect genetic redundancy, genetic adaptation and/or a late onset effect.

666F

The I-PICC Study: Protocol for a point-of-care randomized controlled trial of statin pharmacogenetics in a learning healthcare system. *S.J. Miller[1], R.E. Ferguson[1,2], J.M. Gaziano[1,3,4], R.C. Green[3,4,5,6], J.L. Vassy[1,3,4].* 1) VA Boston Healthcare System, Jamaica Plain, MA; 2) Boston University School of Public Health, Boston, MA; 3) Brigham and Women's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Partners HealthCare Personalized Medicine, Cambridge, MA; 6) Broad Institute of MIT and Harvard, Cambridge, MA.

   Background: The pharmacogenetic (PGx) association between *SLCO1B1* rs4149056 and simvastatin myopathy is well validated, but the clinical utility of its implementation in patient care is unknown. Here we describe the protocol of the I-PICC Study (Implementing Pharmacogenetics in Clinical Care), a randomized controlled trial (RCT) of *SLCO1B1* genotyping in primary care. Design/Results: The I-PICC Study is enrolling 400 statin-naïve primary care and women's health patients across the Veteran Affairs Boston Healthcare System. Eligible patients are aged 45-70 and have elevated risk of cardiovascular disease (CVD) according to American College of Cardiology/ American Heart Association (ACC/AHA) guidelines. Primary care providers (PCPs) are also research subjects and consent via electronic health record (EHR) alerts. To model PGx genotyping at the point of care, we are enrolling patients when their PCPs order cholesterol testing, indicating a moment of clinical decision-making about CVD risk. Given this design, we discussed with the institutional review board options for patient consent that minimize PCP burden, including 1) waiving consent and allowing PCPs themselves to enroll patients; 2) mailing letters to eligible patients at study launch, giving them the chance to contact the staff to opt out of enrollment but otherwise be considered consented; and 3) mailing letters and requiring patients to contact staff expressly to opt in to enrolling. A consented patient is then enrolled only if and when their PCP co-signs an order for *SLCO1B1* testing, performed on a blood sample already collected in clinical care. Enrolled patients are randomized to have their PCPs receive results through the EHR immediately (PGx+) vs. after 1 year (PGx-). We will query clinical and pharmacy data for 1-year outcomes: myopathy and concordance with Clinical Pharmacogenetics Implementation Consortium (CPIC) simvastatin guidelines (drug safety) and cholesterol levels and concordance with ACC/AHA guidelines (CVD risk reduction). We are validating a natural-language processing algorithm for myopathy and have 80% power to detect a 15% difference in the proportions meeting CPIC and ACC/ AHA guidelines. Discussion: A point-of-care RCT, the I-PICC Study eliminates the need for separate study visits or blood draws while minimizing the burden to participating PCPs. Such studies pose special challenges for balancing human subjects protection with trial efficiency in a learning healthcare system.

667W

Genetic variants associated with statin-induced changes in gene expression in human lymphoblastoid cell lines. *E. Theusch, R.M. Krauss, M.W. Medina.* Children's Hospital Oakland Research Institute (CHORI), Oakland, CA.

   Statins are widely prescribed to lower plasma low-density lipoprotein cholesterol levels. They also have pleiotropic effects seemingly unrelated to their cholesterol-lowering properties, but the mechanisms are poorly understood. Though statins reduce cardiovascular disease risk overall, statin efficacy varies, and some people experience adverse side effects while on statin treatment.   To identify potential genetic modulators of clinical statin response, we looked for genetic variants associated with statin-induced changes in gene expression (differential eQTLs or deQTLs) in lymphoblastoid cell lines (LCLs) derived from participants of the Cholesterol and Pharmacogenetics (CAP) 40 mg/day 6 week simvastatin clinical trial. We exposed CAP LCLs to 2 μM simvastatin or control buffer for 24 hours and performed polyA-selected, strand-specific RNA-seq. Statin-induced changes in gene expression from 259 European ancestry or 153 African American ancestry LCLs were adjusted for potential confounders prior to association with genotyped and imputed genetic variants within 1 Mb of each gene's transcription start site.   From the meta-analysis of the two ancestral populations, we identified deQTLs for 7 genes (*TBC1D4*, *MDGA1*, *TDRD12*, *SERPINB1*, *GATM*, *GLUL*, and *DTD1*) that had unadjusted p-values under 5.8×10^{-9} (a conservative threshold based on 13,955 genes tested and 1 million independent markers per genome). deQTLs for another 4 genes (*CHI3L2*, *OAS3*, *OAS1*, and *DFNA5*) were identified in the European ancestry subset alone.   The lead deQTLs for *OAS3* and *OAS1* were in strong linkage disequilibrium (LD; $r^2 > 0.9$) and, unlike most of the other deQTLs identified, were not strong eQTLs for endogenous gene expression levels in our data. They were also in more modest LD ($r^2 > 0.8$) with a known *OAS1* splice site mutation and were strongly correlated with endogenous, but not statin-induced changes in, *OAS1* splicing in our data.   *OAS1* and *OAS3* are interferon-induced genes that play major roles in antiviral response. Interestingly, *OAS1* has been a target of intense natural selection in humans. This could explain why most *OAS1/3* deQTL variants are common (35% MAF) in Europeans but rare (2% MAF) in Africans and why there was no *OAS1/3* deQTL signal observed in the CAP African Americans. There is some epidemiological evidence that statins are protective against certain viruses, thus supporting future investigation of whether antiviral activity of statin treatment varies by *OAS1/3* deQTL genotype.

668T

Swedegene: Genome-wide association study of ACE inhibitor-induced cough in Sweden. *M. Wadelius[1], N. Eriksson[1,2], T. Axelsson[1], Q-Y. Yue[3], P.K.E. Magnusson[4], P. Hallberg[1].* 1) Department of Medical Sciences and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Uppsala Clinical Research Center, Uppsala, Sweden; 3) Medical Products Agency, Uppsala, Sweden; 4) Swedish Twin Registry, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

   Background: Angiotensin-converting enzyme (ACE) inhibitors are used to manage hypertension and heart failure. Treatment is usually successful, but a substantial number of patients experience drug intolerance in the form of persistent dry cough. Candidate gene studies have tested the association between cough and genes in the bradykinin pathway, but findings have been inconsistent. A recent genome-wide association study (GWAS) of patients with cough induced by ACE inhibitors identified associations with intronic SNPs in the potassium channel gene *KCNIP4* on chromosome 4 [1].   **Method**: We performed a GWAS on 124 patients with ACE inhibitor-induced cough and 4890 population controls collected by the Swedish adverse drug reaction bio-bank Swedegene (www.swedegene.se). Cases and controls were genotyped using the Illumina arrays HumanOmni2.5 and HumanOmniExpress 700K. After quality control, the merged genotyped set contained 600 000 single nucleotide polymorphisms (SNPs). After phasing and imputation at the Michigan Imputation Server that uses the software Eagle and Minimac3 with the Haplotype Reference Consortium panel, the dataset contained 8.6 million SNPs. We corrected for principal components 1-4. The genome-wide significance p-value threshold was set to p <5 E-08 to correct for multiple testing.   **Results:** ACE inhibitor-induced cough was significantly associated with SNPs on chromosomes 1, 3, 12 and 14 at a genome-wide level. The minor allele frequencies ranged from 0.04 to 0.09 in cases and 0.01 to 0.02 in controls. The strongest association was with rs36058006 located in an intergenic region on chromosome 1 (odds ratio 6.93 [95% confidence interval 3.65, 13.17], p = 3.42 E-09). The second strongest signal came from rs115510347, an intronic SNP in *KCNMB2* on chromosome 3 (odds ratio 6.25 [95% confidence interval 3.37, 11.59], p = 6.31 E-09). There was no association with the previously identified gene *KCNIP4* on chromosome 4 (p=0.982).   **Conclusion:** We found novel associations between ACE inhibitor-induced cough and uncommon SNPs that need to be replicated. One of the interesting associations was with the potassium channel gene *KCNMB2*. The association between cough induced by ACE inhibitors and another potassium channel gene, *KCNIP4,* was not replicated [1].   **Reference:** 1. Mosley, J.D., et al., *A genome-wide association study identifies variants in KCNIP4 associated with ACE inhibitor-induced cough.* Pharmacogenomics J, 2016; **16**(3): 231-7.

669F

Genetic variants in the lipoprotein (a) gene and atherothrombosis in Mexican population. *L.I. López-Salazar[1], E.A. Hernández-Tobías[1], R. Camacho-Mejorado[1], G. Noris[2], C. Santana[2], R. Gómez[1].* 1) Departamento de Toxicología, Cinvestav-IPN, Mexico City, Mexico; 2) Laboratorio Biología Molecular Diagnóstica, Querétaro, México.

   Atherothrombosis (ATT) is a complex disease that involves atherosclerotic lesions disruption and clot formation causing myocardial infarction and stroke. Therefore, ATT is the major cause of death worldwide. ATT is a highly heritable and polygenic condition with differential ethnic-susceptibility. Consequently, genetic variants in candidate genes could act as potential predictors. Lipoprotein (a) gene (*LPA*) has been related to cholesterol levels, pro-thrombotic effects and inflammation, all of them related to ATT development. Hence, the aim of the present study was to evaluate the contribution of three SNPs (rs9457951, rs10455872 and rs3798220) in *LPA* to ATT development in Mexican population.   204 cases and 204 controls were genotyped using real-time PCR through TaqMan assays. In addition, a genomic control (n=300) was included to reduce the spurious associations because Mexican population is an admixed one. An ancestral control (n=140) was also included (Mazahuas, Me'Phaas and Nahuas). Population genetic parameters were calculated through Arlequin software.   A nuance association between rs9457951-GC and hypercholesterolemia (*P*=0.02) was found. In addition, the genotype CC in rs3798220 showed important differences when Mexican populations (Mestizo) were compared to worldwide populations from 1000 genomes (0.08 *vs.* ≤ 0.0001, respectively). This finding was reinforced to compare with Native American populations, which exhibited the highest frequencies of this genotype (0.37 in Nahuas, 0.26 in Me'Phaas and 0.30 in Mazahuas). Hardy-Weinberg equilibrium was found in all populations. However, Mazahuas exhibited important but not significant homozygous excess (F_{is} = 0.345).   Our results suggest a possible contribution of Amerindian ancestry to ATT development, given that rs3798220-CC genotype has been related to cardiovascular risk. Further, heterozygous genotype in rs9457951 might be associated with hypercholesterolemia. Nevertheless, further studies increasing sample size should be realised to confirm these findings.

670W

Identifying protein quantitative trait loci for plasma and serum proteins in the Tromsø Study. *T. Solomon[1], E. Smith[2], J. Lapek[3], K. Hindberg[4], S.B. Jensen[5], N. Latysheva[5], S.K. Brækkan[5,6,7], T. Wilsgaard[8], D. Gonzalez[3], J.B. Hansen[5,6,7], K.A. Frazer[2,5,9].* 1) Biomedical Science Graduate Program, University of California, San Diego, La Jolla, CA; 2) Department of Pediatrics and Rady's Children's Hospital, University of California, San Diego, La Jolla, CA; 3) Department of Pharmacology, University of California, San Diego, La Jolla, CA; 4) Department of Mathematics and Statitics, UiT The Arctic University of Norway, Tromsø, Norway; 5) K.G. Jebsen Thrombosis Research and Expertise Center (TREC), Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway; 6) Hematological Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway; 7) Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway; 8) Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway; 9) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA.

   Statement of purpose: Genetic variation can be used to study causal relationships between biomarkers and diseases. We have identified new common and rare genetic variants associated with cardiovascular-related protein levels (protein quantitative trait loci; pQTLs). We functionally annotated these pQTLs, predicted and experimentally confirmed a novel molecular interaction and determined which pQTLs are associated with diseases and physiological phenotypes.   **Methods and results:** As part of a larger case/control study of VTE, serum levels of 51 proteins implicated in cardiovascular diseases were measured in 330 individuals from the Tromsø Study. Exonic genetic variation near each protein's respective gene (*cis*) was identified using sequencing and arrays. Using single site and gene-based tests, we identified 27 genetic associations between pQTLs and the serum levels of 20 proteins: 14 associated with common variation in *cis*, of which six are novel (i.e. not previously reported); seven associations with rare variants in *cis*, of which four are novel; and six associations in *trans*. Of the 20 proteins, 13 were associated only with single sites, five with only rare variation, and two with both single site and rare variation. *cis*-pQTLs for kallikrein and F12 also show *trans* associations for proteins (uPAR, kininogen) known to be cleaved by kallikrein as well as with NTproBNP. We experimentally demonstrate that kallikrein can cleave proBNP (NTproBNP precursor) *in vitro*. Nine of the pQTLs have previously identified associations with 17 diseases and/or physiological phenotypes. To further validate these findings and identify novel associations, we have performed and are currently analyzing un-targeted mass spectrometry on plasma from an additional 160 individuals in the Tromsø Study.   **Conclusions:** We have identified *cis* and *trans* genetic variation associated with the serum levels of 20 proteins and utilized these pQTLs to study molecular mechanisms underlying diseases and/or physiological phenotypes.

671T

Fetal microsatellite in the *HMOX-1* promoter is associated with early-onset and severe preeclampsia. *T. Kaartokallio[1], S. Utge[1], M.M. Klemetti[2,3], J. Paananen[4], S. Heinonen[2], E. Kajantie[5,6,7], J. Kere[8,9,10], K. Kivinen[11], A. Pouta[7,12], P. Lakkisto[13,14], H. Laivuori[1,2,15].* 1) Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 2) Obstetrics and Gynaecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Obstetrics and Gynecology, South-Karelia Central Hospital, Lappeenranta, Finland; 4) Bioinformatics Center, University of Eastern Finland; 5) Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland; 6) Children's Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland; 7) PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland; 8) Department of Biosciences and Nutrition, and Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; 9) Molecular Neurology Research Program, University of Helsinki, Helsinki, Finland; 10) Folkhälsan Institute of Genetics, Helsinki, Finland; 11) Division of Cardiovascular Medicine, University of Cambridge, Cambridge, UK; 12) Department of Government services, National Institute for Health and Welfare, Oulu, Finland; 13) Minerva Foundation Institute for Medical Research, Helsinki, Finland; 14) Clinical Chemistry and Hematology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 15) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

   Introduction: Preeclampsia is a potentially severe and clinically heterogeneous vascular pregnancy disease characterized by systemic inflammation, endothelial dysfunction and often impaired placental development. Hemeoxygenase-1 gene (*HMOX-1*) encodes the HO-1 protein, a stress response enzyme with antioxidant and anti-inflammatory functions. The promoter region of *HMOX-1* contains a regulatory GT_n microsatellite repeat, which long version is associated with reduced *HMOX-1* expression. We have previously shown that pregnant women with the long repeat are at higher risk of developing late-onset, less severe form of preeclampsia. In this study, our aim was to investigate whether fetal genotype of the repeat also affects mother's preeclampsia risk. **Materials and methods:** We genotyped the GT_n repeat in cord blood DNA from 609 preeclamptic and 745 non-preeclamptic pregnancies from the Finnish Genetics of Preeclampsia Consortium cohort using DNA fragment analysis. **Results:** The long allele of the repeat was associated with preeclampsia (p=0.041, OR 1.18, 95% CI 1.00-1.38) and early-onset preeclampsia (p=0.026, OR 1.39 (1.03-1.86)). Moreover, the long genotype was associated with preeclampsia (additive model: p=0.024, OR 1.51 (1.05-2.16); dominant model: p=0.031, OR 1.45 (1.03-2.03)), early-onset preeclampsia (additive model: p=0.041, OR 2.08 (1.03-4.21)) and severe preeclampsia (additive model: p=0.021, OR 1.62 (1.07-2.44); dominant model p=0.021, OR 1.58 (1.07-2.33)). Based on interaction analysis the effect of child's genotype is independent of the mother's genotype. **Conclusions:** Maternal and fetal *HMOX-1* risk alleles seem to predispose to two preeclampsia subphenotypes. Decreased HO-1 level may affect endothelial adaptation to pregnancy and placental development, two mechanisms central in the development of preeclampsia.

672F

The effect of *NR3C1* and *VDR* polymorphisms on stable warfarin doses in patients with mechanical cardiac valves. *H. Kim[1], Y. Cho[1], B. Yi[1], H. Gwak[2], K. Lee[1].* 1) College of Pharmacy, Chungbuk National University, Cheongju, Republic of Korea, South Korea; 2) College of Pharmacy & Division of Life and Pharmaceutical Sciences, Ewha Womans University, Seoul.

   Background It is widely recognized that genotype significantly influences pharmacokinetics and pharmacodynamics of warfarin and contributes to the interpatient variability in warfarin dose requirements. The aim of this study was to evaluate the effects of polymorphisms of CYP2C9 transcription factor genes, NR3C1 and VDR, on stable warfarin doses in Korean patients with mechanical heart valves. METHODS Two hundred and four patients were enrolled including patients who had mechanical heart valve prosthesis and maintained a stable warfarin dose with INR of 2.0-3.0 for three consecutive times. In order to investigate the influences of genetic polymorphisms on the inter-individual variability of daily warfarin dose requirements, 17 single nucleotide polymorphisms (SNPs) were genotyped; VKORC1 (rs9934438), CYP2C9 (rs1057910), CYP4F2 (rs2108622), NR3C1 (rs41423247, rs1800445, rs56149945, rs10052957, rs6198, rs33388, rs6196, rs244465), and VDR (rs1544410, rs11568820, rs731236, rs757343, rs7975232, rs2228570). Statistical analyses were conducted to evaluate the effects of NR3C1 and VDR allelic variations on stable warfarin doses. RESULTS VDR SNP combination of rs7975232 and rs2228570 showed significant associations with stable warfarin doses, along with VKORC1, CYP2C9 and CYP4F2 polymorphisms. Patients with the combination of GT,TT/CT,CC genotypes of VDR rs7975232/rs2228570 required significantly higher warfarin stable doses (5.79 ± 2.02 mg) than those with the other combined genotypes (5.19 ± 1.78 mg). Multivariate analysis showed that VDR rs7975232/rs2228570 contributed to the overall warfarin dose variability by 1.7% out of 42.9% in total.   CONCLUSION This study demonstrated that stable warfarin dose is associated with VDR SNPs in combination as well as VKORC1, CYP2C9, and CYP4F2 .

673W

Identification of novel heart failure plasma biomarkers using the Hybrid Mouse Diversity Panel (HMDP) cardiac transcriptome data. *A. Huertas-Vazquez[1], M. Seldin[1], P. Gupta[1], J. Hsiao[1], G. Stolin[1], A. Lusis[1,2,3], J. Wang[1].* 1) Department of Medicine/Division of Cardiology, David Geffen School of Medicine, University of California, Los Angeles 650 Charles E. Young Drive South. Room A2-237 CHS; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles; 3) Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles.

   Heart Failure (HF) is a heterogeneous condition characterized by a large number of pathological abnormalities with vast differences in outcomes. HF occurs in over 6 million people in the US and it is projected that at least 8 million will have HF by 2030. Therefore, the discovery of novel biomarkers for more accurate and predictive risk assessment remains a priority. The purpose of the study was to identify novel plasma biomarkers for HF by using cardiac transcriptome data. We carried out cardiac transcriptome profiling using Illumina arrays among 100+ mouse strains from the Hybrid Mice Diversity Panel (HMDP), including both, controls and mice sustaining chronic heart injury due to systemic isoproterenol (ISO) treatment. We identified differentially expressed genes using the statistical package LIMMA implemented in R. For further validation, we performed Western blots to evaluate whether the differentially expressed cardiac transcripts were also differentially abundant at the plasma protein level in C57BL/6J mice under control or ISO treatment for 3 weeks. We observed an enrichment for genes known to be involved in HF including *Lgals3*, P=9.73 X 10-22 (galectin-3) and *Timp1*, 1.01 X 10-22 (TIMP metallopeptidase inhibitor 1) as well as newly identified proteins, including rgs1 (regulator of G-protein signaling 1). Based on known transcript abundances in the HMDP, we observed the expected differences in plasma protein expression for LGALS3 and RGS1 between control and ISO treated mice over 3 weeks of ISO treatment. Our findings showed the promise and feasibility of novel biomarker discovery using mice raised in a controlled environment from the HF-HMDP cardiac transcriptome data. The contribution of identified proteins in HF human samples is under investigation.

674T

Gene-alcohol interactions identify novel blood pressure loci in African Ancestry using 1000G imputations: The CHARGE Gene-Lifestyle Interactions Working Group. *M.F. Feitosa[1], A.T. Kraja[1], T. Winkler[2], A.R. Bentley[3], M. Brown[4], K. Schwander[5], Y.J. Sung[5], S.K. Musani[6], S.M. Tajuddin[7], K. Lohman[8], M. Richard[9], T. Bartz[10], N. Franceschini[11], X. Guo[12], E. Fox[6], S.L.R. Kardia[13], C. Rotimi[3], M. Fornage[8], M.A. Province[1], A. Morrison[4] on behalf of the CHARGE Gene-Lifestyle Interactions Working Group.* 1) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 3) Center for Research on Genomics and Global Health, NHGRI, NIH, Bethesda, MD; 4) Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health Science Center, Houston, TX; 5) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; 6) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 7) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD; 8) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 9) Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX; 10) Department of Epidemiology, University of Washington, Seattle, WA; 11) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 12) Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA; 13) Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI.

Genetic and lifestyle factors and their interactions contribute to the risk for hypertension. Heavy alcohol consumption is associated with high blood pressure (BP) and cardiovascular disease; however, it remains controversial whether light-to-moderate consumption is beneficial or harmful. We performed two genome-wide meta-analyses of SNP-alcohol interactions using two exposure types: current drinker (Yes/No) and, among drinkers, if light (1-7 drinks/week) or heavy (8-more drinks/week) consumption. Our analyses included 21,417 individuals (11,224 non-drinkers and 10,193 current drinkers) from 12 African Ancestry studies. We conducted SNP-alcohol interaction associations on systolic, diastolic, mean arterial, and pulse BPs. Studies calculated the joint 2 degree of freedom test for each SNP and SNP-alcohol interaction effects. We applied study- and meta-based genomic control adjustments. SNP results are reported if: (*a*) minor-allele count >10; (*b*) df>20 (compound statistic for allele frequency, N exposed, imputation quality); (*c*) >50 individuals alcohol-exposed and >5,000 N individuals, (*d*) >2 studies for each SNP, and (*e*) meta-heterogeneity p >0.1. Besides replicating several known BP loci, we identified 18 novel loci for BP (p<5E-08): eight interacting with light-heavy alcohol (*KCNMB2, USP15, DNAJC3, ZMAT4, TACC2, TUSC1, LOC100996669, SP100*), and ten interacting with current alcohol consumption (*CHRM2, LINGO2, PTPRE, GALNT18, FAT3, PCDH9, LDLRAD4, FAM19A2, LINC00689, TCF21*). Our findings show that *CHRM2* (Cholinergic-Receptor, Muscarinic 2) and *LINGO2* (Leucine Rich Repeat And Ig Domain Containing 2), which have been reported in association with alcohol dependence, also interact with alcohol consumption influencing BP. Associations have also been reported with diabetes, glucose metabolism (*ZMAT4, TACC2, LINGO2, GALNT18, FAT3*), heart disease (*TUSC1, TCF21, CHRM2*), angiotensin II receptor blocker response (*TCF21*), neurological psychiatric disorders (*CHRM2, KCNMB2, PTPRE, ZMAT4, GALNT18, FAT3*), and calcium or potassium channel mediation (*FAT3, PTPRE, KCNMB2, CHRM2, PCDH9*). Bioinformatics from GeneGO analyses showed enrichment for urinary tract (p_{FDR}=6.4E-11), heart rhythm (p_{FDR}=2.2E-09) and personality disorders (p_{FDR}=2.1E-07). These findings identify both novel and previously known BP genes which demonstrate significant interactions with alcohol consumption which may provide insights into mechanisms of high BP, therapeutic interventions, and novel drug targets.

675F

Rare coding variants associated with blood pressure in ~16,000 individuals of African ancestry. *P. Nandakumar[1], D. Lee[1], M.A. Richard[2], F. Tekola-Ayele[3], B.O. Tayo[4], E. Ware[5,6], Y.J. Sung[7], B. Salako[8], A. Ogunniyi[8], C.C. Gu[7], M.L. Grove[9], M. Fornage[2], S. Kardia[5], C. Rotimi[3], R.S. Cooper[4], A.C. Morrison[10], G. Ehret[1,11], A. Chakravarti[1].* 1) McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 3) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Department of Public Health Sciences, Stritch School of Medicine, Loyola University Chicago, Maywood, IL; 5) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 6) Institute for Social Research, University of Michigan, Ann Arbor, MI; 7) Division of Biostatistics, Washington University School of Medicine in St. Louis, St. Louis, MO; 8) Department of Medicine, University of Ibadan, Ibadan, Nigeria; 9) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 10) Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health Science Center at Houston, Houston TX; 11) Department of Specialties of Medicine, Geneva University Hospitals, Geneva, Switzerland.

Genome-wide association studies (GWAS) of blood pressure (BP) have implicated common variants at ~80 loci for systolic (SBP) and diastolic (DBP) pressure across multiple populations, but the specific genes involved are unknown. This study focuses on identifying these and other specific genes associated with BP traits in individuals of African ancestry using eight cohorts (AADM, ARIC, CARDIA, GenNet, GENOA, HUFS, HyperGEN, LUC) genotyped on the Illumina HumanExome BeadChip v1.0 (n=247,870 variants) or v1.1 (n=242,901). Data quality control yielded a final set of 15,914 individuals and 170,540 variants for analysis. Fixed-effects meta-analyses were performed with the seqMeta R package for SBP and DBP, each adjusted for age, age-squared, sex, BMI and the first 10 principal components of ancestry from each cohort. These analyses include single variants analyzed by frequency class (common: MAF≥0.05, low frequency: 0.01≤MAF<0.05 and rare: MAF<0.01), gene-based analysis with SKAT and the T1 burden test on variants with MAF<0.01, and those considered "damaging" (intronic splice, NMD-compatible stop-gain, and conserved missense (phyloP>4)), multiple test corrected by the Bonferroni method. **Results**: Associations of rare variants in *COL6A1* (rs150432347, SBP, P=1.19 x 10^{-7}), encoding a collagen VI protein providing structural support in the heart, and *SLC28A3* (rs11568416, SBP, P=8.92 x 10^{-8}), encoding a sodium-dependent nucleoside transporter which may regulate vascular tone, were identified, with 19 and 6 copies across the six cohorts, respectively. Further, associations of rare variants with 2 to 15 carriers in each of *SEL1L3, KRBA1, YOD1, CRYBA2, GAPDHS,* and *AFF1* were identified for SBP and DBP. The T1 burden test on damaging variants identified two significant associations: *CCDC13* (SBP, P=3.54 x 10^{-7}), a centriolar satellite protein playing a role in primary ciliogenesis, relevant to the poor functioning of cilia resulting in weak vascular tone characteristic of polycystic kidney disease; the *QSOX1* gene (DBP, P=3.86 x 10^{-6}) encodes a sulfhydryl oxidase enzyme involved in vascular smooth muscle cell migration and proliferation *in vitro*. These results suggest credible candidate genes for BP regulation in African-ancestry individuals.

676W

Relationship between C-reactive protein serum concentration and the 1846 C>T polymorphism in elderly patients with Acute Coronary Syndrome from Western Mexico. *G.L. Reynoso Villalpando[1,2,3], J.R. Padilla Gutierrez[1,2], A. Valdez Haro[1,2,3], F. Casillas Muñoz[1,2,3], H.E. Flores Salinas[4], J.F. Muñoz Valle[1,2], Y. Valle[1,2].* 1) GENETICS, UNIVERSIDAD DE GUADALAJARA , Guadalajara , Jalisco, Mexico; 2) Instituto de Investigacion en Ciencias Biomedicas; 3) Doctorado en Genetica Humana, Centro universitario de Ciencia Salud, Guadalajara; 4) Unidad Medica de alta especialidad, Centro Medico Nacional de Occidente, Guadalajara.

  Background: C-reactive protein (CRP) is an inflammation biomarker and risk predictor in cardiovascular disease. Variants of the *CRP* gene are associated with serum protein concentration. Current evidence supports a relationship between elevation on CRP concentration and genetic polymorphisms thus increasing risk of Acute Coronary Syndrome (ACS). We evaluate the association between the 1846C>T (rs1205) polymorphism with CRP protein concentration in ACS. Methods: A total of 300 controls and 300 patients with ACS from Western Mexico were included. Genotyping was performed with polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). High sensitivity CRP concentration (hsCRP) was measured by immunoturbidimetry. For IL-6 measurement we used the Quantikine Human IL-6 Immunoassay (Quantikine ELISA, R&D), a solid phase sandwich ELISA, according to manufacturer instruction. Results: Serum CRP concentration was increased in patients (19.80 mg/L) compared with controls (4.10 mg/L, p=<0.0001). By clinical spectrum, ST-segment elevation myocardial infarction had increased CRP concentration than unstable angina patients (21.8 vs 5.9 mg/L, p=<0.0001). The 1846C>T *CRP* polymorphism was not associated with ACS. However, T/T carriers had lower CRP concentration than C/C (2.5 *vs* 5.1 mg/L, p=0.032) in CG. IL-6 showed a low positive correlation with CRP concentration in ACS and controls (r=0.20-0.44, p<0.05). Conclusion: ACS patients had higher cardiovascular risk as denoted their increased CRP levels. CRP concentration seems to be related with ACS severity. The 1846C>T polymorphism is not a susceptibility genetic marker to ACS in western Mexico population. However, the allele 1846T is associated with lower CRP concentration regardless disease.

677T

Multiple independent variants at the *MCPH1/ANGP2T* locus influence circulating concentrations of serum angiopoietin-2. *S. Salimi[1], K. Ryan[2], V. Thomeas[3], T. Dave[2], J.R. O'Connell[2], J. Perry[2], A.R. Shuldiner[2], M.L. Maitland[3], B.D. Mitchell[2].* 1) Department of Epidemiology and Public Health, University of Maryland School of Medicine; 2) Department of Medicine, Division of Endocrinology, University of Maryland School of Medicine; 3) The University of Chicago Medicine.

  Background Angiopoietin-2 (Ang-2) is an endothelium-derived growth factor and the concentration of Ang-2 [Ang-2] measured in circulation is a candidate biomarker for multiple cardiovascular pathophysiologic processes. To elucidate the underlying genetic determinants of variation in [Ang-2], we performed a genome-wide association (GWA) analysis of serum [Ang-2] in 1,172 healthy Old Order Amish individuals. **Methods:** Subjects were genotyped on the Affymetrix 500K/6.0 platforms and imputed to the 1000G reference panel. Following GWAS, we performed sequential conditional analysis at associated loci to explore whether multiple SNPs contributed independently to residual variation in [Ang-2] after accounting for age, sex, and body mass index. We then identified all SNPs in high linkage disequilibrium with these SNPs and annotated their potential effects on gene function using ROADMAP Epigenomics. **Results:** The heritability of serum [Ang-2] was 0.38, $p < 4×10^{-12}$. On initial GWA analysis, rs2440393, at the *MCPH1/ANGPT2* overlapping locus, was significantly associated with [Ang-2] ($p=3.47×10^{-8}$) with a minor allele frequency of 0.06. This SNP had a 5-fold enrichment in Amish compared to European Caucasians from 1000G. Conditional analysis revealed a total of 3 SNPs at this locus (rs2440393, rs17625642, rs1968586) to be independently associated with [Ang-2] at $p < 0.001$, and these collectively accounted for 20% of the residual variation in [Ang-2]. These SNPs are in high linkage disequilibrium (LD) ($r^2 > 0.8$) with 46, 8 and 2 other SNPs, respectively. The rs2440393 and rs17625642 variants are in high LD with SNPs that are enriched enhancers in 25 and 31 tissues, respectively. The rs2440393 variant was associated with expression levels of multiple genes expressed in peripheral blood monocytes, including *FER, RMND5A, and RNF215* and *DEFA4*. **Conclusion:** Our results suggest that 20% of the variation of serum [Ang-2] is accounted by at least three independent signals at the *MCPH1/ANGPT2* locus. These variants are independently in high LD with numerous other tissue–specific putative functional SNPs with enhancers and DNase activities.

678F

Causal inference of platelet reactivity on cardiovascular disease risk: A Mendelian randomization study. *C. Song, A.D. Johnson, CHARGE Consortium Stroke Working Group.* Framingham Heart Study, Framingham, MA.

Background: Antiplatelet medication, *e.g.*, aspirin, is widely used in the prevention and treatment of cardiovascular disease. However, population studies report conflicting results for the association between variation in baseline platelet reactivity and incident cardiovascular disease (CVD). In this study, we aimed to understand the causal influence platelet reactivity on arterial CVD, including coronary heart disease (CHD) and ischemic stroke using a Mendelian randomization (MR) approach. Methods: Platelet reactivity was determined via platelet rich plasma measures of light transmission aggregometry in the Framingham Heart Study Offspring cohort. Measures included lag time in aggregation to collagen (Type 1 calf skin), and a range of ADP and epinephrine doses. ADP and epinephrine threshold concentrations (EC50) were the minimal concentration of each agonist to produce a >50% aggregation response. We selected the top locus for each platelet reactivity trait from the largest genome-wide association study (GWAS; Johnson et al. *Nat Genet* 2010). Then, we created genetic risk scores for each trait using SNPs from the same locus with low-to-moderate linkage disequilibrium (LD; $r^2 < 0.5$). We used these genetic risk scores as instrumental variables (IV), correcting for LD, in MR analyses. The association between IVs and platelet reactivity traits were obtained from the Framingham Offspring study while the association between IVs and CVD outcomes were obtained from previous Consortia GWAS. We then tested whether elevated platelet reactivity was associated with higher CVD risk by MR using summary statistics. Results: Our results show that shorter collagen lag time, which indicates increased platelet reactivity, had a causal effect on CHD risk [Odds ratio (OR) =2.53; P-value =0.007], but not significant on ischemic stroke risk [OR =2.23; P-value =0.630]. Decreased threshold concentrations of ADP or epinephrine indicate increased platelet reactivity and also showed a similar trend of influence on CHD. Conclusion: Our study indicates that increased platelet reactivity is a causal risk factor for arterial CVD, in line with therapeutic treatment strategies. This highlights potential clinical utility of platelet reactivity tests in early detection of CVD risk, though defining a precise test remains a significant future challenge. Further study is ongoing to understand the causal influence of platelet reactivity on venous CVD.

679W

A gene-by-environment interaction study of peripheral arterial disease identifies novel loci. *N.R. van Zuydam[1,2], M. de Andrade[3], A.I. Stilby[4], E. Vlachopoulou[5], E. Ahlqvist[6], E. Dahlström[7], V. Salomaa[8], A.P. Morris[1,9], C.N.A Palmer[2], J.C. Hopewell[4], H.M. Colhoun[2], I.J. Kullo[3], GoLEAD and SUMMIT.* 1) Wellcome Trust centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Medical Research Institute, University of Dundee, UK; 3) College of Medicine, Mayo Clinic, Rochester MN, USA; 4) CTSU, Nuffield Department of Population Health, University of Oxford, Oxford, UK; 5) University of Helsinki, Finland; 6) Lund University, Malmö, Sweden; 7) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki University Central Hospital, Helsinki, Finland; 8) National Institute for Health and Welfare, Helsinki, Finland; 9) Institute of Translational Medicine, University of Liverpool, Liverpool, UK.

Peripheral arterial disease (PAD) affects ~200 million people world-wide and is associated with increased mortality. PAD is a complex, heritable disease (21-58%) that is strongly influenced by environmental factors such as smoking and diabetes. Disease mechanisms remain poorly understood but genetic association studies may provide insights into disease pathology. We aimed to identify novel loci and to assess known loci for PAD over all, and in the context of smoking and diabetes. PAD was defined as atherosclerosis of the lower limbs and controls had no history of PAD. We combined summary statistics from 1000G imputed or CardioMetabochip genotypes, derived from subjects of European ancestry, in a fixed-effects inverse-variance weighted meta-analysis. We performed 3 separate meta-analyses: 1) All subjects (7,942 cases and 176,594 controls); 2) Diabetes stratified (3,228 cases/21,013 controls with diabetes and 4,714 cases /135,581 controls without diabetes); and 3) Smoking stratified (5,759 cases/85,854 controls from ever smokers and 1,782 cases /88,692 controls from never smokers). We tested loci for interaction with diabetes and smoking status. We identified novel loci, at genome-wide significance ($p < 5 \times 10^{-8}$), associated with PAD in all subjects (rs10850034, OR[95%CI]=1.18[1.11-1.24], $p=2 \times 10^{-11}$, near *HECTD4*) and in non-smokers (rs10894808, OR[95%CI]=1.43[1.31-1.55], $p=2 \times 10^{-8}$, $phet=2 \times 10^{-7}$) near *B3GAT1*. Rs10850034 is a proxy for rs11066301 ($r^2=0.97$), a locus associated with haemoglobin concentrations and haematocrit ($p=8 \times 10^{-12}$). Rs10894808 is an eQTL for *B3GAT1* in lung and pituitary tissues ($p < 3 \times 10^{-8}$). We detected associations for known PAD loci: rs10738610 (OR[95%CI]=1.16[1.12-1.21], $p=5 \times 10^{-16}$), in the 9p21.3 region; rs951266 (OR[95%CI]=1.13[1.09-1.17], $p=2 \times 10^{-8}$), an eQTL for *CHRNA5* in a range of tissues including tibial nerve and skeletal muscle ($p < 2 \times 10^{-13}$); and rs10455872 (OR[95%CI]=1.31[1.24-1.39], $p=2 \times 10^{-12}$), near *LPA*. We present the largest genetic study of PAD that identifies two novel loci, one specific to non-smokers with evidence of a smoking interaction, and strengthens the association of three known loci. Some loci are linked to gene expression, pointing to potential effector transcripts for experimental follow-up.

680T

Gene, gender, environment interactions between serum homocysteine and rs1801133 polymorphisms of methylenetetrahydrofolate reductase (MTHFR). *X. Xu[1], J.C. Lovejoy[1,2], A.T. Magis[1].* 1) Arivale Inc., 710 2nd Avenue, Suite 410, Seattle, WA, 98104; 2) Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109.

  Background: MTHFR is the rate-limiting enzyme in S-adenosyl-methionine biosynthesis and regeneration cycle, catalyzing the re-methylation of homocysteine to methionine. Hyperhomocysteinemia leads to endothelial cell injury and blood vessel inflammation, which greatly increases the risk of cardiovascular events. Genetic mutations in *MTHFR* are the most commonly known inherited risk factor for elevated homocysteine levels. The aim of this study was to assess the effect of MTHFR variants on homocysteine levels, and to explore potential interaction patterns. **Methods and Results:** 656 individuals (51.4% females, 94.2% whites, average age: 51.3 years, body mass index-BMI: 26.3 kg/m^2) were recruited from the general population in Seattle, WA, and were offered standard lifestyle intervention (i.e., exercise and nutritional counseling) beginning July 2015. Serum homocysteine level was measured up to 3 times (T1, T2, T3) over a 12-month period (average time interval: 6-month). Significant decreases were observed following supplementation with active methyltetrahydrafolate (MTHF) (T1: 10.8 umol/L, T2: 10.0 umol/L, T3: 9.8 umol/L, p<0.01). *MTHFR* rs1801133 C/T was detected by qPCR, with genotype frequency of TT 10.2% and CT 42.2% in the cohort. Generalized estimating equation accounting for within-individual correlations over time was used to examine the relationships and interactions, adjusted for age, BMI and lab vendor. No main effect of *MTHFR* genotype on homocysteine level was found. However, a significant gene-by-gender interaction was observed (p<0.01). In males, TT homozygotes had lower homocysteine level than CT+CC group (i.e., T1: 11.0 vs. 12.1 umol/L). Whereas, female TT homozygotes had significantly higher level than women carrying one or no T allele (i.e., T1: 10.1 vs. 9.6 umol/L). Further longitudinal analysis identified a gene-by-environment interaction pattern (rs1801133*time^2, p<0.01). Compared to CC group, T-carriers exhibited a faster decrease with MTHF supplementation from T1 to T2. Additionally, a gender-gene-environment interaction (rs1801133*gender*time) also occurred (p<0.01), with males carrying T-alleles exhibited the largest-trajectory of homocysteine change from T1 to T3. **Conclusion:** We observed significant gene, gender, environment interactions between *MTHFR* rs1801133 variation and homocysteine levels, which seems to affect trajectories in participants exposed to lifestyle interventions.

681F

Copy number variants in the nuclear genome and mitochondrial copy number. *N. Pankratz[1], J. Lane[1], D. Arking[2], R. Longchamps[2].* 1) Dept Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 2) Mckusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

  Introduction: Mitochondrial DNA copy number (mtDNA CN), which is a surrogate measure of mitochondrial function, declines with age and is associated with multiple traits and outcomes, including all-cause mortality. Heritability estimates for mtDNA range from 35-65%, indicating that these levels are under genetic control. In the present study, we have performed a genome-wide association study in the ARIC study to search for copy number variants (CNVs) that are associated with mtDNA CN. Methods: A raw estimate for mitochondrial copy number (mtDNA CN) was obtained from the median log R ratio (LRR) of mitochondrial markers on the Affymetrix 6.0 SNP array. To correct for batch effects, total amount of nuclear DNA, and sample quality, we performed a principal component analysis (PCA) of LRR on select autosomal markers and regressed out the most informative PCs. This method has been implemented in Genvisis (http://www.genvisis.org), and there is evidence that it may be more accurate than qPCR derived estimates. Estimates were further adjusted for age, sex, study center, and smoking status before being inverse normalized. Nuclear copy number variants (CNVs) were typed in the 9,312 samples of European Ancestry that had high quality intensity data (standard deviation of LRR < 0.35 and call rate > 0.96). CNVs were called on PC-corrected LRRs using PennCNV, filtered down to a high quality set of large (>200kb) variants using Genvisis, and underwent genome-wide segmental analyses using PLINK. Results: While no region reached genome-wide significance, the top two ranking genes (*TFB1M* and *MTHFD2*) are both known to localize to the mitochondria. A large duplication surrounding the *TFB1M* gene, which is known to be involved in mitochondrial biogenesis, led to a mtDNA CN estimate that was 3.0 SD above the population mean. Conclusions: The genetic mechanisms influencing mtDNA CN are not well known, however genome-wide analyses such as these may shed some light on the genetic factors influencing this trait. Since deleting or duplicating a gene will often lead to a powerful loss of function or gain of function, these events are likely rare due to natural selection, and therefore a genome-wide significant association might only be achieved through the meta-analysis of a hundred thousand samples.

682W

Exome-wide association analysis of cardiac structural traits in a large healthcare provider organization identifies genetic heterogeneity underlying left ventricular structure and overlapping genetic architecture with cardiomyopathy genes. *J.H. Chung[1], J. Robishaw[2], C. O'Dushlaine[1], R. Metpally[2], H.L. Kirchner[2], N.A. Manus[2], D. Hartzel[2], A. Lopez[1], J. Penn[1], J.D. Overton[1], J.G. Reid[1], A.R. Shuldiner[1], A. Baras[1], I. Borecki[1], O. Gottesman[1], V. Mehra[2], F.E. Dewey[1].* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

Understanding the genetic architecture of human cardiac structure can aid in the treatment of diseases related to abnormal structure and function, such as cardiomyopathy (CM). Previous studies of left ventricular (LV) structure have identified several candidate loci but were limited to common variants which may not be causal. Whole exome sequencing could reveal causal protein-coding variation. We performed an exome-wide association study (EXWAS) in a large clinical population with echocardiographic traits in the Regeneron Genetics Center-Geisinger Health System DiscovEHR study. We analyzed 6 LV traits from echocardiographic data captured in electronic health records in 10,639 individuals without ischemic heart disease. Association tests were performed for single variants with minor allele count ≥ 10 using linear mixed models adjusted for age, height, sex, systolic blood pressure, ancestry, and genetic relatedness. We also performed gene-burden association of rare (minor allele frequency [MAF] ≤ 0.01) predicted loss-of-function (pLOF) variants for all genes with ≥ 4 pLOF carriers. We found 6 variants in 5 genes and 11 variants in 11 genes with exome-wide (P ≤ 1E-7) and suggestive (P ≤ 1E-6) association, respectively. The top locus was rs75431703, a synonymous variant in *PLEKHG4B*, associated with LV mass (P = 6.7E-11, β = 106.5 ± 16.3g, MAF = 7E-4). We also found rs2234962, a missense variant in *BAG3* found to be protective for dilated cardiomyopathy (DCM), associated with LV internal diameter (LVIDd) (P = 3.3E-08, β = -0.064 ± 0.013cm, MAF = 0.2). Gene-burden analysis found 3 and 8 genes with gene-wise (P ≤ 1E-6) and suggestive (P ≤ 1E-5) associations, respectively. The top finding was for *SPON2*, an extracellular matrix protein, with interventricular septum thickness (P = 4.2E-8, β = 0.536 ± 0.098cm). pLOFs in TTN, a cardiac structural protein implicated in DCM, were associated with LVIDd (P = 1.0E-07, β = 0.26 ± 0.05cm). Overall, there was a high degree of genetic heterogeneity with 14/17 variants and 10/11 genes associated with a single trait, most of which were novel. In the largest EXWAS of cardiac structural traits to date, we found exome-wide significant associations with 6 variants and 3 genes with LV structure; two of which were in familial CM genes. These results suggest a high degree of genetic heterogeneity underlying cardiac structure in humans, and shared genetic architecture between rare coding variation in large populations and familial CM.

683T

Rare Variants in Fox-1 homolog A (*RBFOX1*) are protectively associated with blood pressure. *K.Y. He[1], H. Wang[1], B.E. Cade[2,3], P. Nandakumar[4], A. Giri[5], E.B. Ware[6,7], J. Haessler[8], J. Liang[1], J.A. Smith[7], N. Franceschini[9], C. Kooperberg[8], T.L. Edwards[5], S.L.R. Kardia[7], X. Lin[10], A. Chakravarti[4], S. Redline[2,3,11], X. Zhu[1].* 1) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 3) Division of Sleep Medicine, Harvard Medical School, Boston, MA; 4) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 5) Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 6) Biosocial Methods Collaborative, Institute for Social Research, University of Michigan, Ann Arbor, MI; 7) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 8) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 9) Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC; 10) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 11) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA.

Hypertension is a risk factor for cardiovascular diseases and the most important risk factor for stroke. Family studies suggest that hypertension related traits are heritable. Many genome-wide association studies (GWAS) have identified common blood pressure (BP) variants; however, each variant only accounts for a small amount of variation in BP. In particular, many BP variants identified by large, powerful genome-wide association studies do not overlap with the linkage evidence observed from family studies. In this study, we performed linkage analysis using 517 individuals in 130 European families from the Cleveland Family Study (CFS) who have been genotyped by Illumina OmniExpress Exome array. The largest linkage peak was observed on chromosome 16p13 (MLOD=2.81) for systolic blood pressure (SBP). Follow-up conditional linkage and association analyses in the linkage region identified multiple rare, coding variants in *RBFOX1* to be protectively associated with SBP. In a 17-member CFS family, carriers of a missense variant rs149974858 are normotensive even though their BMIs were extremely high (average BMI=60). Gene-based association of rare variants using SKAT-O showed significant association with SBP (p-value=0.00403) and DBP (p-value=0.0258) in the discovery data set and the association was further observed in large independent European replication cohorts (N=57,234, p-value=0.013 for SBP, 0.0023 for PP). These rare exonic variants are not polymorphic in African ancestry populations and they may contribute to the hypertension disparity between African Americans and European Americans. *RBFOX1* is expressed in multiple brain tissues, as well as the atrial appendage and left ventricle in the heart, and skeletal muscle tissues, which are highly related to blood pressure. Our study showed that associations of rare, functional variants could be detected using family information with blood pressure traits in the discovery family data.

684F

Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian randomization study. *S. Lindstroem[1], M. Germain[2,3], M. Crous-Bou[4,5], E. Smith[6], P.E. Morange[7], A. van Hylckama Vlieg[8], H.G. de Haan[8], D. Chasman[9,10], P. Ridker[9,10], J. Brody[11], M. de Andrade[12], J.A. Heit[13], I. DeVivo[4,5], F. Grodstein[4], N.L. Smith[1,14,15], D. Tregouet[2,3], C. Kabrhel[10,16], INVENT Consortium.* 1) University of Washington, Seattle, WA; 2) Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France; 3) ICAN Institute for Cardiometabolism and Nutrition, Paris, France; 4) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 5) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA; 6) Department of Pediatrics, Rady Children's Hospital, University of California, San Diego, La Jolla, CA, USA; 7) INSERM, UMR_S1062, Nutrition Obesity and Risk of Thrombosis, Aix-Marseille University, Marseille, France; 8) Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 9) Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA USA; 10) Harvard Medical School, Boston MA, USA; 11) Department of Medicine, University of Washington, Seattle WA, USA; 12) Biomedical Statistics and Informatics, Mayo Clinic and College of Medicine, Rochester, MN, USA; 13) Division of Cardiovascular Diseases, Mayo Clinic and College of Medicine, Rochester, MN, USA; 14) Group Health Research Institute, Group Health Cooperative, Seattle WA USA; 15) Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle WA USA; 16) Department of Emergency Medicine, Massachusetts General Hospital, Boston MA, USA.

Venous thromboembolism (VTE) is the collective term for deep vein thrombosis (DVT) and its life-threatening complication pulmonary embolism (PE). Between 300,000-600,000 Americans are diagnosed with VTE every year, and it has been suggested that 60,000-100,000 Americans die of VTE annually. Observational studies have shown an association between obesity and venous thromboembolism (VTE) but it is not known if observed associations are causal, due to reverse causation or confounding bias. We leveraged information from two recent genome-wide association studies (GWAS) of childhood and adult body mass index (BMI) to explore the causal relationship between childhood and adult obesity and adult VTE. Using a two-sample Mendelian Randomization (MR) approach, we assessed the associations of 15 childhood BMI single nucleotide polymorphisms (SNPs) and 95 adult BMI SNPs in a VTE study comprising 7,507 cases and 52,632 controls. We observed a significant association between genetically predicted childhood BMI as defined by a composite risk score and VTE (OR: 1.30, 95% CI: 1.03-1.63 per s.d. increase in childhood BMI, p=0.02). The overall association between genetically predicted adult BMI and VTE was 1.59 per s.d. increase in adult BMI (95% CI: 1.30-1.93, p=5.8x10⁻⁶). We reran the analysis including only the 75 SNPs found to be genome-wide significant in European ancestry individuals only and observed a similar association (OR=1.58, 95% CI: 1.28-1.95 per s.d. increase in BMI, p=2.02x10⁻⁵). In sensitivity analysis, we removed 10 SNPs that have been associated with both childhood and adult BMI, leaving five SNPs for childhood BMI and 90 SNPs for adult BMI. After removal of the overlapping SNPs, the childhood BMI-VTE association was no longer apparent (OR= 1.17, 95% CI: 0.74-1.87 per s.d. increase in BMI, p=0.50), whereas the results for adult BMI remained largely unchanged (OR=1.54, 95% CI: 1.21-1.96 per s.d. increase in BMI, p=0.0005). Thus, our overall results supporting a causal link between childhood BMI and VTE should be interpreted with caution, but it is possible that the lack of association observed when removing adult BMI SNPs is due to low statistical power. In summary, our results show evidence for a causal association between higher BMI and VTE, lending support to previous observational studies.

685W

Defining updated risk profiles by linking contemporary cohorts with historical cohorts based on shared genetics. *J.D. Mosley[1], S.L. Van Driest[1,2], Q.S. Wells[1], C.M. Shaffer[1], T.L. Edwards[3], L. Bastarache[4], C.A. McCarty[5], W. Thompson[6], C.G. Chute[7], G.P. Jarvik[8], D.R. Crosslin[9], E.B. Larson[10], I.J. Kullo[11], J.A. Pacheco[12], P.L. Peissig[13], M.H. Brilliant[14], J.G. Linneman[13], J.C. Denny[1,4], D.M. Roden[1,4,15].* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University, Nashville, TN, USA; 3) Vanderbilt Epidemiology Center, Vanderbilt University Nashville, TN, USA; 4) Biomedical Informatics, Vanderbilt University, Nashville, TN, USA; 5) Essentia Institute of Rural Health, Duluth, MN, USA; 6) Center for Biomedical Research Informatics, North Shore University Health System, Evanston, IL; 7) Schools of Medicine, Public Health, and Nursing, Johns Hopkins University, Baltimore, MD, USA; 8) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA, USA; 9) Departments of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA; 10) Group Health Research Institute, Seattle, WA, USA; 11) Division of Cardiovascular Diseases, Mayo Clinic, Rochester MN, USA; 12) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 13) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI, USA; 14) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI, USA; 15) Department of Pharmacology, Vanderbilt University, Nashville, TN, USA.

Background: Longitudinal studies have identified and validated many risk factors for chronic diseases, and management of these factors has led to declines in morbidity and mortality. Limitations of longitudinal studies include long latencies and high costs. We hypothesized that we could use genetic correlations, which quantitate the shared genetic architectures of phenotype pairs, to efficiently define updated risk profiles for a contemporary cohort using risk factors measured in a historical prospective study. **Methods:** We used 37 baseline risk factors measured in the Atherosclerosis Risk in Communities (ARIC) study (n=7,716 European ancestry subjects) and cases and controls for type 2 diabetes (T2D) (1,321 cases and 6,395 controls) and ischemic heart disease (IHD) (5,114 cases and 8,789 controls) from a contemporary electronic health record (EHR) data set (n=19,093 EAs). We measured pairwise genetic correlations (rG) between the ARIC and EHR phenotypes using linear mixed models and compared these rG estimates to hazard ratio estimates for longitudinal outcomes of incident T2D and IHD within the ARIC population. **Results:** For T2D, the rG estimates were strongly correlated with hazard ratio estimates (Pearson's correlation [r]=0.8), indicating that the EHR and ARIC T2D phenotypes have similar risk profiles. The linear correlation between these measures was weaker when comparing the EHR to ARIC IHD phenotype (r=0.6), suggesting that the phenotypes have differing risk profiles. The EHR IHD phenotype was most strongly genetically correlated with ARIC metabolic risk factors including total-to-HDL cholesterol ratio (rG=-0.44, p=0.005), HDL (rG=-0.48, p=0.005), systolic blood pressure (rG=0.44, p=0.02) and triglycerides (rG=0.38, p=0.02). We measured genetic correlations between these 4 ARIC IHD risk factors and 158 EHR-derived diagnoses, and found they were significantly genetically correlated with T2D and atherosclerotic and hypertension-related disease. **Conclusions:** Genetic correlations measured across populations were linearly correlated with longitudinal hazard ratio estimates for T2D and IHD. For IHD, contemporary treatment and prevention efforts in the EHR population should target hypertensive and metabolic disease. Linking outcomes from EHR data sets with biomarkers from epidemiological cohorts using genetic correlations will facilitate the identification of risk factors without decades of follow-up.

686T

Missense variant in *GDF15* explains 3.7-5.3% of the variance in GDF-15 in two large CV outcome trials. *N. Eriksson[1], A. Johansson[2], B.J. Barratt[3], R.C. Becker[4], A. Budaj[5], E. Hagström[1,6], C. Held[1,6], A. Himmelmann[7], H.A. Katus[8], W. Koenig[9,10,11], A. Siegbahn[1,12], P.G. Steg[13,14,15,16], S.K. James[1,6], T. Axelsson[17], A-C. Syvänen[17], R.A.H. Stewart[18], A. Akerblom[1,6], R.F. Storey[19], D. Waterworth[20], H.D. White[18], L. Wallentin[1,6] on behalf of the PLATO and STABILITY Investigators.* 1) Uppsala Clinical Research Center, Uppsala, Uppsala, Sweden; 2) Department of Immunoloy, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 3) AstraZeneca R&D, Alderley Park, Cheshire, UK; 4) Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; 5) Postgraduate Medical School, Grochowski Hospital, Warsaw, Poland; 6) Department of Medical Sciences, Cardiology, Uppsala University, Uppsala, Sweden; 7) AstraZeneca Research and Development, Mölndal, Sweden; 8) Medizinishe Klinik, Universitätsklinikum Heidelberg, Heidelberg, Germany; 9) Department of Internal Medicine II-Cardiology, University of Ulm Medical Center, Ulm, Germany; 10) Deutsches Herzzentrum München, Technische Universität München, Munich, Germany; 11) DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany; 12) Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden; 13) INSERM-Unité 1148, Paris, France; 14) Assistance Publique-Hôpitaux de Paris, Département Hospitalo-Universitaire FIRE, Hôpital Bichat, Paris, France; 15) Université Paris-Diderot, Sorbonne-Paris Cité, Paris, France; 16) NHLI Imperial College, ICMS, Royal Brompton Hospital, London, UK; 17) Department of Medical Sciences, Molecular medicine, Uppsala University, Uppsala, Sweden; 18) Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand; 19) Department of Cardiovascular Science, University of Sheffield, Sheffield, UK; 20) Genetics, GlaxoSmithKline, King of Prussia, PA, USA.

Background: Growth differentiation factor 15 (GDF-15) is an independent prognostic biomarker for cardiovascular events and major bleeding during antithrombotic treatment. Whether there is a causal association between GDF-15 and clinical outcomes is still unknown. Genetic markers located in *GDF15* are associated with altered circulating GDF-15 concentrations. In this meta-analysis, we assess the genetic effects on GDF-15 in 19,910 patients from two large cardiovascular outcome trials. **Methods:** GWAS was performed in 10,013 patients from the PLATelet inhibition and patient Outcomes study (PLATO, ClinicalTrials.gov number, NCT00391872) and 10,788 patients from the Stabilization of Atherosclerotic Plaque by Initiation of Darapladib Therapy trial (STABILITY, ClinicalTrials.gov number, NCT00799903). Genotype data was generated on the Illumina HumanOmni2.5 and OmniExpressExome arrays. In total 9,448 patients from PLATO and 10,462 patients from STABILITY passed QC and had GDF-15 measurements available at baseline. Imputation was performed using 1000 genomes as reference. GWAS analyses of GDF-15 were performed study wise using PLINK, adjusted for established cardiovascular risk factors and genetic principal components and meta-analyzed using PLINK. Analyses were performed sequentially adjusting for each genome-wide significant variant until no signals were left. **Results:** Overall the median (interquartile range) level of GDF-15 was 1528 (1139 – 2144) in PLATO and 1248 (909 – 1814) ng/L in STABILITY. Within each study we identified three independent genome-wide significant SNPs associated with log(GDF-15). All were within or close to the *GDF15* gene on chromosome 19 with the top hit in PLATO being rs17725099 (PLATO beta=0.161, p=1.47×10^{-107}, STABILITY beta=0.187, p=3.06×10^{-158}). Combining the GWAS results in a meta-analysis pinpointed the missense variant rs1058587 (beta=0.179, p =2.36×10^{-270}) in *GDF15* as the driving variant of the top signaling cluster of SNPs. Follow-up analyses adjusting for rs1058587 gave additional signals. The top hit SNP rs1058587 explained 3.7% and 5.3% of the variation in log(GDF-15) and the increase in median GDF-15 from the major genotype C/C to the minor G/G was 36% (1406 to 1918 ng/L) and 48% (1121 to 1663 ng/L) in PLATO and STABILITY, respectively. **Conclusion:** We identified a missense SNP affecting GDF-15 concentrations. The SNP explained up to 5.3% of the variance in GDF-15 and may be suitable for future Mendelian randomization analyses.

687F

Functional enrichment analysis of genome-wide association study identifies novel pathways influencing blood pressure in African Americans. *J. Liang, COGENT-BP Consortium of African Ancestry.* Wolstein Research BLDG, Rm 1306, Case Western Reserve University, Cleveland, OH.

Blood pressure (BP) is a highly heritable trait influenced by many biological pathways and functional categories encoded in the genome. Individuals of African ancestry share a disproportionate burden of hypertension and its complications compare to other ethnic groups in the U.S. To better understand the biological pathways and functional categories of the genome, we performed meta-analyses of BP genome-wide association studies (GWAS) of 31,968 individuals of African ancestry from 21 studies using high density imputed genotypes from the 1000 Genomes Project for gene discovery. We further explored genetic contributions of BP traits with respect to the functions involved. We used the Meta-Analysis Gene-Set Enrichment variaNT Association (MAGENTA) software on the GWAS data from systolic BP (SBP), diastolic BP (DBP), pulse pressure (PP) and hypertension (HTN) to identify GSK3 pathway ($P_{GWAS-GSEA}$ =1.2×10^{-3} for SBP), Th1/Th2 immune pathway ($P_{GWAS-GSEA}$ =1.6×10^{-3} for DBP) and SHH pathway ($P_{GWAS-GSEA}$ =2.1×10^{-3} for HTN) using the BIOCARTA database, and membrane depolarization ($P_{GWAS-GSEA}$ =6.0×10^{-4} for DBP), regulation of BP ($P_{GWAS-GSEA}$ =6.0×10^{-4} for SBP) and histone deacetylation ($P_{GWAS-GSEA}$ =7.0×10^{-4} for PP) using the GOTERM database, as the most significantly enriched biological pathways. Functional annotation and cell-type group enrichment analysis were performed using a stratified LD score regression method by partitioning heritability from GWAS summary statistics of the four BP traits while accounting for linkage disequilibrium among markers. For functional categories enrichment, super enhancer (P_{Enrich} = 6.6×10^{-6} for DBP), enhancer (P_{Enrich} = 3.8×10^{-4} for SBP) and H3K27ac (P_{Enrich} = 9.0×10^{-4} for HTN) are significantly enriched passing the multiple test threshold. For cell types enrichment analysis, immune (P_{Enrich} = 1.4×10^{-9} for DBP), kidney (P_{Enrich} = 2.4×10^{-6} for DBP) and cardiovascular (P_{Enrich} = 2.3×10^{-5} for SBP) are the three most significantly enriched systems. Our findings from analyzing pathway and functional elements enrichment using African American BP GWASs data suggest novel pathways for BP regulation in a highly susceptible population to hypertension.

688W

Meta-analysis of genome-wide association studies in different ethnicities identifies new loci regulating Factor VII, Factor VIII and von Willebrand factor. *M. Sabater-Lleal[1], P.S. de Vries[2], J. Marten[3], J.E. Huffman[4], C.J. O'Donnell[4], N.L. Smith[5,6], CHARGE Hemostatic Working Group.* 1) Cardiovascular Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 3) MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh, United Kingdom; 4) National Heart, Lung, and Blood Institute, Framingham, MA; The Framingham Heart Study, NHLBI Population Sciences Branch, Framingham, MA; 5) Seattle Epidemiologic Research and Information Center, Veterans Affairs Office of Research and Development, Seattle, WA; 6) University of Washington, Seattle, WA.

  Background: Coagulation Factors VII (FVII), VIII (FVIII) and von Willebrand factor (vWF) are relevant proteins in hemostasis that have been strongly associated to risk of thrombosis and hemorrhagic disorders. In the present study we aim to identify novel loci regulating plasma levels of FVII, FVIII and vWF by meta-analyzing GWAS results from up to 46,232 individuals of different ethnicities.**Methods:** All cohorts performed linear regression analysis according to an additive model of inheritance between genome-wide 1000 Genomes imputed SNPs (about 14 million markers after filtering) and phenotype levels, with adjustments for age, sex, and ancestry-informative principal components. Case/control status, family structure and other cohort-specific covariates were added when necessary. Individuals of European, African, Asian, and Hispanic ancestry were included, and separate analyses were run for each ethnicity.**Results:** We identified significant associations with vWF at loci near *ABO, vWF, STXBP5, STAB2, SCARA5, TCN2, FCHO2, TAB1, HLA-C, STX2, CLEC4M, SYNGR1, ST3GAL4, SLC39A8, DAB2IP, PXK, GIMAP7, C2CD4B,* and *OR13C5,* the first 9 representing overlapping loci between the FVIII and vWF phenotypes. Among these, *FCHO2, SYNGR1/TAB1, ST3GAL4/KIRREL3, BANK1/SLC39A8* (near *NFKB1), DAB2IP, HLA* region*, PXK, GIMAP* gene cluster, and *C2CD4B* represent novel findings. Three novel loci, located near *SOX17* gene and two non-coding RNA respectively, were uniquely associated with FVIII. For FVII levels, loci containing *F7, PROC, GCKR, MS4A5A* and *ADH4* represented replications of previously described loci, whereas 2 loci were novel, near *REEP3* and *TBL2.***Conclusions:** Our study represents the largest meta-analysis to unravel genetic determinants of FVII, FVIII and vWF and identify novel genetic determinants regulating plasma levels of these proteins. Our preliminary results show limited relevance of rare variation in the regulation of these phenotypes. Functional characterization of the novel loci regulating these proteins might reveal interesting new therapeutic targets for thrombosis and bleeding disorders.

689T

Genotype-phenotype in Marfan syndrome patients with causative mutations from the 6 cysteines in the 43 cbEGF-like domains in fibrillin-1 (FBN1). *J.A. Aragon-Martin[1], D. Akinfenwa[1], Y. Isekame[1], C.S. Gracias[1], Y.B.A. Wan[1], P. Mazadza[1], T. Thayalacumar[1], G. Arno[2], A. Chandra[3], D.G. Charteris[4], A. Saggar[5], M. Jahangiri[6], M.T. Tome Esteban[1], A.H. Child[1].* 1) Cardiovascular Sciences Research Centre, St George's University of London, London, England, United Kingdom; 2) Institute of Ophthalmology, University College London, London, United Kingdom; 3) Southend University Hospital, Southend on Sea, United Kingdom; 4) Vitreoretinal Surgery, Moorfields Eye Hospital, London, United Kingdom; 5) Clinical Developmental Sciences, St George's University of London, London, United Kingdom; 6) Department of Cardiothoracic Surgery, St. George's Healthcare NHS Trust, Cranmer Terrace, London SW17 0RE, United Kingdom.

  Background: Marfan syndrome (MFS) is an autosomal dominant connective tissue condition (OMIM #154700) that is phenotypically heterogeneous [cardiovascular (Thoracic Aortic Aneurysm – TAA), ocular (Ectopia Lentis – EL), skeletal (Scoliosis)] and is reported to be due to mutations in the *FBN1* gene. Through out the FBN1 protein there are 43 cbEGF-like domains, which contains 6 highly conserved cysteine residues that form disulphide bonds between themselves: C1-C3, C2-C4 and C5-C6. Calcium ions play a key role in the stability of these domains and increase resistance to proteolysis degradation. Methods: Patients (n=101) with mutations in one of the 6 cysteine residues in any of the 43 cbEGF-like domains were entered in the study. Clinical data was obtained from the Marfan clinic at St George's Hospital and the genotypes from the *FBN1* mutation-screening database in Sonalee Laboratory. A genotype-phenotype database was created with this data. Results: This study found that MFS phenotypes may arise due to mutations in any of the 6 cysteine residues: – C1 showed Skeletal (82%), EL (65%), TAA (47%); – C2 showed Skeletal (75%), TAA (69%), EL (38%); – C3 showed TAA (57%), EL (50%), Skeletal (50%); – C4 showed TAA & EL (56%), Skeletal (44%); – C5 showed TAA (79%), Skeletal (71%), EL (57%); – C6 showed Skeletal (94%), EL (83%), TAA (61%); – C1^C3 showed Skeletal (68%), EL (58%), TAA (52%); – C2^C4 showed Skeletal & TAA (64%), EL (44%); – C5^C6 showed Skeletal (84%), EL (72%), TAA (69%); – C1_C6 showed Skeletal (89%), EL (74%), TAA (54%). Data showed that phenotypes found in cysteine residues of the cbEGF-like domains were Skeletal (73%), TAA (61%), EL (59%). Females were more likely to have Skeletal (58%) or EL (64%) phenotypes than men TAA (52%) phenotypes. Conclusions: Not counting skeletal phenotypes, the study shows a strong genotype-phenotype relationship with mutations in C1 & C6 with EL and C2 & C5 with TAA. When combining data from C1^C3 + C5^C6 the strongest association was with EL & C2^C4 with TAA. This maybe due to the proximity of C1^C3 + C5^C6 to the calcium binding region and it could have less influence in a TAA phenotype. The different phenotypes arising from mutations in the cysteine could be due to a random assembly of these bonds in the protein. Instability of the calcium binding region due to mutations in cysteine from cbEGF-like domains are more likely to express an EL than a TAA phenotype, but the risk of TAA is till very high.

690F

Genetic correlation of cardiovascular and metabolic traits with anthropometric measures in the Busselton Health Study. *G. Cadby[1], P.E. Melton[1], N.S. McCarthy[1], J. Hui[2,3], J. Beilby[2,3], A.W. Musk[2,4,6], A.L. James[2,5,6], J. Hung[6,7], J. Blangero[8], E.K. Moses[1].* 1) The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and The University of Western Australia, Western Australia; 2) Busselton Population Medical Research Institute Inc, Western Australia; 3) PathWest Laboratory Medicine WA, Western Australia; 4) Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Western Australia; 5) Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Western Australia; 6) School of Medicine and Pharmacology, The University of Western Australia, Western Australia; 7) Department of Cardiovascular Medicine, Sir Charles Gairdner Hospital, Western Australia; 8) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley, Texas.

Anthropometric measures genetically correlated with cardiovascular (CVD) and metabolic traits may provide insight into their pathways and aetiologies. Using subjects from the Busselton Health Study (n=4671), we employed empirically derived identity by descent (IBD) measures to estimate the genetic correlation between CVD and metabolic traits (HDL-C, LDL-C, triglycerides, insulin, total cholesterol, coronary heart disease and diabetes) and anthropometric measures (BMI, waist-hip ratio, weight, and abdominal, arm, thigh, calf and neck circumferences). IBD estimates were derived from genome-wide data using LDAK and genetic correlations were calculated in GCTA. Transformed residuals (after adjustment by age and sex) were used in all analyses. For significantly associated genetic correlations, anthropometric measures were also adjusted by BMI to determine whether the correlation was independent of obesity. We used the false discovery rate to correct for multiple testing, with q≤0.05 considered statistically significant. Diabetes was genetically correlated with BMI (r_g=0.59) and abdominal circumference (r_g=0.63). Coronary heart disease was genetically correlated only with arm circumference (r_g=0.55). HDL, triglycerides and insulin were genetically correlated with BMI (HDL r_g=-0.29; triglycerides r_g=0.28; insulin r_g=0.57), waist-hip ratio (HDL r_g=-0.38; triglycerides r_g=0.42; insulin r_g=0.57) and circumference of the abdomen (HDL r_g=-0.33; triglycerides r_g=0.42; insulin r_g=0.77) and arm (HDL r_g=-0.29; triglycerides r_g=0.31; insulin r_g=0.57). LDL and total cholesterol were not genetically correlated with any anthropometric measure. After adjustment for BMI, three genetic correlations remained statistically significant: waist-hip ratio and triglycerides (r_g=0.26); waist-hip ratio and HDL (r_g=-0.23); abdominal circumference and insulin (r_g=0.26). These results reveal that the genetic correlation between CVD and metabolic traits and anthropometric measures are largely due to the role of obesity. However, central adiposity, rather than obesity, remains genetically correlated with these traits, indicating possible pleiotropic genes that affect both central adiposity and CVD and metabolic traits.

691W

Pharmacogenetics of time to Acute Coronary Syndrome recurrence (PhACS): A UK cohort study. *P. Yin[1], A. Jorgensen[1], A. Morris[1], R. Turner[2], R. Fitzgerald[2], R. Stables[3], A. Hanson[2], M. Pirmohamed[2].* 1) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 2) Department of Molecular & Clinical Pharmacology, University of Liverpool, United Kingdom; 3) Liverpool Heart and Chest Hospital, United Kingdom.

Coronary heart disease is a major global public health burden. It is a recurrent disease that is treated by multiple drugs, including aspirin, clopidogrel, statins, beta-blockers, ACE inhibitors and aldosterone antagonist. There is significant inter-patient variability in the response to cardiovascular drugs, which may have a heritable basis. We have therefore undertaken a genome-wide association study to identify loci associated with response to cardiovascular drugs in 1470 patients recruited to a UK prospective pharmacogenetic study of acute coronary syndrome (ACS). Index hospital admission was defined non-ST elevation acute coronary syndrome. Patients were treated with a range of cardiovascular drugs and followed up prospectively for up to 5 years from hospital discharge. Despite treatment, approximately 14% of patients had recurrence of ACS during the following up, defined by myocardial infarction (MI), stroke or cardiovascular death. We began by considering clinical risk factors for response to cardiovascular drugs (time to recurrence of ACS) in a Cox proportional hazard regression framework. Several risk factors were nominally related (p<0.05): age (p=1.3x10^{-14}), body mass index (p=0.015), prior MI (p=2.1 x10^{-8}), ACE inhibitor use pre-admission (p=0.0049), and aldosterone antagonist use (p=0.0058). Patients were genotyped using the Illumina Omni Express array. After quality control, the genotype scaffold was imputed up to the 1000 Genomes Phase I reference panel (all ancestries, March 2012 release). We tested for the association of SNPs with time to event under an additive dosage model after adjusting for the clinical factors identified above and principal components to account for population structure. We identified a genotyped intronic variant (rs2660044) in the *CTNNA3* gene with strong evidence of association: minor allele frequency 0.15, hazard ratio (95% confidence interval) 2.47 (1.67-3.65), p=3.4 x10^{-7}. This gene encodes a protein that plays a role in cell-cell adhesion in muscle cells. Mutations in this gene are associated with arrhythmogenic right ventricular dysplasia, familial 13. Our study highlights that variants mapping to *CTNNA3* are associated with response to cardiovascular drugs in ACS recurrence time. .

692T

Genetic factors predicting Lp-PLA₂ activity and their association with cardiovascular-related mortality and morbidity. *M.K. Siddiqui[1], G. Kennedy[2], F. Carr[1], A. Doney[1], C.N.A. Palmer[1].* 1) Molecular and Clinical Medicine, University of Dundee, Dundee, United Kingdom; 2) Immunoassay Biomarker Core Laboratory, Ninewells Hospital & Medical School Dundee, United Kingdom.

Lipoprotein-associated phospholipase A₂ (Lp-PLA₂), is a vascular-specific, proinflammatory enzyme that is known to increase the risk of CVD events and stroke. Chu *et al.* conducted a meta-analysis of genome-wide association studies across the CHARGE consortium and JUPITER study to identify genetic variants associated with the mass and activity of the enzyme. We sought to replicate the findings in a cohort of 7,000 Scottish Caucasian individuals in the GoDARTS study for whom Lp-PLA₂ activity levels had been measured. We replicate one of five loci associated with Lp-PLA₂ mass and eight of nine loci associated with Lp-PLA₂ activity, with the ninth variant showing borderline significance ($p = 0.07$). Notably, the variants in *PLA2G7* (rs1362931, $\beta = 3.24$, $P = 4 \times 10^{-5}$) and *APOC1-APOE* (rs7412, $\beta = -7.4$, $P = 2 \times 10^{-10}$ and rs445925, $\beta = -5.0$, $P = 1.6 \times 10^{-6}$) were strongly associated with Lp-PLA₂ activity in models adjusted for BMI, age, sex and smoking status, similar to the original study. In context of Lp-PLA₂'s role in cardiovascular health, we decided to examine the impact of these variants on related mortality and morbidity. We classified statin users who were subsequently hospitalized due to an ischemic event or coronary artery disease or had the same events recorded as their cause of death, as having statin failure. We find the variant in *PLA2G7* is associated with this outcome, both cross-sectionally and longitudinally. In the highest quartile of Lp-PLA₂ levels homozygous carriers of the variant (T/T) had 1.7 times the hazards of having statin failure compared to homozygous carriers of the ancestral allele (p value = 0.035).

693F

Fine-scale genetic population structure and genetic risk for coronary artery disease in Finland. *S. Kerminen[1], I. Surakka[1], A-P. Sarin[1], E. Tikkanen[1], G. Hellenthal[2], A. Havulinna[3], M. Perola[3], V. Salomaa[3], S. Ripatti[1], M. Pirinen[1].* 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) University College London, London, UK; 3) National Institute of Health and Welfare, Helsinki, Finland.

Finland is known for its relatively small founder population and strong genetic isolation, which have made it a valuable source for genetic studies of diseases and traits. Previous studies of Finnish population structure have reported genetic differences between South-West (SW) and North-East (NE) Finland and a similar division can be seen, e.g., in the incidence of coronary artery disease (CAD). Here we refine the current knowledge of the genetic structure in Finland and study whether it explains regional differences in the incidence of CAD. We identify Finnish population structure by applying haplotype-based methods ChromoPainter and fineSTRUCTURE to 2376 such individuals from the FINRISK Study survey of 1997 whose both parents were born close to each other. The high-level population structure reveals where exactly the previously undetermined "genetic border" between SW and NE Finland lies and shows strong similarities both to the known history of inhabitation of Finland and to the geographic incidence of CAD. At a lower-level, we identify several geographically clustered genetic populations which correspond well to detailed dialectal regions. To study the possible impact of the genetic structure to the regional differences in CAD, we generate two genetic risk scores: (1) lead-SNP score using 153 independent lead SNPs and (2) polygenic score using 53,000 LD-pruned SNPs with p-value below 0.05 from CARDIoGRAMplusC4D Consortium. Among the genetically identified subpopulations, our lead-SNP score for CAD yields an odds ratio of 1.11 ($p = 0.025$) between NE region of Kainuu and SW Finland possible explaining some of Kainuu's 1.8 fold incidence in CAD compared to SW Finland. The geographic distribution of our polygenic score also clearly reflects the observed incidence: the highest genetic risk is found from East and the lowest from the SW. The results demonstrate that the haplotype-based methods can detect unprecedented details of population structure in Finland reflecting population history and dialect patterns. The similarity between distributions of genetic risk scores and incidence of CAD suggests that some of the regional differences in risk of CAD are genetic.

694W

Using big data to interpret genomes with medical records for therapeutic target discovery. *R. Chen.* Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Millions of individuals have been sequenced or genotyped and linked with medical records, providing an exciting opportunity for therapeutic target discovery. My lab has been using a resilience approach to learn from success instead of failure, and analyzed hundreds of thousands of genomes with electronic medical records to drive target discovery for childhood Mendelian diseases, Alzheimer's disease, and cardiovascular traits. I will describe how we analyzed 589,305 genomes and identified 13 genetic super heroes who carry fully penetrant childhood Mendelian diseases but are healthy in our recent resilience project, and how we decoded individuals who carry APOE e4/e4 risk variants but are resilient for Alzheimer's disease. I will further describe how we analyzed individuals who have a loss of gene but carry favorable cardiovascular traits, identified potential therapeutic targets to lower fasting glucose levels, and validated the effects in mice.

695T

A reference-agnostic and rapidly queryable NGS read data format allows for flexible analysis at scale. *N. Shekar[1], W.J. Salerno[2], A.C. English[2], A. Mangubat[1], J. Bruestle[1], E. Boerwinkle[2,3], R.A. Gibbs[2].* 1) Spiral Genetics, Seattle, WA; 2) Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX; 3) University of Texas Health Science Center, Human Genetics Center, Houston, TX.

In identifying the complement of genetic variants that are associated with complex disease, larger sample sizes increase power. Studies such as the Alzheimer's Disease Sequencing Project and the CHARGE Consortium where samples are collected from a range of centers show heterogeneous data, requiring informatics that can additively scale to thousands of samples and analytics that go beyond identifying small variants in NGS data. At scale, the challenge of evaluating SNPs, indels and SVs becomes the "N+1" problem of incrementally adding samples without having to perpetually reevaluate petabytes of population read data stored in BAM files. The Biograph™ Analysis Format (BAF) is a method of indexing NGS data that extends the Burrows Wheeler Transform to allow for multiple paths, effectively creating a read overlap graph of the data. A BAF of HiSeq X 30x WGS data is 20 Gb, 80% smaller than the original BAM. Generated from the BAM in 14 hours, the BAF can be queried up to 200,000 times a second. Multiple BAFs can be combined, resulting in a file size of approximately 10GB per individual. With multi-individual BAFs, query time grows less than linearly with the number of individuals. For example, with 30,000 putative SV sites to be queried, SV-typing these sites across 10,000 HiSeq X WGS samples in BioGraph Analysis Format would require less than 100 TB of storage (for all the read data), 16 CPU hours, and 10 minutes (using 100 machines). Additionally, the data are reference-agnostic, so variants can be called against any reference or against the read graph of any other set of individuals, dramatically reducing the time for data harmonization. Further, information is divided such that the "read overlap graph" created from all the individuals is separate from the information indicating that path through the graph for each individual. This allows a search for a particular variation of interest directly from the read data remotely and rapidly, without the opportunity to reveal the exact individual(s) from that the variant originates. Because the data are essentially a read overlap graph, it is possible to accurately characterize SVs by traversing the graph from a particular location or search for a particular sequence associated with the SV. So, fast querying of small files with reasonable compute requirements provides an N+1 solution for SNPs, indels and SVs. We describe how the BAF API allows users to construct specific queries for a range of applications.

696F

Rare variants in the non-coding sequence and plasma lipids: An analysis of deep, whole genome sequence data in 2,255 Estonians. *S. Zekavat[1,2,3], P. Natarajan[1,2,4], A. Ganna[1,5,6,7], J. Ernst[8,9,10,11], A. Metspalu[12], B. Neale[1,5,6], T. Esko[1,12], S. Kathiresan[1,2,4].* 1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 2) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, 02114, USA; 3) Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA; 4) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm 171 77, Sweden; 8) Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA; 9) Departments of Biological Chemistry and Computer Science, University of California, Los Angeles, CA, USA; 10) Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, CA, USA; 11) Jonsson Comprehensive Cancer Center, University of California, Los Angeles 12 Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA; 12) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.

Plasma lipids are heritable risk factors for coronary heart disease. Common DNA sequence variants in non-coding regions (analyzed individually) and rare mutations in coding regions (analyzed after collapsing by gene) are known to contribute to variability in plasma lipids. High-coverage whole genome sequences (WGS) allow for detection of rare variants in non-coding regions. The analysis of such variants poses considerable challenges due to issues of power and insufficient knowledge of non-coding functional interpretation with WGS. In 2255 Estonians whole-genome sequenced with mean coverage 30X, we evaluated consequences of individual variants and rare (MAF <1%) variants aggregated within functional units. We defined three functional units: 1) protein-coding regions of a gene; 2) annotated enhancer regions of HepG2 liver cells linked to genes by gene expression data (Ernst et al. Nature 2011); and 3) annotated promoter and enhancer regions of HepG2 liver cells within 20kb of the transcription start site (TSS) at DNAse hypersensitive sites (from Roadmap and ENCODE). We limited our analyses to 301 genes previously associated with plasma lipids by large-scale GWAS. Single variant analyses recapitulated known associations: *APOE* rs7412 for LDL-C (p = 3.2e-70), *CETP* rs7205804 for HDL-C (p = 7.4e-14), *LPA* rs74617384 for Lp(a)-Ch (p=3.9e-60), *LPL* rs115849089 for TG (p=2.3e-8), and *APOA5* rs3741298 for TG (p=3.9e-8). Gene-based analyses of rare null protein-coding variants identified previously reported associations: *LDLR*, *APOB* for LDL-C (p = 1.9e-23 and 2.4e-7, respectively). Across 31.5M non-coding variants, we observed 12,443 rare variants (MAF < 1%) at HepG2 promoters and enhancers within 20KB of the TSS of 301 lipid genes. Non-coding burden analyses demonstrated suggestive association (p<0.05) at many lipid genes. Non-coding burden analyses of rare variants in enhancers linked to genes showed suggestive association at *R3HDM2* and *FAM117B* for LDL-C; *LIPG* and *APOA4* for HDL-C; and *VEGFA*, *HIST1H1B*, *ARHGEF15*, and *ABCG5* for triglycerides. Non-coding burden analyses of rare variants in promoters and enhancers near transcription start sites showed suggestive association at *NF1*, *SBNO1*, *APOA5*, and *APOE* for LDL-C; *SERINC3*, *ABCG5*, *NBEAL2*, *LRP4*, and *ANGPTL3* for HDL-C; and *PLCB3*, *HIST1H1B*, *APOB*, *SERBP1*, *FADS1*, and *SLC22A18AS* for triglycerides. Rare variants grouped within regulatory functional units may impact blood lipid variation; replication of these results is ongoing.

697W

Evaluation of coronary artery disease risk loci as targets of current and prospective drugs. *V. Tragante do O[1], F. Asselbergs[1,2,3].* 1) Heart & Lungs, University Medical Center Utrecht, Utrecht, Utrecht, Netherlands; 2) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands; 3) Faculty of Population Health Sciences, Institute of Cardiovascular Science, University College London, London, UK.

Genome-wide association studies (GWAS) have so far identified around 60 loci associated with coronary artery disease (CAD) and myocardial infarction (MI). Only a few of these loci are currently targets of on-market medications. Here, we systematically evaluated these risk loci for their druggability, either theoretically (with databases of small molecules) or based on known and predicted interactions (with databases of drug-gene interactions), including their potential side effects. Our results show that dozens of drugs potentially target pathways affected by CAD/MI associated loci. These drugs can be grouped for predominantly treating neoplastic disorders, blood and blood forming organs and the cardiovascular system. However, 80% of pathways affected by CAD/MI loci are not targeted by any drug. The overlap of indications and side effects of the drugs that target these genes evidence diabetes and hypertension as the main risk factors for CAD/MI. After filtering for undesired side effects, we discuss the possibilities of atrasentan, imatinib and tocilizumab as candidates for repurposing for treatment of CAD/MI.

698T

Haplotype phasing of key cardiac disease genes at genome and transcriptome levels using long-read sequencing technologies. *A. Dainis[1], E. Tseng[2], T. Clark[2], T. Hon[2], M. Wheeler[3], E. Ashley[1,3].* 1) Genetics, Stanford University, Stanford, CA; 2) Pacific Biosciences, Menlo Park, CA; 3) Cardiovascular Medicine, Stanford University, Stanford, CA.

Cardiac diseases, including hypertrophic cardiomyopathy (HCM [MIM 192600]), can be caused by rare genetic mutations. While current clinical sequencing strategies have ever improving power to find these disease-causing mutations, mutation detection often lacks additional information about the haplotype or chromosome of origin of these mutations. Without this information, critical knowledge about the disease-causing genes, including resolution of compound heterozygotes, is lost. Haplotype phasing using short-read sequencing often requires additional sequencing of close family members or inferential phasing using common genetic variants, a process which may provide poor resolution for rare mutations. New long-read sequencing technologies, including Pacific Biosciences SMRT Sequencing and Oxford Nanopore MinION Sequencing, can allow us to capture multiple SNPs in a single sequencing read, reducing the need for familial sequencing or inferential phasing. We present the results of sequencing multiple cardiac disease related genes, including *MYH7* (MIM 160760) and *MYBPC3* (MIM 600958), using three long-read haplotyping strategies: gene-specific gDNA capture and sequencing using PacBio SMRT sequencing, targeted-primer cDNA sequencing using PacBio SMRT sequencing, and tiling amplification of gDNA for Oxford Nanopore sequencing. We compare all three methods for ease and efficiency, sequencing error-rates, and downstream analysis. These three haplotyping strategies may lend themselves best to their own individual problems: when rapid haplotyping is needed, a tiling gDNA strategy on the MinION may be preferred, but when economical sequencing of many genes is required, a capture strategy using PacBio may be the best course of action.

699F

Label-free quantitative proteomic profiling of discordant monozygotic twin blood serum identifies Fibulin-1 as a candidate biomarker for ischaemic stroke. *N. Vadgama[1], R.C. Lovering[2], D. Lamont[3], J. Hardy[1], J. Nasir[4].* 1) Instituite of Neurology, University College London, London, United Kingdom; 2) Institute of Cardiovascular Science, University College London, London, United Kingdom; 3) College of Life Sciences, University of Dundee, Dundee, United Kingdom; 4) Molecular & Clinical Sciences Research Institute, St. George's University of London, United Kingdom.

Background: Stroke is a common disorder with significant morbidity and mortality, and complex aetiology involving both environmental and genetic risk factors. Although some of the major risk factors for stoke, such as smoking and hypertension, are well-documented, the underlying genetic and detailed molecular mechanisms remain elusive. Thus, identification of biomarkers may contribute to the clinical diagnosis of stroke. **Methods**: We performed a comparative proteomic analysis of blood serum of monozygotic twins discordant for ischaemic stroke using a label-free proteomics method to identify qualitative and quantitative differences at the protein level. To explore the biological processes associated with the differentially expressed proteins, we performed pathway analysis and Gene Ontology classification using g:Profiler. **Results**: MS/MS data for candidate peptides were exported to the MASCOT database for comparison to known peptides. Proteins were identified and quantified with high confidence (MASCOT ion score of 35 and $p<0.05$) and reimported into Progenesis software for further filtering. 190 proteins were identified in the plasma and a functional analysis was performed on those which were differentially expressed. These proteins were classified according to biological function and canonical pathway. The categories overrepresented in the affected twin related to vesicle mediated transport (n=12), response to stress (including wound heading) (n=16), protein activation cascade (including complement activation) (n=11), and beta-amyloid clearance (n=2). Further, the expression of Fibulin-1, an extracellular matrix protein in cardiovascular tissues, was consistently up-regulated in both replicate datasets of the affected twin. **Conclusion**: The identification of cellular pathways enriched in ischaemic stroke as well as the up-regulation of Fibulin-1 sheds new light on the underlying disease-causing mechanisms at the molecular level, which could lead to the development of new drug targets and biomarkers in the future.

700W

The gut microbiome associates with lifetime cardiovascular disease risk profile among Bogalusa Heart Study Participants. *T.N. Kelly[1], L.A. Bazzano[1,2], N. Ajami[3], H. He[1], J. Zhao[1], J. Petrosino[3], A. Correa[4], J. He[1,2].* 1) School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA; 2) School of Medicine, Tulane University, New Orleans, LA; 3) The Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX; 4) University of Mississippi Medical Center, Jackson, MS.

   With growing evidence for a critical role of the gut microbiome in energy harvest, glucose and lipid metabolism, and systemic inflammation, a potential influence of this supraorganism on cardiovascular health has been hypothesized. The objective of the current analysis was to examine the association between the gut microbiome and lifetime cardiovascular disease (CVD) risk profile among 57 Bogalusa Heart Study (BHS) participants with the highest and 55 with the lowest lifetime burdens of CVD risk factors. Lifetime CVD risk profile was estimated using a z-score based data reduction technique that leveraged fasting plasma glucose, systolic blood pressure, and low-density lipoprotein cholesterol values collected from an average of 9 study visits obtained over 33 years of follow-up. 16S rRNA sequencing was conducted on microbial DNA extracted from stool samples of the BHS participants. Alpha diversity, including measures of richness and evenness, and individual genera were tested for associations with lifetime CVD risk profile. Multivariable regression techniques were employed to adjust for age, gender, and race (Model 1), and additionally body mass index (Model 2). Increases in microbial richness, but not evenness, were associated with decreased odds of high lifetime CVD risk profile. In Model 1, odds ratios (95% confidence intervals) for each standard deviation increase in richness, measured by the number of observed operational taxonomic units, Chao 1 index, and abundance-based coverage estimator, were 0.62 (0.39, 0.99), 0.61 (0.38, 0.98), and 0.63 (0.39, 0.99), respectively. Associations were consistent in Model 2. Eight microbial taxa associated with lifetime CVD risk profile in Model 1 including: *Alloprevotella* ($P=2.96 \times 10^{-6}$), *Paraprevotella* ($P=6.48 \times 10^{-4}$), *Prevotella 7* ($P=7.95 \times 10^{-5}$), *Catenibacterium* ($P=1.82 \times 10^{-4}$), *Megamonas* ($P=4.18 \times 10^{-5}$), *Tyzzerella* ($P=4.39 \times 10^{-4}$), *Tyzzerella 4* ($P=1.51 \times 10^{-4}$), and *Enterobacter* ($P=6.41 \times 10^{-8}$). Five additional microbial taxa were identified by Model 2 including: *Coprococcus 2* ($P=1.38 \times 10^{-4}$), *Megasphaera* ($P=1.38 \times 10^{-4}$), *Ruminococcus* ($P=5.01 \times 10^{-4}$), *Thalassospira* ($P=3.44 \times 10^{-4}$), and *Methanobrevibacter* ($P=1.20 \times 10^{-4}$). Findings from this study suggest that increased microbial richness but not evenness are associated with lower lifetime CVD risk. Thirteen taxa also associated with lifetime CVD risk. These cross-sectional data add to the accumulating evidence that microbiota may play an important role in CVD risk.

701T

Transcriptome signatures and structural variations in clinical subtypes of congenital heart disease (CHD) based on next generation sequencing technologies. *Y. Liu[1], A. Desai[1], K. Nguyen[1], X. Chang[1], E. Goldmuntz[2], H. Hakonarson[1].* 1) Center for Applied Genomics CAG, The children's hospital of philadelphia CHOP, Philadelphia, PA; 2) Division of Cardiology The Children's Hospital of Philadelphia CHOP, Philadelphia, PA.

   Congenital heart disease (CHD) constitutes defects in the structure of the heart and the great vessels that is present at birth. CHD is among the most common birth defects with approximately 9 out of 1000 newborns affected and it's the leading cause of birth defect-related deaths. Clinically, there are multiple subtypes of CHD. However, previous genetic studies have usually treated different clinical subtypes of CHD as a single identity or focused on individual subtypes without comparison between them. In this study, we observed transcriptome differences among 4 common CHD subtypes of 13 tissue types based on 145 genomic RNA-seq samples from 118 individuals. We uncovered differentially expressed genes (DEGs) between conotruncal cardiac defects (CTD) and left ventricular outflow tract obstruction (LVOT) defects that involve genes related to embryonic organ development processes/vascular processes, and these expression profiles were distinct from those observed for heterotaxy (Het) and atrial septal defect (ASD), which involved genes regulating energy processes such as oxidative phosphorylation. The differentially expressed genes did not show tissue specificity but gene expression levels in general were different among different cardiac tissues. To further explore the potential underlying causes for the DEGs, whole exome sequencing (WES) was performed from blood-derived DNA samples for those 118 individuals. Using multiple filters, recurrent CNV regions were detected for some of the CHD subtypes. Overall, we identified 8 transcriptome signature genes of differential expression impact: *C5orf46, ODF2L, MAP7D3, PSIP1,CLGN, ABI3BP, CCDC88A*, and *AGPS*, based on integrative analysis of DGEs, gene module networks, and CNVs. Those signature genes were further validated using 3D protein structure modeling. Acknowledgment: We thank the PCGC consortium for providing access to the data used for this study.

702F

Analysis of the human kidney transcriptome reveals transcriptional regulation of genes crucial for blood pressure regulation. *C. Mercado[1], X. Wang[1], M. Bell[1], G. Roberts[2], Y. Chang[1].* 1) Medicine, University of Maryland, Baltimore, Baltimore, MD; 2) University of Maryland, Baltimore County, Baltimore, MD.

Serine/threonine kinases (WNK1, WNK4, SPAK, OSR1) and cation co-transporters (NKCC2, NCC) are members of a multi-kinase network that determines renal Na$^+$ reabsorption and blood pressure (BP) regulation. Environmental stressors such as hypovolemia activate WNK kinases to phosphorylate SPAK and OSR1, which in turn phosphorylate NKCC2 and NCC, allowing for the influx of extracellular Na$^+$ and Cl$^-$ into renal epithelial cells to restore blood volume. Rare mutations in these genes cause monogenic forms of hyper- and hypotension, which are well characterized in mouse models. While post-translational regulation of these proteins has been well established, regulation of these genes at the transcript level is not completely understood. In this study, we performed RNA-seq of the human kidney transcriptome, complemented with 5' and 3' RACE, to uncover pre- and post-transcriptional regulation of several genes in this pathway. First, *STK39* encodes for Ste20-related proline alanine rich kinase (SPAK), which phosphorylates and activates cation co-transporters. Variants within *STK39* are associated with susceptibility to essential hypertension, and SPAK null mice are hypotensive and mimic Gitelman syndrome (GS), a rare monogenic salt-wasting human disorder. In mice, SPAK activity is determined by the nephron segment-specific expression of full length SPAK and truncated SPAK isoforms with impaired kinase function. In humans, we identified N-terminally truncated *STK39* transcripts transcribed with human-specific promoters not found in mice. The relative abundances of these alternative isoforms are also distinct from rodent models, suggesting differential regulation of Na$^+$-reabsorption between humans and mice. Second, *SLC12A3* encodes for the thiazide-sensitive Na$^+$/Cl$^-$ cotransporter (NCC), and rare mutations in this gene cause GS. *SLC12A3* undergoes alternative splicing (AS) and alternative polyadenylation at multiple sites. AS within the final exon leads to an exon-exon junction >50 nucleotides downstream of the termination codon and generates a potential substrate for nonsense-mediated decay (NMD). By suppressing NMD, we demonstrated that *SLC12A3* transcript abundance is partially determined by a mechanism termed "regulated unproductive splicing and translation" (RUST). In summary, genes in this pathway are under complex pre and post-transcriptional regulation, resulting in the differential expression of alternative transcripts that may contribute to the fine-tuning of BP control.

703W

Molecular characterization of pediatric restrictive cardiomyopathy from integrative genomics. *T.N. Rindler[1], R.B. Hinton[1], N. Salomonis[1], S.M. Ware[2].* 1) Pediatrics, Cincinnati Children's Hospital, Cincinnati, OH; 2) Pediatrics and Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

Pediatric restrictive cardiomyopathy (RCM) is a genetically heterogeneous heart muscle disease with limited therapeutic options. RCM cases are largely idiopathic; however, even within families with a known genetic cause for cardiomyopathy, there is striking variability in the disease severity. Although accumulating evidence suggests a strong role for regulation of both gene expression and alternative splicing in the development of dilated cardiomyopathy (DCM), there have been no detailed molecular characterizations of the underlying pathways dysregulated by RCM. By applying RNA-Seq to a cohort of pediatric RCM patients and comparing to other forms of adult cardiomyopathy and controls, we have identified transcriptional differences highly common to all examined cardiomyopathies, as well as those unique to RCM. RCM-selectively induced transcripts include a number of known and novel G-protein coupled receptors linked to calcium handling and contractile regulation. An in-depth comparison of alternative splicing revealed splicing difference shared among cardiomyopathy subtypes as well as splicing events linked solely to RCM. Genes identified with altered alternative splicing implicate direct regulation by the DCM splicing factor, RBM20. In conclusion, we present the first comprehensive report on the molecular pathways dysregulated in pediatric RCM including shared and unique pathways identified in comparison to other subtypes of cardiomyopathy. Analyses identify RBM20 as a potential mediator of alternative splicing dysregulation and demonstrate that disruption of the alternative splicing patterns in pediatric RCM is in the inverse direction as occurs in DCM.

704T

Assessing chromatin marks to unveil the pleiotropic mechanisms of 122 loci associated with blood pressure. *D. Hemerich[1,2], J. Pei[3], M. Harakalova[1], T. Lumbers[4], T. Treibel[4], J. van den Velden[5], I. Ifimov[6], B. Boukens[6], C. Cheng[3], H. Kerstens[7], M. Mokry[8], J.C. Moon[4], J. Moore[9], P.B. Munroe[10], V. Tragante[1], F.W. Asselbergs[1,4,11].* 1) University Medical Center Utrecht, Utrecht, Utrecht, Netherlands; 2) CAPES Foundation, Ministry of Education of Brazil, Brasília, Brazil; 3) Department of Nephrology, UMC Utrecht, Utrecht, Netherlands; 4) Institute of Cardiovascular Science, UCL, London, UK; 5) Department of Physiology, VU UMC, Amsterdam, Netherlands; 6) Department of Biomedical Engineering, WUSTL, USA; 7) Princess Máxima Center for Pediatric Oncology, Department of Research, Utrecht, Netherlands; 8) Department of Pediatrics, Wihelmina Children's Hospital, Utrecht, Netherlands; 9) Department of Biostatistics and Epidemiology, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, USA; 10) Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, UK; 11) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, Netherlands.

   Background: High blood pressure (BP) is a major risk factor for cardiovascular disease and premature death, that can lead to a variety of outcomes (aneurysm, left ventricular hypertrophy, stroke, kidney failure or coronary artery disease, for instance). To better understand how BP-associated variants influence different phenotypes, we aim at assessing to which extent regulatory elements are affected by 122 BP-associated loci in different tissues. **Materials and Methods**: We tested the phenotypic cell-type specificity of the 122 independent single nucleotide polymorphisms (SNPs) and vicinity ($r^2 > 0.8$) with H3K27ac peaks across 48 datasets of different tissues (43 from public sources, and 5 recently generated in-house from myocardial biopsy tissues). Statistical significance was evaluated by permuting 10,000 matched sets of SNPs not associated with the phenotype to calculate a 95th-percentile threshold as cell-type specificity score, resulting in specific tissues, best scoring (BS) SNPs and their respective index SNPs influencing each specific tissue. We intersected both sets of index and BS SNPs with 342 datasets of DNase data. Both DNase and H3K27ac are marks of active transcription. We used 43 datasets on H3K27me3, a histone mark known to silence genes, for comparison. **Results**: The most significant cell-types were right ventricle (RV) ($p=3\times10^{-4}$), kidney ($p=3\times10^{-3}$), both public ($p=0.01$) and in-house ($p=4\times10^{-3}$) datasets of left ventricle, aorta ($p=0.02$) and adrenal gland, both adult ($p=0.02$) and fetal ($p=0.05$), all tissues related to possible different outcomes induced by BP. The intersection of the two sets of BS (74) and index (63) SNPs with DNase data showed that 45 BS SNPs (60,8%) caused 2998 overlaps with peaks, while 19 index SNPs (30,1%) implicated 616 overlaps, a 4.8 fold enrichment. A similar rate was observed with H3K27ac peaks in the most significant tissue, RV, with 23 overlaps by 23 BS SNPs (31,1%) and 4 by 4 Index SNPs (6,3%) (fold enrichment=5.7). When testing intersection with peaks of an opposite mark (H3K27me3), 21 Index SNPs (33,3%) had 147 overlaps, while 33 BS SNPs (44,6%) had 222, a fold enrichment of only 1.5. **Conclusion**: These preliminary results point to an enrichment of BP-related SNPs with chromatin marks, highlighting variants in LD that may play a causal role. Further analysis may help clarify the mechanisms through which these SNPs affect genes, resulting in different outcomes.

705F

Plasma levels of miR-19a, miR-26a and miR-584 are associated with severity of coronary artery disease in Turkish men. *N. Coban[1], D. Pirim[2], A.F. Erkan[3], F. Guclu Geyik[1], B. Ekici[3], N. Erginel Unaltuna[1].* 1) Department of Genetics, Institute for Experimental Medicine, Istanbul University, Istanbul, Turkey; 2) Department of Molecular Biology and Genetics, Mus Alparslan University, Mus, Turkey; 3) Department of Cardiology, Medical Faculty, Ufuk University, Ankara, Turkey.

   MicroRNAs (miRNAs) have been shown to be associated with various physiological and pathological conditions, including inflammation and cardiovascular disease, but little is known about their relationship with the presence of coronary artery disease (CAD) and disease severity. We hypothesized that particular miRNAs may be differentially expressed in patients with severe coronary artery disease (SYNTAX score ≥33). In order to test the hypothesis, ninety nine patients with angina or acute myocardial infarction who underwent coronary angiography were recruited and divided into 2 groups: normal coronary arteries (coronary lesion with ≤30% stenosis) and critical disease (≥1 coronary lesion with ≥70% stenosis). Of the 99 total subjects, 57 were diagnosed as having CAD. Blood samples were drawn before coronary angiography. Gensini and SYNTAX scores and myocardial blush grade (MBG) were assessed. MiRNA expression profiles were analyzed by using the Agilent's miRNA microarray system in selected 4 male patients with proven severe atherosclerotic CAD (SYNTAX score ≥33) and 4 without CAD. MiRDB target prediction tool was utilized to identify miRNA target genes involved in atherosclerosis pathway for differently expressed miRNAs. Expressions of selected miRNAs were analyzed by using real-time PCR in 57 patients with CAD and of 42 healthy controls. Microarray results indicate that miR-584 was significantly down-regulated in male patients compared to non-CAD group (fold change>1.5, *p*<0.05). In addition, among CAD patients with three-vessel disease, higher SYNTAX scores were significantly associated with higher levels of miR-19a and miR-26a (fold change>1.5, *p*<0.05). Based on miRDB target prediction scores, we identifed *ABCA1, ADM and PHACTR1* genes were strong candidates (target score≥93) for miR-19a, miR-26a and miR-584, respectively. Currently, expression profiles of miR-19a, miR-26a, and miR-584 are being analyzed by qRT-PCR in 57 patients with CAD and of 42 healthy controls including male and female subjects. Lower plasma levels of miR-584 and higher levels of miR-19a and miR-26a were significantly associated with the presence as well as severity of CAD in Turkish men. Plasma miR-584, miR-19a and miR-26a could be potential biomarkers for CAD, and thus may be utilized in predicting CAD risk and its severity in patients presenting with chest pain.

706W

RNA sequencing analyses reveal eight miRNA/mRNA pairs that may mediate the effects of cocaine in cardiovascular disease. *C.M. Dong, W. Zhu, H.L. Wang, G.C. Sartor, M. Bao, A.J. Griswold, C.R. Wahlestedt.* University of Miami, Miami, FL.

Introduction: Cocaine abuse is a significant risk factor for cardiovascular (CV) mortality and morbidity. The molecular mechanisms underlying the cocaine effects, however, remain elusive. **Goal:** MicroRNAs (miRNAs) are short, noncoding RNAs that bind to the 3'-UTR of the target mRNA, resulting in translational repression or degradation. This study aimed to identify miRNA/mRNA pathways that may mediate cocaine's effects in the CV system. **Methods:** C57BL/6 mice were treated daily with cocaine (20mg/kg body weight, ip injection), cocaine methiodide (CM, a cocaine analogue that does not enter the CNS), or saline for consecutive 10 days. Total RNA was extracted from the aortas (n=3) and subjected to HTSeq-based small RNA sequencing and RNA sequencing containing >41,000 probes. Gene expression profiles between treatment groups (fold changes ≥2) were analyzed, and candidate genes were validated by qRT-PCR. Putative pathways were identified by Gene Ontology and KEGG pathways analyses. miRNA's target mRNAs were predicted using miRDB. The relationship between miRNAs and mRNAs were determined by co-transfection of HEK293 cells using wt and mutant candidate gene 3'UTRs. **Results:** When compared to saline treatment, 8 miRNAs were downregulated by both cocaine and CM, and 4 miRNAs were upregulated by both treatments. Relative to saline, 48 genes were upregulated by both cocaine and CM, and 267 genes were downregulated by both agents. MiRNA prediction analysis revealed 2070 genes to be targets for the 4 upregulated miRNAs, and 3205 genes to be targets for the 8 downregulated miRNAs. Superimposition of the dataset from miRNA prediction with the RNAseq data revealed 71 genes predicted to be targets of the 4 upregulated miRNAs showed decreased expression (anti-correlation) in cocaine and CM-treated aortas, whereas 18 genes predicted to be the targets of the 8 downregulated miRNAs showed increased expression by cocaine and CM. Based on their relevance to cellular physiology and CV diseases, we ranked 8 top miRNA/mRNA pairs and established causal relationship between the miRNAs and their target genes. **Conclusions:** Cocaine appears to affect the homeostasis and functional integrity of the aortas by affecting the expression of miRNAs and their target genes. Further functional analyses of the candidate pathways in vitro and in vivo will lead to the discovery of novel molecular targets for intervention for cocaine-induced CV toxicities.

707T

Interpretable deep learning approaches to understand the genetic and regulatory basis of coronary artery disease. *P. Greenside, J. Israeli, C. Miller, M. Pjanic, T. Quertermous, A. Kundaje.* Stanford University, Stanford, CA.

The majority of genetic variants associated with coronary artery disease (CAD), such as those identified through genome-wide association studies, are enriched in non-coding regulatory elements that are preferentially active in CAD-relevant cell types. However, deciphering causal variants and the context-specific regulatory mechanisms through which they manifest their effects remain key challenges. To address these issues, we present a deep learning framework to learn context-specific de-novo sequence features that can accurately predict transcription factor (TF) binding from ChIP-seq experiments and chromatin accessibility dynamics from ATAC-seq experiments in coronary artery smooth muscle cells in healthy, stimulated and diseased states. We develop a novel method for efficiently computing feature importance in deep neural networks to highlight binding sites of individual TFs, homotypic and heterotypic sequence grammars with spatial constraints and important nucleotides in flanking sequences of chromatin accessible regulatory elements. We focus our study on the TCF21 transcription factor in the context of these dynamic accessibility profiles as TCF21 binding events are highly enriched for statistically significant CAD-associated variants. We discover a novel cis-regulatory grammar involving TCF21 and previously unidentified cofactors that is highly predictive of chromatin accessibility changes across normal and diseased states. Finally, we use our model to score and identify putatively causal non-coding CAD-associated variants and obtain novel hypotheses about the regulatory mechanisms through which they disrupt intermediate molecular phenotypes.

708F

Cigarette smoking changes CpG methylation and thus increases the expression of *GPR15* **- A potential candidate gene for plaque development?**
T. Haase[1,2], J. Krause[1,2], C. Mueller[1,2], J. Stenzig[2,3], C. Roethemeier[1], P. Wild[4,5,6], S. Blankenberg[1,2], T. Zeller[1,2]. 1) University Heart Center Hamburg, Hamburg, Germany; 2) DZHK (German Center for Cardiovascular Research), partner site Hamburg/Kiel/Lübeck, Germany; 3) Department of Experimental Pharmacology and Toxicology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4) Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; 5) Center for Thrombosis and Hemostasis, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; 6) DZHK (German Center for Cardiovascular Research), partner site Rhine-Main, Mainz, Germany.

   Purpose Cigarette smoking increases the risk for cardiovascular disease. Differentiation of monocytes to proinflammatory macrophages is an important step in plaque development, but the mechanisms remain not fully understood. *GPR15*, coding for the G protein-coupled receptor 15, is an interesting new cardiovascular candidate gene as the CpG locus cg19859270 within the *GPR15* gene has been identified to be hypomethylated in smokers. The aim of this study was to analyze the connection between smoking, *GPR15* and plaque development.   **Methods** To determine the effect of changes in smoking over time, monocytic *GPR15* mRNA expression was measured by qPCR in smokers (n=179), ex-smokers (n=369) and never-smokers (n=480) from the population-based Gutenberg Health Study at baseline and five-year follow-up. The methylation status of the *GPR15* cg19859270 locus was determined in PBMCs from 262 subjects by bisulphite conversion and high resolution melting and mRNA expression was measured by microarray. To evaluate the role of *GPR15* in plaque development, monocytes from 4 smokers and 5 never-smokers were differentiated *in vitro* into proinflammatory macrophages and *GPR15* expression was measured by qPCR.   **Results** Monocytic *GPR15* expression was significantly higher in smokers compared to ex- (p<0.001) and never-smokers (p<0.001). Ex-smokers had significantly higher *GPR15* expression levels than never-smokers up to ten years after quitting (p<0.001). Smoking initiation between baseline and follow-up (n=17) led to an increase in *GPR15* expression, whereas smoking cessation resulted in decreased *GPR15* expression (n=42, p<0.001). In PBMCs, the correlation between smoking and *GPR15* expression (p<0.0001) was associated with cg19859270 hypomethylation (p<0.05). Likewise, smoking cessation was linked to cg19859270 methylation as well as reduced *GPR15* expression (p<0.05). In macrophages, *GPR15* expression was significantly lower compared to monocytes in smokers (p<0.001) and never-smokers (p<0.01).   **Conclusion** This is the first longitudinal study showing that smoking changes *GPR15* expression, mediated by changes in DNA methylation. Whether *GPR15* plays a role in cardiovascular disease development needs to be further analyzed.

709W

Epigenome-wide association study for myocardial infarction identified a DNA methylation site on the loci related to atrial fibrillation and ischemic stroke. *M. Nakatochi[1], S. Ichihara[2], K. Yamamoto[3], T. Matsubara[4], M. Yokota[5].* 1) Statistical Analysis Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 2) Graduate School of Regional Innovation Studies, Mie University, Tsu, Japan; 3) Department of Medical Biochemistry, School of Medicine, Kurume University, Kurume, Japan; 4) Department of Internal Medicine, School of Dentistry, Aichi Gakuin University, Nagoya, Japan; 5) Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya, Japan.

   Aims Development of myocardial infarction (MI) depends on environmental and genetic factors. To investigate epigenetic basis for MI, we performed an epigenome-wide association study for MI in an elderly Japanese. We also explored DNA methylation (DNAm) sites associated with single nucleotide polymorphisms (SNPs) related to cardiovascular disease (CVD). *Methods* 192 case subjects with MI and 192 control subjects were enrolled, and whole-blood DNA samples were obtained from them. Genome-wide DNAm profiles for their DNA samples were obtained by analysis with an Infinium HumanMethylation450 BeadChip. Genome-wide SNPs were genotyped using Illumina HumanOmniExpress-12 BeadChip. Unassayed genotypes were imputed using the 1000 Genome Project phase 3 reference panel. From measured and imputed genotypes, we extracted genotypes of 145 SNPs reported to be associated with the CVD including MI. First, the association of MI with DNAm status was performed with a general linear model, and then the association of DNAm status with CVD-associated SNPs was assessed. *Results* Three DNAm sites showed genome-wide significant associations with MI. Out of these DNAm sites, one DNAm site, which indicated most significant association ($\beta = 0.018 \pm 0.003$, $p = 3.96 \times 10^{-10}$), is located within the *ZFHX3* gene. Also, we identified 228 combinations of CVD-associated SNP and DNAm site that showed significant association ($p < 1 \times 10^{-8}$). The above DNAm site was not included in the combinations. *Conclusions* We identified three DNAm sites significantly associated with MI. Out of these, one DNAm site is located within the *ZFHX3* gene. Although SNPs in *ZFHX3* on 16q22 has been reported to be associated with atrial fibrillation and ischemic stroke, there is no report that *ZFHX3* was associated with MI. Our results suggest a novel insight that the development of MI was caused by changes in DNAm within *ZFHX3* via a different pathway with SNPs in *ZFHX3*.

710T

GWAS candidate gene for coronary artery disease *TCF21* **interacts with the aryl-hydrocarbon receptor to modify coronary smooth muscle cell response to pro-atherogenic stimuli.** *M. Pjanic[1], J.B. Kim[1], O. Sazonova[1], T. Nguyen[1], T. Wang[1], C.L. Miller[1], L. Maegdefessel[2], U. Hedin[2], T. Quertermous[1].* 1) Department of Medicine, Cardiovascular Institute, Stanford University, Stanford, CA; 2) Karolinska Institute, Stockholm, Sweden.

Genome-wide association studies (GWAS) for coronary artery disease (CAD) have discovered and validated 48 loci genome-wide and recent 1000-Genomes based meta-analyses have discovered additional 8 loci with significant association, however, detailed follow-up studies on the mechanisms of association are still scarce. Here we analyze the relationship of two transcription factors, TCF21, one of the lead GWAS candidate genes for CAD, and AHR, aryl hydrocarbon receptor, which previously was not implicated as fundamental for the atherosclerotic process. We show that both the predicted and in-vivo ChIP-Seq binding sites for TCF and AHR colocalize in the human genome with over 400 predicted sites and 119 ChIP-Seq sites directly overlapping. TCF and AHR-ARNT predicted binding sites longitudinal phasing was observed near the transcription start sites, outlining the position of +1 nucleosome. AHR-ARNT matrix was highly enriched in both TCF21 ChIP-Seq peaks and ATAC-Seq open chromatin regions in human coronary artery smooth muscle cells (HCASMC), implicating the role of AHR in this important vascular cell type. Co-expression modules of TCF21 and AHR showed high degree of connectivity. Separation of rotationally phased and unphased predicted binding sites for TCF and AHR resolved the roles of direct and indirect TCF-AHR interactions, and implicated as highly-modulated various inflammatory, interleukin and cytokine related processes, while ChIP-Seq co-localization of TCF21 and AHR/ARNT emphasized the role of calcium related processes. We performed TCF21 overexpression analysis in HCASMC and obtained GO ontologies that recapitulate in vivo pathophysiology of the atherosclerotic vessel wall, and a set of chronic inflammatory ontologies was established with the colocalization of TCF21 and AHR ChIP-Seq sites as well as with the binomial testing of GWAS SNP overrepresentation. We experimentally confirmed that TCF21 binds to and regulates AHR gene expression in HCASMC, as well as to elements near AHR downstream genes, such as CYP1A1, using reporter assays. The functional relevance of AHR pathway in HCASMC is confirmed using AHR ligands TCDD and oxidized LDL. Finally, we show that AHR is elevated in atherosclerotic arteries using laser capture microdissection in mice in vivo and in human ex vivo. In conclusion, we extend GWAS results to functional assays and show that TCF21 and AHR functional connectivity provides a novel mechanism for diseased coronary vessel wall biology.

711F

Effect of lipoprotein (a)-associated genetic variants on plasminogen levels and fibrinolysis. *X. Wang[1,2], C. Hong[1], H. Wang[1], J. Lewis[1], Y. Zhu[1], X. Chu[3], J. Backman[1], Z. Hu[1], P. Yang[4], C. Still[3], G. Gerhard[3,5], M. Fu[1].* 1) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, baltimore, MD; 2) Department of Orthopedic Surgery, Second Affiliated Hospital of Chongqing Medical University, Chongqing 400010, China; 3) Geisinger Obesity Institute, Geisinger Clinic, Danville. PA 17822, USA; 4) Department of Obstetrics, Gynecology & Reproductive Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA; 5) Penn State Institute for Personalized Medicine, Penn State College of Medicine, Hershey, PA 17033, USA.

Two genetic variants (rs3798220 and rs10455872) in the apolipoprotein (a) gene (*LPA*) have been implicated in cardiovascular disease (CVD), presumably through their association with Lipoprotein (a) (Lp(a)) levels. While Lipoprotein (a) (Lp(a)) is recognized as a lipoprotein with atherogenic and thrombogenic characteristics, it is unclear whether or not the two Lp(a)-associated genetic variants in the *LPA* gene are also associated with markers of thrombosis (i.e. plasminogen levels and fibrinolysis). In the present study, we genotyped the two genetic variants in 2919 subjects of the Old Order Amish (OOA) and recruited 146 subjects according the carrier and non-carrier status for rs3798220 and rs10455872, and also matched for gender and age. We measured plasma Lp(a) and plasminogen levels in OOA subjects, and found that the concentrations of plasma Lp(a) were 2.62 and 1.73 fold higher in minor allele carriers of rs3798220 and rs10455872, respectively, compared with non-carriers ($P = 2.04 \times 10^{-17}$ and $P = 1.64 \times 10^{-6}$, respectively). By contrast, there was no difference in plasminogen concentrations between carriers and non-carriers of rs3798220 and rs10455872. Furthermore, we observed no association between carrier status of rs3798220 or rs10455872 with clot lysis time. Finally, plasminogen mRNA expression in liver samples derived from 76 Caucasian subjects was not significantly different between carriers and non-carriers of these two genetic variants. Our results provide further insight into the mechanism of action behind two genetic variants previously implicated in CVD risk and show that these polymorphisms do not modulate plasma plasminogen levels and fibrinolysis.

712W

Cocaine may induce hypertension and aortic stiffness via the miR-30c-5p-Me1-NADPH oxidase-ROS pathway. *W. Zhu, H.L. Wang, G.C. Sartor, J.Q. Wei, M. Bao, C.T. Pierce, C.R. Wahlestedt, C.M. Dong.* University of Miami, Miami, FL.

   Introduction: Cocaine abuse is a significant risk factor for hypertension (HTN) and vascular aging/aortic stiffness. It is known that cocaine stimulates sympathetic nervous system by inhibiting catecholamine reuptake at sympathetic nerve terminals; however, this has not been shown to be the major mechanism of action in cocaine-induced HTN. Using RNA sequencing, we have demonstrated that cocaine treatment results in altered expression of microRNAs and target genes that segregate into eight molecular pathways. One of them is the microRNA-30c-5p (miR-30c)—malic enzyme 1 (Me1)—NADPH oxidase—reactive oxygen species (ROS) pathway. Me1 is a cytosolic protein important for oxidation-reduction and ROS production. **Goal:** Our goal was to establish if the miR-30c—Me1—NADPH oxidase—ROS axis mediates, at least in part, the effects of cocaine in cardiovascular disease (CVD). **Methods:** C57BL/6 mice were treated daily with cocaine (20mg/kg body weight, ip injection), cocaine methiodide (CM, a cocaine analogue that does not enter the CNS), or saline for consecutive 10 days. Blood pressure (BP) and pulse wave velocity (PWV) for aortic stiffness were obtained. miR-30c and Me1 expression was determined by qRT-PCR and/or immunohistochemistry. The causal relationship between miR-30c and Me1 was established using lusiferase assay with wild type (wt) and mutant Me1 3'UTR (disrupting the binding sites for miR-30c) in HEK293 cells. ROS production was detected by dihydroethidium (O_2^-), chloromethyl-2',7'-dichlorofluorescein diacetate (H_2O_2), and aminophenyl fluorescein ($ONOO^-$). **Results:** When compared with saline treatment, BP and PWV were significantly increased by cocaine and CM over the entire treatment course. Cocaine and CM also caused downregulation of Me1 and upregulation of miR-30c expression in the aortas, resulting in increased O_2^-, H_2O_2, and $ONOO^-$ production. Importantly, transfection of miR-30c dose-dependently inhibited the luciferase activities of wt Me1 3'UTR, whereas miR-30c had no effects on the luciferase activity of the mutant Me1 3'UTR, indicating that miR-30c directly interacts with Me1 3'UTR, suppressing Me1 expression and leading to enhanced ROS production. **Conclusions:** Our data indicate that cocaine may induce HTN and aortic stiffness in mice via, at least in part, the miR-30c—Me1—NADPH oxidase—ROS pathway. Further characterization of this pathway *in vivo* may lead to the development of novel therapeutic strategies to combat drug abuse-related CVD.

713T

Integration of allele-specific chromatin signatures with GWAS findings to localise functional variation for complex traits. *A.J.P. Smith[1,3], F. Drenos[2,3], K. Li[1,3], J.A. Cooper[3], J. Palmen[3], P.B. Munroe[1], P. Deloukas[1], P.J. Talmud[3], S.E. Humphries[3].* 1) Clinical Pharmacology, Queen Mary University of London, London, United Kingdom; 2) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 3) Institute of Cardiovascular Sciences, University College London, London, United Kingdom.

   Genome-wide association studies have found many loci associated with complex diseases and traits. These studies usually identify SNPs that are proxies for the causal variant due to strong linkage-disequilibrium (LD) that is present in the human genome. Fine-mapping may aid in the reduction in the number of potential causal variants, particularly in conjunction with trans-ethnic fine-maping studies, but this may still not identify the functional variant. The majority of GWAS associations occur outside of coding regions and are therefore likely to have roles in modulating gene regulation. We previously developed a high-throughput method, termed FAIRE-gen, to localize potential causal variants using FAIRE-enrichment of open chromatin followed by allelic quantification using fine-mapping genotyping chips. Allelic imbalance in heterozygotes following enrichment of accessible chromatin can identify variants that influence chromatin structure and therefore regulatory potential. To address the localization of functional regulatory variants, we have integrated allele-specific chromatin signatures with public eQTL databases and regulatory annotation projects (ENCODE and Roadmap Epigenomics) with GWAS-associated variants using the GWAS catalog and 1000 Genomes Project to examine those in strong LD. Variants demonstrating potential for functionality using these tools were examined using *in vitro* methods, and revealed strong potential for effects on transcription factor binding and reporter gene expression. Using this methodology we demonstrate the identification of potentially functional variants for cardiovascular and autoimmune traits.

714F

A study of the effect of dietary behaviour on DNA methylation in children in the Avon Longitudinal Study of Parents and Children (ALSPAC). *M. Ward[1,2], T. Gaunt[1,2], C. Relton[1,2,3].* 1) Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, U.K; 2) School of Social and Community Medicine, University of Bristol, Bristol, U.K; 3) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, U.K.

Background: Previous studies have found evidence of associations between DNA methylation and various aspects of diet, including fruit and vegetable, alcohol, B vitamin and protein intake and general nutritional status [1, 2]. Methods: We investigated the relationship between DNA methylation derived from peripheral blood and dietary patterns in 7 year old children in ALSPAC. Dietary patterns were quantified using principal components generated from a food frequency questionnaire [3]. The study identified three distinct dietary patterns: 'junk', 'traditional' and 'health conscious'. We performed epigenome-wide association analyses (EWASs) on those three dietary principal components, adjusting for age, sex and BMI. Findings: CpG sites with p-values below a Bonferroni-corrected threshold of $p<1.04e-7$ were observed for the 'junk' and 'health conscious' principal components, with smallest p-value around 5e-17. No CpG sites reached this threshold for the 'traditional' principal component. Interpretation and future research: We have found evidence of associations between DNA methylation and dietary patterns in children. We plan to perform two-sample Mendelian randomisation [4] to investigate whether the observed associations between diet and DNA methylation represent causal relationships. References: [1] Haggarty P. The Proceedings of the Nutrition Society. 2013;72(4):363-71. [2] Zhang FF et al. The Journal of nutrition. 2011;141(6):1165-71. [3] Northstone K et al. European Journal of Clinical Nutrition. 2005;59:751-60. [4] Pierce B and Burgess S. American Journal of Epidemiology. 2013;178(7):1177-1184.

715W

Genetic variation influences the gene expression response to hypoxia and oxidative stress in human iPSC-derived cardiomyocytes. *M.C. Ward[1], N.E. Banovich[1], D.A. Knowles[2], A. Mitrano[1], J.K. Pritchard[3,4,5], Y. Gilad[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Computer Science, Stanford University, Stanford, CA; 3) Department of Genetics, Stanford University, Stanford, CA; 4) Department of Biology, Stanford University, Stanford, CA; 5) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

The basis for complex polygenic diseases such as cardiovascular disease (CVD) is not fully understood, however the importance of non-coding regions has been highlighted. Myocardial infarction, a prevalent sub-class of CVD, is caused by oxygen deprivation leading to damage of heart tissue (ischaemia). Re-oxygenation by reperfusion can result in further damage, however the severity of ischaemia-reperfusion (I/R) injury is known to differ between individuals. To develop an understanding of the genetic determinants underlying this difference in phenotypic outcome, we have established an *in vitro* I/R model in induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs). iPSCs from 15 genotyped Yoruba HapMap individuals were differentiated into *TNNT2* positive iPSC-CMs and maintained under normoxic and aerobic metabolism-promoting culture conditions. To mimic I/R injury, iPSC-CMs under normoxia (condition A) were subjected to hypoxia (B) followed by short (C)- and long-term (D) re-oxygenation. Global gene expression levels were determined by RNA-seq. A total of 5982 genes are differentially expressed between all four oxygen conditions combined and this includes, both genes known to be involved in the hypoxic and oxidative stress response pathways such as *VHL* and *TET1*, as well as novel targets. The secretion of the cell damage marker, lactate dehydrogenase (LDH), indicates that re-oxygenation of oxygen-deprived iPSC-CMs is cytotoxic, and that there is inter-individual variation in the degree of response. A combined analysis of gene expression levels and LDH release identifies co-expressed modules of genes associated with overall cell stress and highlights the role of the autophagy gene *GABARAPL2*. In order to map genetic loci associated with the gene expression response, recently developed methods were used to identify expression quantitative trait loci (eQTLs) from a small sample size. Over 500 genes with a putative *cis* eQTL in at least one condition were identified, as well as a number of genes with response eQTLs - i.e. eQTLs with a difference in effect size between conditions. These data suggest the importance of genetic variation in mediating the variable gene expression, and subsequent cellular stress response to hypoxia and oxidative stress in a disease-relevant cell type.

716T

Transcriptome analysis of human explanted hearts identifies cardio-myopathy-specific mechanisms. *M. Sweet[1], A. Cocciolo[2], D. Slavov[1], S.L. Graw[1], K. Jones[1], L. Mestroni[1], M.R.G. Taylor[1].* 1) Cardiovascular Institute and Adult Medical Genetics Program, University of Colorado, Aurora, CO; 2) Ospedali Riuniti and University of Trieste, Trieste, Italy.

Cardiomyopathies are diverse in their genetic and phenotypic architecture but have been proposed to evolve via a "final common pathway" towards heart failure. The goal of this study was to use RNA-sequencing (RNA-seq) and pathway analysis to identify differentially regulated pathways in 42 samples of left ventricular heart tissue from cardiomyopathy patients. We identified differentially expressed genes between ischemic cardiomyopathy (ICM, n=14) and dilated cardiomyopathy (DCM, n=28) as well as between DCM sub-phenotypes, arrhythmogenic DCM (aDCM, n=13) and non-arrhythmogenic DCM (naDCM, n=15). Pathway analysis was performed using Ingenuity Pathway Analysis (IPA). Between ICM and DCM, 878 genes were differentially expressed ($p < 0.05$, $q < 0.05$). Pathway analysis showed enrichment for mTOR signaling ($p = 4\times10^{-13}$), epithelial adherens junction remodeling ($p = 3\times10^{-10}$), and Rho signaling ($p = 1\times10^{-7}$) with functional enrichment in immune response ($p = 7\times10^{-7}$). These results agree with previous models of chronic myocardial infarction that demonstrate increased mTOR signaling, which is known to aid in myocardial remodeling, and inflammatory response to the site of myocardial injury, which requires immune cell motility via Rho signaling and infiltration via cell-cell junction remodeling. Sub-phenotyping of the DCM samples into patients with and without a history of ventricular tachycardia revealed 280 genes differentially expressed between aDCM and naDCM ($p < 0.05$, $q < 0.20$). Pathway analysis demonstrated downregulation of oxidative phosphorylation ($p = 1\times10^{-13}$) and genes involved in muscle contraction ($p = 3\times10^{-5}$) in aDCM versus naDCM. These results suggest that aDCM patients may have reduced ATP production and contractility, which could contribute to arrhythmogenesis as compared to naDCM. More broadly, these results suggest that a 'final common pathway' model may need to be amended based on biological differences among DCM sub-phenotypes in advanced heart failure.

717F

Endothelial activating *KRAS* mutations in brain arteriovenous malformation. *S. Nikolaev[1,2], S. Vetiska[3], X. Bonilla[1], S. Antonarakis[1], J. Fish[4], I. Radova-novic[3].* 1) HUG, Geneva, geneva, Switzerland; 2) Service of Genetic Medicine, University Hospitals of Geneva, Switzerland; 3) Department of Fundamental Neurobiology, Krembil Neuroscience Institute, University of Toronto, Canada; 4) Department of Laboratory Medicine and Pathology, Toronto General Research Institute, University of Toronto, Canada.

Brain Arteriovenous Malformations (BAVMs) are focal developmental disorders of the brain vasculature with a prevalence of about 1 in 100'000 people and are a leading cause of hemorrhagic stroke in young adults and children. A BAVM is composed of entangled and morphologically abnormal vascular channels. In order to reveal the genetic cause of sporadic BAVMs we have performed deep exome sequencing of 27 BAVMs and 18 matching blood samples. In the majority of BAVMs we have detected recurrent putative somatic mutations p.G12D and p. G12V at low variant allele frequency (VAF) varying between 0.9% and 4.2%. Droplet Digital PCR (ddPCR) for these and other KRAS driver mutations in 32 BAVM patients (including 27 EXOME-sequenced) confirmed all predicted KRAS variants and revealed new ones totaling to 75% of patients (15 p.G12D, 8 p. G12V and 1 p.Q61H). No KRAS mutations were detected in normal blood neither with the exome sequencing no with ddPCR. Endothelial and smooth muscle cell cultures from 3AVMs with KRAS mutations and from 3 normal brain vessels were established using CD31 and SMA markers, respectively. KRAS mutations from DNA and RNA sequencing were detected with VAFs 2.8 – 9.3 times higher than in original BAVMs in endothelial but were absent in smooth muscle or normal brain endothelial cells cultures. Moreover we show that KRASmut transfected HUVEC cell lines and AVM endothelial cells exhibit higher levels of phospho-ERK, but not phospho-AKT or phospho-p38 MAPK, as compared to normal endothelial cells. Lastly we demonstrate on HUVEC cells lines that VEGF stimulation results in high pospho-ERK, however treatment with Ras inhibitor Farnesyl transferase prevents ERK phosphorylation. This work reveals the causative mutations and proposes the mechanistic explanation for BAVM, thus categorizing this disease as a new type somatic RASopathies.

718W

Genetic analysis of transcription factors, *GATA4* and *NKX2-5* in patients with non-syndromic congenital heart defect from India. *R. Dixit[1], A. Kumar[2], B. Mohapatra[1].* 1) Dept of Zoology, Institute of Science, Banaras Hindu University, Varanasi, Uttarpradesh, India; 2) Dept of Pediatrics, Institute of medical Sciences, Banaras Hindu University, Varanasi, Uttarpradesh, India.

Congenital Heart Disease (CHD) is the leading cause of infant morbidity and mortality, occurring approximately in 1% of all live births. Transcription factors are known to govern heart development. Among them NKX2-*5* (MIM: 600584) and *GATA4* (MIM: 600576) are two critical cardiac specific transcription factors, those are most frequently associated with mutations in familial as well as sporadic CHD cases. Over past couple of decades around 74 mutations in *NKX2-5* and 69 mutations in *GATA4* have been identified globally. However no such studies have been conducted in Indian population so far and etiology of these defects in India is not precisely known. Therefore, in present study we sought to find the association of mutations in *NKX2-5* and *GATA4* among CHD patients from India and validating the observed variants by *in silico* and *in vitro* methods. These genes were sequenced in 280 CHD patients and 200 controls (400 chromosomes). Functional significance of observed genetic variants was analyzed by *in silico* (Polyphen, SIFT and Mutation Taster) and *in vitro* (western blotting, luciferase assay and immunostaining) approaches. By Sanger sequencing in *GATA4*, 5 novel sequence variants (c.23C>A; p.A8D, c.25G>A; p.A9T, c.383A>T; p.E128V, c.397A>T; p.S133C and c.682T>A; p.W228R) and 4 known variants (c.223G>T; p.A75S, c.1073G>C; p.S358T, c.1220C>A; p.P407Q and c.1064C>G; p.T355S) were identified in 20 CHD cases (7.1%). Simultaneously, *NKX2-5* revealed 3 novel non-synonymous genetic variants (c.182C>G; p.A61G, c.391G>A; p.E131K and c.443C>A ; p.A148E) and 2 known variants (c.284G>T; p.R95L and c.739C>G; p.P247A) in 5 CHD cases (1.78%). None of these were found in 200 ethnic matched controls (400 Chromosomes). *In silico* predictions showed that the mutated amino acid residues were conserved across the species and also predicted to be disease causing by various *in silico* tools. The mutant proteins showed *in vitro* functional deficits demonstrated by western blot, immunocytochemistry and transactivation assay of downstream ANF and cActin promoter activity. This study further expands the mutational spectrum of these two widely studied transcription factors, where novel mutations affecting the protein's function are identified in 25 (8.9%) non-syndromic CHD patients from Indian cohort. *In vitro* studies provide strong evidence of their role in disease pathology and encourage us to conduct more detailed animal model studies for understanding the molecular basis of disease mechanism.

719T

***SOX7* deficiency impairs embryonic vasculogenesis and epithelial-to-mesenchymal transition during atrioventricular endocardial cushion development.** *A. Hernandez-Garcia[1], M. Wat[1], R. Udan[2], A. Renwick[2], Z. Yu[1], C.A. Shaw[1], M. Dickinson[2], D.A. Scott[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030.

SOX7 is located on chromosome 8p23.1 in a region that is commonly deleted in individuals with complex cardiovascular malformations. SOX7 is highly expressed in vascular endothelial cells and in the endocardium but its role in these cells remains unclear. To determine if SOX7 plays a critical role in cardiovascular development, we generated standard and conditional *Sox7* knockout mice. *Sox7[-/-]* embryos die around E11.5 with signs of cardiac failure including pericardial edema and failure of vasculature remodeling. The same phenotype was observed when *Sox7* was ablated in endothelial cells using Tie2-Cre transgenic mice. Immunohistochemistry studies revealed normal levels of apoptosis and an increased level of endothelial cell proliferation in the absence of SOX7. Transcriptome analysis carried out at E8.5, revealed a set of genes involved in vasculogenesis whose expression is decreased in the absence of SOX7 including *Gja4*, *Gja5*, *Ptrf*, *Hey1*, *Dll4* and *Cxcr4*. We then investigated the role of SOX7 during endocardial cushion development. *Sox7[-/-]* embryos at E10.5 had markedly hypocellular endocardial cushions with severely reduced numbers of mesenchymal cells. This suggests a possible defect in epithelial-to-mesenchymal transition (EMT). These embryos also had hypoplasia of ventricular trabeculae. Rare *Sox7[-/flox]*;Tie2-Cre embryos that were recovered at E15.5 had pericardial effusions, vascular hemorrhage and ventricular septal defects. Our results suggest that SOX7 is an important regulator of vascular development and most likely plays a key role in the regulation of endothelial cell proliferation and specification. Our studies also suggest that SOX7 has an essential role in regulating atrioventricular canal and septal development and may act in a cell autonomous fashion to regulate EMT in the endocardium.

720F

LRP6 mutation causes coronary artery disease by activating noncanonical Wnt signaling. *R. Srivastava, J. Zhang, G.W. Go, A. Narayanan, T.P. Nottoli, A. Mani. Yale University, New Haven, CT.*

Mutations in Wnt signaling coreceptor LRP6 have been linked to coronary artery disease (CAD) in humans. Mice carrying *LRP6^R611C* mutation exhibit loss of vascular smooth muscle cell (VSMC) differentiation and aortic medial hyperplasia. *LRP6^R611C* mice on high-fat diet display dramatic obstructive CAD and exhibit accelerated atherosclerotic burden on LDLR knockout background. Pathway analysis showed that impaired LRP6/canonical Wnt activity leads to enhanced non-canonical Wnt signaling, diminishes TCF7L2 expression and increases activation of PDGF signaling. Hence, LRP6 mutations are on one hand loss of function mutations in the canonical and a gain of function mutations in the noncanonical Wnt signaling pathways. Strikingly, Wnt3a administration to *LRP6^R611C* mice improved LRP6 activity, reduced noncanonical Wnt activity, led to TCF7L2-dependent VSMC differentiation, and rescued post-carotid-injury neointima formation. Paradoxically, excess Wnt signaling has been also associated with diverse vascular diseases. Accordingly, wild-type mice administered with Wnt3a exhibited reduced VSMC differentiation and increased post carotid injury neointima formation. These changes were similarly associated with loss of LRP6 activity, reduced TCF7L2 expression and activation of noncanonical Wnt in VSMC. These findings indicate that Wnt3a maintains a delicate balance between canonical and non-canonical Wnt to sustain vascular health, while its excess availability or lack of response to its physiological levels can cause a range of vascular diseases. Our findings identify the underlying pathophysiogical processes that link altered Wnt signaling to diverse vascular diseases and identify new targets for their pharmacological therapy.

721W

PRDM6 mutations impair histone methylation and underlie nonsyndromic patent ductus arteriosus. *N. Li[1], E. Smith[1], M. Bahjati[2], H. Lynch[3], R. Lifton[1,4], A. Mani[1,4]. 1) Cardiovascular Research Center, Department of Internal Medicine, Yale University, New Haven, CT; 2) Isfahan University of Medical Sciences, Isfahan, Iran; 3) Department of Preventative Medicine, Creighton University School of Medicine,Omaha, NE; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT.*

Nonsyndromic patent ductus arteriosus (PDA) is one of the most common congenital heart defects (CHD), which is occasionally inherited as an autosomal dominant disorder with incomplete penetrance. Using combined genome-wide linkage analysis and whole-exome sequencing (WES), we identified independent mutations in PRDM6, a vascular smooth muscle cells (VSMC)-specific histone methyltransferase that transcriptionally suppresses contractile proteins. In vitro assays showed that the mutations cause loss of function either by intracellular redistribution of the protein and/or by alteration of its methyltransferase activities. Wild-type embryonic ductus arteriosus (DA) initially exhibited high expression levels of PRDM6, which rapidly declined postnatally to increase the number of VSMCs necessary for ductus contraction increased. This dynamic change suggests that PRDM6 plays a key role in maintaining VSMCs in an undifferentiated stage in order to promote their proliferation and that its loss of activity results in premature differentiation and impaired remodeling of the DA. PRDM6 knock out mice die prenatally from abnormal vascular development, which prompted generation of conditional KO mice to model the human disease. Our findings identify PRDM6 mutations as underlying genetic causes of nonsyndromic isolated PDA in humans and as a target for treatment of vascular disorders that alter vascular remodeling.

722T

Genome-wide association study of modifiers for tetralogy of Fallot in 22q11DS identifies variants in the *GPR98* locus on 5q14.3. *T. Guo[1], G. Repetto[2], D. McDonald McGinn[3], J. Chung[1], A. Bassett[4], E. Chow[4], E. Mlynarski[3], E. Goldmuntz[3], K. Devriendt[5], A. Swillen[5], D. Gothelf[6], S. Eliez[7], K. Coleman[8], A. Tomita-Mitchell[9], M. Cristina Digilio[10], B. Marino[11], B. Dallapiccola[12], N. Philip[13], C. Bearden[14], W. Hawuła[15], J. Vorstman[16], D. Heine-Suner[17], A. Roberts[18], T. Simon[19], T. van Amelsvoort[20], W. Kates[21], E. Zackai[3], T. Wang[22], B. Emanuel[3], B. Morrow[1].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Center for Human Genetics, Facultad de Medicina Clinica Alemana-Universidad del Desarrollo, Santiago, Chile; 3) Division of Human Genetics CHOP and Department of Pediatrics, U Penn, SOM; 4) Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 5) Center for Human Genetics, University of Leuven, Leuven, Belgium; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 7) Departement of Psychiatry, Office Médico-Pédagogique (OMP), University of Geneva School of Medicine, Geneva, Switzerland; 8) Children's Healthcare of Atlanta, Atlanta, Georgia, USA; 9) Division of Cardiovascular Surgery, Department of Surgery, Medical College of Wisconsin, Milwaukee, WI 53226; 10) Medical Genetics, Bambino Gesù Hospital, Rome, Italy; 11) Department of Pediatrics, La Sapienza University of Rome, Rome, Italy; 12) Dipartimento di Medicina Viale Regina Margherita, Rome, Italy; 13) Department of Medical Genetics, AP-HM and University of Mediterranee, Timone Children's Hospital, Marseille, France; 14) Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, CA, USA; 15) Department of Genetics, Polish Mother's Memorial Hospital - Research Institute, Łód%8;, Poland; 16) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, 3584, Utrecht, The Netherlands; 17) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 18) Department of Cardiology and Division of Genetics, Children's Hospital Boston, Boston, MA, USA; 19) M.I.N.D. Institute & Department of Psychiatry and Behavioral Sciences, University of California, Davis, California, USA; 20) Department of Psychiatry and Psychology, University of Maastricht, Maastricht, The Netherlands; 21) Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY, USA; 22) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA.

A total of 65% of 22q11.2 deletion syndrome (22q11DS, velo-cardio-facial/DiGeorge syndrome) patients have congenital heart disease (CHD), including aortic arch and/or intracardiac malformations. Some of the most serious anomalies include tetralogy of Fallot (TOF) or persistent truncus arteriosus (PTA). We hypothesized that common genetic variants could serve as modifiers of CHD risk, or subtypes of CHD. We performed a GWAS on 1,512 22q11DS subjects using Affymetrix 6.0 microarrays. Imputation was performed based on the 1000 Genome Project data and 6.5 M SNPs were retained for analysis. Logistic regression was used to adjust the first three principal components to assess associations in a discovery cohort of 798 cases and 491 syndromic controls. The controls had 22q11DS but had a normal heart and aortic arch as determined by echocardiography. Findings were validated in a second cohort of 223, 22q11DS subjects from a Chilean population (106 cases and 90 controls). Meta analysis of the two cohorts was performed and none of the loci reached genome wide significance (5×10^{-8}) between cases with CHD and controls. However, six loci reached a p value with suggestive association (1×10^{-5}); they are variants at 5q14.3 (*GPR98*), 4q32.3 (*SPOCK3*), 6p12.1 (*GLYATL3*), 8q24.3 (*ZC3H3*) and 14q24.3 (*POMT2*) and 16q22.3 (close to *PSMD7*). We then examined subtypes of CHD and found a GWAS significant SNP (3.19×10^{-8}) in an intron of *GPR98* with TOF. Variants in another intron of *GPR98* showed suggestive association with PTA ($p=1.81 \times 10^{-6}$). The *GRP98* locus is within 2 Mb of *MEF2C* (Myocyte-specific enhancer factor 2C), which is the closest functional gene known to affect heart development. This data indicates that variants in the 5q14.3 region may alter risk to subtypes of CHD in 22q11DS.

723F

The haploinsufficiency of *Csk* increases blood pressure by modulating sodium resorption and plasma volume. *J. Kang[1], S. Kim[1,2], J. Lim[1], S. Hwang[2], B. Oh[1].* 1) Biochemistry and Molecular Biology, Kyung Hee University, Seoul, South Korea; 2) Department of Chemical Engineering, HanKyong National University, Gyeonggi-Do, South Korea.

Blood pressure is a complex trait regulated by an orchestrated network of physiological pathways that involve modulation of vascular tone, heart rate, and blood volume. The *CSK* locus (15q24) is associated with blood pressure in human samples from various ethnic groups. it was recently reported that the deficiency of the *Csk* gene increases blood pressure through Src, a target protein of the Csk. In this manuscript we have investigated the signaling mechanisms of high blood pressure in *Csk* heterozygote mice. We have shown that high blood pressure in *Csk* heterozygote mice was decreased not only by PP2, an Src inhibitor, but also by losartan, an angiotensin II receptor inhibitor. To determine whether hypertension in *Csk* heterozygotes might be associated with a change in the heart rate or plasma volume, we measured both the heart rate and the plasma volume respectively by observing the electrocardiogram and by using Evans blue dilution in *Csk* heterozygote and wild-type mice. The plasma volume was significantly increased in the *Csk* heterozygote compared to the wild-type while the heart rate was not significantly different between two groups. The increased plasma volume was decreased by treatment of PP2 and losartan. Consistent with the plasma volume, the plasma sodium level was increased in the *Csk* heterozygote compared to the wild-type while the plasma potassium level was not changed in *Csk* heterozygotes. Interestingly, the plasma sodium level was significantly decreased only by PP2, but not by losartan. Our results suggest high blood pressure in *Csk* heterozygotes is induced by increased plasma volume and sodium level that are primarily controlled by the Csk/Src signaling pathway.

724W

Cytogenetic analysis of chromosomal abnormalities in patients with Myelodysplastic Syndrome from Central Brazil. *C.L. Ribeiro[1,2], I.P. Pinto[1,2], D.M.C. Cunha[2], I.B. Minasi[2,3], A.D. da Cruz[1,2,3,4], C.C. da Silva[1,2,3,4,5].* 1) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory (LaGene), Secretary of Goias State for Public Health (SES), Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil.

   Myelodysplastic Syndrome (MDS) is a group of heterogeneous clonal hematopoietic disorders characterized by ineffective myeloid differentiation, dysplasia, karyotypic abnormalities, increased risk for progression to acute leukemia, and frequent progression to acute myeloid leukaemia. MDS incidence has risen dramatically in recent years and is associated with advanced age, shorter telomeres, cancer chemotherapy with alkylating agents, radiation, and inherited syndromes related to abnormalities in DNA repair. Cytogenetic abnormalities are identified in greater than 50% of MDS. Chromosomal analysis can help better determine prognosis and classifying the MDS in different risk groups. The aim of this study was to identify cytogenetic alteration in blood marrow using the karyotype analysis in 10 patients with MDS from the public health service in Central Brazil. Conventional cytogenetics analysis by GTG banding using the software IKAROS® (Metasystems Corporation, Germany) showed a diversity of structural and numerical chromosomal alteration. The cytogenetic abnormalities at diagnosis were identified in 40% of the patients, including deletion at 17p chromosome, trisomy at 21 chromosome, monosomy at X chromosome, gain in the heterochromatic region at 1 chromosome, and polyploid cells. The p53 gene, located at 17p13.1 region, encodes a tumor suppressor protein that responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, and may be involved in the process of neoplastic transformation of hematopoietic cells. On the other hand, the trisomy of the 21 chromosome has been observed in low percentage of cases of MDS, usually in the advanced stages with rapid transformation leukemic. The other chromosomal alterations identified could be intrinsically related to the neoplastic process and could represent both the cause and the consequence of ineffective hematopoiesis in MDS. Cytogenetic analyses are presumed to be strongly predictor of clinical outcome in MDS. Furthermore, the comprehension of pathogenetic mechanisms of MDS can contribute to a better understanding of biological events involved in the initiation, progression and promotion of this disease and help to determine the prognosis, collaborating with an effective, early and individualized treatment.

725T

Aneuploidy detection using multiplex ligation-dependent probe amplification (MLPA), in bone marrow from pre-B ALL patients. *A. Vazquez-Reyes[1,4], L. Bobadilla-Morales[1,2], C. Barba-Barba[2], G. Macías-Salcedo[2], J.R. Corona-Rivera[3], F. Sánchez-Zubieta[5], A. Corona-Rivera[1,2].* 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo "Instituto Dr. Enrique Corona", CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Unidad de Citogenética, Hematología y Oncología Pediátrica, División de Pediatría, Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 3) Servicio de Genética División de Pediatría, Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 4) Doctorado en Ciencias en Biología Molecular en Medicina, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, México; 5) Hematología y Oncología Pediátrica, División de Pediatría, Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México.

   Acute lymphoblastic leukemia (ALL) is a malignancy from T cell and B cell progenitors that proliferate in bone marrow, blood and other organs. Three-quarters of the affected patientswith ALL, show numerical or structural chromosomal alterations, which are important factors in leukemogenesis. The numerical alterations are determined by the golden standard: the karyotype. Nevertheless, there is a new molecular technique that can be used to determine the numerical alterations in ALL: MLPA. The Multiplex Ligation-dependent Probe Amplification (MLPA), is a method based in the multiplex PCR, that detects abnormal changes in copy number of ~45 different genomic DNA sequences. There are studies that have used this technique to detect chromosomal gain or losses in products of conception, copy number alterations in adult malignancies, and partial or complete deletions of genes involved in pre-B ALL. However, there are no studies that have used the MLPA technique to assess chromosomal gains and losses in patients with pre-B ALL. The preliminary aim of this study is to identificate aneuploidy for the first time using the MLPA technique and correlates the results with the routine karyotype analysis from LLA pre-B patients. We analyzed 15 bone marrow samples from children diagnosed with pre-B ALL, from the Unidad de Citogenética, Servicio de Hematología Oncología Pediátrica, Nuevo Hospital Civil "Dr. Juan I. Menchaca". The DNA extraction was carried out using the QIAmp© DNA Blood Mini Kit. The chromosomal gains and/or losses were detected by the SALSA MLPA P036 Subtelomeres Mix 1 probemix, using the MLPA technique. Finally, the MLPA results were correlated with the karyotype results, using the Pearson's correlation test. We analyzed 5 samples from pre-B ALL children with normal ploidy and 10 children with aneuploidy. The chromosomal number detected by MLPA and karyotype matched in 5/5 samples with normal ploidy (r^2=1.000, p<0.05); and the results from MLPA and karyotype matched in 7/10 samples with aneuploidy (r^2=0.8330, p<0.05). In the present study, MLPA technique was capable to detect 7/10 samples with aneuploidy and 5/5 samples with normal ploidy. Preliminary results suggest that MLPA could be a useful tool to aneuploidy identification in children with pre-B ALL, in order to establish a better treatment and a better prognosis.

726F

Over expression of IQGAP1 in patients with mutations in the Armadillo region of APC gene can be a marker of incidence of metastatic Colon Cancer. *H. Azimi[1,3,4], A. Razavi[1,3,8], S. Ghavimi[2], N. Anderson[3,4], J. Brown[1,4], S. Emamian[5], S. Singh[1,6], A. Kolahi[7], A. Afsari[2], R. Jaasbi[8].* 1) Genetic Informatic, GenomicsOnCall, Washington, DC; 2) Howard University Hospital, Department of Medicine, Washington, DC; 3) PsychoGenome, Ottawa, ON, Canada; 4) Carleton University, Department of Integrated Sciences Ottawa, ON, Canada; 5) Indiana University, Bloomington, IN; 6) Howard University School of Medicine, Washington, DC; 7) Shahid Beheshti University School of Medicine, Tehran, Iran; 8) Islamic Azad University, Department of Genomic and Genetics, Iran.

　Background: A large proportion of mutations occurring in patients with colon cancer involve genes encoding proteins of the APC, WNT, EGFR and p53 pathway. The involvement of the majority of these mutations in colon cancer progression is not well known. Next Generation Targeted Exome Sequencing (NGTES) of patients with metastatic colorectal carcinogenesis (mCRC) will lead to the discovery of specific driver mutations that are important in this process. **Objective:** To evaluate patients with mCRC for expression of IQGAP1 and investigate if different mutation sites on the APC protein correlate with the higher incidence of metastasis. **Method:** A total of 100 patients (48 males and 52 females) with mCRC (T2-T3) and matched normal were obtained from the CRC database held at Psychogenome Laboratories in Ottawa, Canada. The age range was from 38 to 92 with a mean of 68. Real-time reverse transcriptase PCR was conducted on frozen tissues of mCRC. Next Generation Targeted Exon Sequencing (Ion Torrent) was performed to analyze mutations in the APC genes of all patients. Results were matched with exon sequencing data of patients with no mCRC [100 patients (53 male and 47 female)] Stage 1 and 2 CRC between age of 43 and 86 with median age of 65. Results were compared to TCGA and published known mutations, COSMIC, dsSNP and false positive was driven out by parallel sequencing by using MuTect. **Result:** Overall 120 different mutations (12 novel and 108 known) were found. There were 48 non-synonymous mutations, 35 synonymous, 12 frameshift, and 25 stop-gain mutations. In mCRC patients with overexpression of IQGAP1, the majority of mutations were in Exon 4 and 5 with 88% of mutations occurring at ARM region of APC protein, between ARM and first 15AA repeat region. Samples that did not have overexpression of IQGAP1 with no mCRC had 15% of mutations in the ARM and 15AA repeat region compared to those who had mCRC. **Conclusion:** Using NGTES, we examined the expression of IQGAP1 in mCRC tissue and found that IQGAP1 levels highly correlate with the degree of malignancy and metastasis in CRC patients. We found that patients with mutations in the ARM region of APC, leading to an upregulation of IQGAP1, have a worse prognosis compared with patients with mCRC with fewer or no mutations in the same region. This is the first Next Generation Sequence analysis of patients with mCRC that has shown an upregulation of IQGAP1 with a majority of the mutations occurring in the ARM region of APC.

727W

Development and validation of a clinical bioinformatics pipeline for comprehensive pediatrics cancer panels. *K. Cao[1], F. Chang[1,2], C. Wu[1], F. Lin[1], M. Li[1,2], M. Sarmady[1].* 1) Pathology, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Dept. of Pathology & Laboratory Medicine, University of Pennsylvania, Philadelphia PA.

　Molecular tumor profiling has been increasingly utilized by oncologists to guide clinical decision making. Next generation sequencing (NGS) has emerged as the method of choice to analyze a wide range of genes known to be associated with various cancers. We have recently launched multiple NGS panels targeting childhood cancers including 13 hereditary cancer panels (128 genes), one hematological cancer panel (99 genes), and one solid tumor panel (237 genes). Development of bioinformatics pipelines that can accurately and efficiently transform numerous genetic alterations in the raw data to a few clinically meaningful somatic/germline variants is an essential component of genomic profiling. To support the cancer diagnostic needs, we have designed, implemented and validated a clinical bioinformatics pipeline for the cancer panel analysis.　Fastq files generated by MiSeq sequencer are first aligned to GRCh37 (hg19) assembly. Variants are called by five variant calling tools to maximize the detection rate. Raw variants are pre-filtered using strand ratios and allele fractions. Variants called by different tools are then merged, annotated and filtered using multiple public, commercial, and in-house databases. Copy number variations (CNVs) and loss of heterozygosity (LOH) are called by using our home-brew software that integrates multiple tools (e.g. VarScan and CNVKit) together. CNVs and LOH calls are calculated and visualized using NGS reads and pre-selected SNPs built in the panels.　To assess the performance of our pipeline, we validated the pipeline using replicates of five cancer cell lines with serial dilutions, three known positive and ten clinical samples for each of the three panels. Our results showed 100% concordance to that of commercially available pipelines on variants with ≥5% allele frequency for somatic and 10% for hereditary panels. Our pipeline allows analyzing multiple samples in parallel, which greatly increases analysis efficiency. To evaluate reproducibility of the pipeline, three libraries were prepared using HapMap sample NA12878. Our results showed 100% intra- and inter-run reproducibility for all panels.　In summary, we have developed a comprehensive bioinformatics pipeline for variant and CNV calling, annotation and filtering for our custom-designed cancer NGS panels. The validation results demonstrate 100% accuracy and reproducibility of the pipeline in identifying cancer associated mutations and CNVs with increased efficiency.

728T

Implementing molecular barcode counting into a comprehensive integrated targeted sequencing portfolio and bioinformatics pipeline. *E. Lader, Z. Wu, S. Tian, Q. Jiang, Y. Wang, J. DiCarlo.* Qiagen Biological Research Center, 6951 executive way, Frederick MD 21703.

We recently completed the development of a comprehensive targeted sequencing portfolio making use of random molecular barcoding and an integrated bioinformatics package which makes use of a machine learning algorithm for mutation calling. These applications include targeted somatic mutation and in/del detection in DNA, fusion transcripts and other mutations and rearrangements (both known and novel) in RNA, high throughput targeted quantitive RNA expression analysis, and miRNome profiling. For RNA and miRNA quantification and characterization, molecular barcoding allows extremely accurate and unbiased quantification of a few to up to thousands of targets simultaneously. For mutation, rearrangement, and fusion calling, we employ a combination of single primer extension and molecular barcoding to achieve extremely uniform coverage and high sensitivity mutation detection, even in challenging FFPE and biofluid samples. The highly integrated sample to insight workflow and analysis package, and flexibility of panel design.

729F

Neuroblastoma with *MYCN* amplification and ALL pre-B with 46,XY,t(4;11). Case report. *S. Brukman[1,5], H. Romo-Rubio[4], L. Bobadila-Morales[1,2,3], G. Macias-Salcedo[2], A. Marquez-Mora[2], M.C. Almodovar-Cuevas[2], R. Corona-Rivera[1,2,3], F. Sanchez-Zubieta[4], A. Corona-Rivera[1,2].* 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo, CUCS, Universidad de Guadalajara; 2) Unidad de Citogenética, Servicio de Hemato-Oncología Pediátrica, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca". Guadalajara, Jalisco, México; 3) Servicio de Genética División de Pediatría, Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 4) Hematología y Oncología Pediátrica, División de Pediatría, Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 5) Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, México.

Common childhood malignancies include leukemias (30-40%), brain tumors (20%) and lymphoma (12%) followed by neuroblastoma (NB)(1). Acute lymphoblastic leukemia (ALL) and ALL pre-B cell represent 80-85%; the most frequent cytogenetics abnormalities are hyperdiploidy-25% and MLL-5%. The most common extracranial cancer in childhood is Neuroblastoma, with approximately 15% of deaths during childhood. The amplification of the *MYCN* oncogene locus on 2p24-25, occurs in approximately 25% of primary NB, and it is a biologic marker for an unfavorable prognostic in patients. The finding toghether of this two neoplasic abnormalities in the same patient is unusual. The objective is present a case of patient with two different primary cancer tumor with their respective laboratory diagnostics. **Clinical report:** 23 months old male, at 1 year old presented stiffness in lower limbs, MRI reported spinal cord compression by tumor diagnosed as Neuroblastoma in Biopsy report. Immediately start treatment with radiotherapy and chemotherapy. 11 months after treatment presented weakness, hematomas in lower libs, hyporexia and fever; leukocytosis of 222,200 10^3/μL, lymphocytes 83% and platelets 12.4×10^3. FISH positive for *MYCN* gene amplification: nuc ish(MYCN amp)[100]. In bone marrow aspirate was found 100% of blastic lymphocytes L1 by FAB, inmunophenotype Pre-B (CD19, CD79, MLL+), Karyotype: 46,XY,t(2;?)(q36;?),t(4;11)(q21;q23)[3]/46,XY[5]. FISH: nuc ish(MLLx2)(5´MLL sep 3´MLLx1)[169/200]. **Conclusion:** Germline MLL locus 11q23; MLL gene is rearranged most often in de novo infant leukemia, which results from t(4;11)(q21;q23) associated with a dismal prognosis, therapy refractoriness and central nervous system (CNS) infiltration. The simultaneous occurrence of 2 different kinds of neoplasms in a child is infrequent, just one case are reported in the literature by D'angelo P. et. al. Common mechanism may be involved in this two early on-set neoplasias. .

730W

Mutation detection in circulating tumor DNA in early stage cancer samples. *A. Chen, M. Anselmo, J. Jaboni, S. Li, G. Braunstein, A. Kammesheidt.* Pathway Genomics, San Diego, CA.

Background: Circulating Tumor DNA (ctDNA), a subset of cell-free DNA (cfDNA), are small DNA fragments in the blood stream released by tumor cells, that reveal somatic mutations from existing tumors. Pathway Genomics CancerIntercept™ is a CLIA validated, laboratory-developed blood test that can be used by physicians as a screening tool for the detection of biomarkers associated with specific cancer types in patients without a cancer diagnosis but at high risk for developing a cancer or as a test to monitor patients who have been diagnosed with cancer. The assay detects the presence of 96 mutations in nine genes (*BRAF, CTNNB1, EGFR, FOXL2, GNAS, KRAS, NRAS, PIK3CA,* and *TP53*). In this study, we sought to determine the frequency of ctDNA detected in stage I and II cancer patients. Methods: To date this study series has analyzed 85 plasma samples from individuals with stage I (n=81) and stage II cancers (n=4). Cancer types included breast, colorectal, gastric, lung, ovarian, pancreatic, thyroid, head and neck, and prostate cancer. Clinical information such as grade, TNM, tumor size, lymph node status was available for most samples. For each sample, cfDNA was isolated from 1-2 ml of frozen plasma, and analyzed using the CancerIntercept™ assay. Results: In this series, all the breast, ovarian, prostate, and head and neck cancer samples were screened negative. A total of 15 of the 85 stage I and II specimens (17.6%) had at least one positive mutation. The positive samples included 5/24 gastric, 3/7 pancreatic, 3/9 colon, 2/10 lung, and 2/11 thyroid cancer specimen. Three cancer samples (one each from pancreatic, lung and colon) had two mutations. The 18 mutations detected in the stage I and II cancers included 15 unique mutations across *BRAF, EGFR, GNAS, KRAS, PIK3CA,* and *TP53*. Conclusion: In this series across multiple cancer types, the combined ctDNA mutation rate in the stage I and II cancers detected was 17.6%. Three of the 15 positive samples had more than one detectable mutation in this screening assay.

731T

Recurrent mutations of *BRCA1* **in Algerian triple-negative breast cancer patients: Implications for genetic screening criteria selection and counseling.** *F. Cherbal[1], H. Gaceb[1], C. Mehemmai[1], K. Yatta[1], M. Belkacemi[1], A. Aguentil[1], R. Bakour[1], K. Boualga[2].* 1) Unit of Genetics, LMCB, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Radiation Therapy Services, Anticancer Center of Blida, Blida, Algeria.

Background Breast cancer is currently the leading cause of cancer deaths among Algerian women. Triple-negative breast cancer (TNBC) shows substantial overlap with basal-type and BRCA1-related breast cancers. In the present study, we screened for *BRCA1* germline mutations in 141 women with TNBC including all exons where a mutation was previously found in Algerian population (exons 2, 3, 5, 11, 18, 20). **Materials and Methods** 141 TNBC patients and their families were referred through several public hospitals and private medical clinics which provide oncology services throughout Algeria. *BRCA1* gene was screened by PCR-direct sequencing in 56 TNBC patients with family history of breast cancer and 85 sporadic TNBC patients including all exons where a mutation was previously found in Algerian population (exons 2, 3, 5, 11, 18, 20). *In silico* analyses have been performed using different bioinformatics programs to individualize genetics variations that can disrupt the *BRCA1* function. **Results** The analysis of DNA samples of 141 women with TNBC revealed that 10 patients carried pathogenic germline mutations in *BRCA1* gene and unclassified variants (UVs) (7%), respectively. Most of TNBC patients with mutations had family history of breast and were diagnosed at age ≤ 45 years. 4 distinct pathogenic mutations: c.83_84delTG, c.181T>G, c.798_799delTT and c.2125_2126insA and two unclassified variants p.His-41Arg and p.Arg1753Gly, respectively, were identified in this study. The recurrent *BRCA1* mutation c.83_84delTG has been identified in 3 patients with a frequency of 2.1%. The c.181T>G/p.Cys61Gly mutation has been detected here in one patient. To date, C61G mutation has been detected already in two Algerian families. The *BRCA1* mutation c.798_799delTT has been identified in a young TNBC patient diagnosed at age 34 years. Interestingly, the *BRCA1* c.2125_2126insA deleterious mutation was detected for the first time in 3 patients (and in some first relatives of two patients) with a family history of breast cancer and/or prostate cancer.The two *BRCA1* UVs H41R and R1753G were found to be low and high probability pathogenicity by AGVGD (C25 and C65), deleterious and disease causing by SIFT and Mutation Taster, respectively. **Conclusions** Recurrent mutations of *BRCA1* have been detected in TNBC patients diagnosed at earlier age.TNBC immunophenotype should be considered as an additional criterion for genetic screening in Algerian women with early onset breast cancer.

732F

Prognostic value of syntenin in colorectal adenocarcinoma. *Y. Cui[1], J.H. Shin[2], E.S. Jung[3], S.T. Oh[4], J.G. Kim[4], S.Y. Kim[1], S.H. Lee[3].* 1) Department of Pathology, College of Medicine, The Catholic University of Korea,, Seoul, Seoul, South Korea; 2) Department of Hospital Pathology, St. Paul's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 3) Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 4) Department of Surgery, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea.

Background: Syntenin is reported to modulate migration and invasion in several tumors. To date, the expression has not been clearly elucidated. In this study, we compared syntenin expression with various clinicopathological parameters and molecular markers to evaluate its clinical significance in colorectal adenocarcinoma. **Method:** We evaluated the immunoreactivity of syntenin using tissue microarray in 202 surgical specimens of primary colorectal adenocarcinoma from patients consecutively treated between 2008 and 2011 at Seoul St. Mary's Hospital, The Catholic University of Korea. The relationship between syntenin expression and various clinicopathological parameters and molecular markers was analyzed. **Results:** Of the 202 cases, 28 cases show cytoplasmic syntenin positivity in tumor cells. Syntenin expression is significantly associated with increased lymphatic invasion (p=0.038), perineural invasion (p=0.005) and higher nodal stage (p=0.013). In addition, syntenin expression has weak positive correlation with syndecan expression (γ=0.167, p=0.019). However, syntenin expression is not significantly correlated with tumor stage, metastasis, vascular invasion, tumor differentiation, K-ras mutation and/or EGFR expression status. **Conclusions:** The expression of syntenin may be of clinical value in colorectal adenocarcinoma and may help in identifying aggressive forms of colorectal adenocarcinoma.

733W

Identification of prostate cancer driver mutations and transcriptomic profiling from cultured circulating tumor cells. *B.R. Downie[1], B. Adams[1], C. Ryan[2], J. Lim[1].* 1) Xcell Biosciences, San Francisco, CA; 2) UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA 94115.

Biopsies are labor-intensive, invasive procedures that capture tumor tissue for downstream processing. While liquid biopsy technologies relying on amplification of cell-free DNA particles has promised to assist clinical decisions, no technology to date has enabled the direct capture and expansion of living cancer cells due to the extremely low fraction of circulating tumor cells (CTCs) in blood. We present here a cell culturing workflow for the isolation and expansion of CTCs followed by next-generation sequencing and analysis. As proof of principle, we successfully cultured 4-22 blood-derived colonies of cells per prostate cancer patient under tumor-relevant culturing conditions. RNA from selected colonies was extracted, sequenced, and profiled for prostate cancer relevant signaling pathways. Disruptive driver mutations with established pathogenicity in ClinVar could be identified in the DNA repair genes BRCA1, BRCA2, and PARP1. Additionally, androgen receptor (AR) downstream target genes co-express together with constitutively active AR splice isoforms. These data, together with expression of neuroendocrine differentiation marker genes are used to generate a composite score which could be used to better inform clinical decisions. Together, these results demonstrate that colonies of tumor-derived cells can be expanded and interrogated using a range of integrative next-generation technologies.

734T

Identification of somatic mutations and copy number variations in multiple myeloma using next generation sequencing. *H. Fukushima[1,2], Y. Sasaki[2], M. Tamura[2], H. Ikeda[3], K. Ishiguro[4], T. Tokino[2], A. Sakurai[1].* 1) Department of Medical Genetics, Sapporo Medical University, Sapporo, Hokkaido, Japan; 2) Department of Medical Genome Sciences, Research Institute for Frontier Medicine, Sapporo Medical University, Sapporo, Hokkaido, Japan; 3) Department of Hematology, Sapporo Medical University, Sapporo, Hokkaido, Japan; 4) Department of Gastroenterology, Sapporo Medical University, Sapporo, Hokkaido, Japan.

    Introduction of novel agents such as bortezomib, thalidomide, and lenalidomide have remarkably improved the overall survival in multiple myeloma (MM [MIM 254500]), but the prognosis of patients with relapse and refractory MM remains poor. The comprehensive analysis of genetic alterations in tumor by next generation sequencing can allow for the prediction of drug resistance and facilitate improvements in the treatment of MM.We tried to define genetic changes specific for each myeloma sample using targeted semiconductor sequencing technology. DNA was extracted from magnetic bead-enriched bone marrow CD138-positive malignant plasma cells from 11 cases of MM, and CD138-negative cells were used as matched non-tumor cells. Forty nanograms of DNA were used for multiplex PCR amplification with an Ion Ampliseq Comprehensive Cancer Panel that offers targeted coverage of all exons in 409 tumor suppressor genes and oncogenes frequently cited and frequently mutated in human cancers. (covered regions: 95.4% of total). We sequenced 15,992 regions which obtained more than 1.5 megabases of target sequence. Each sample underwent on average 8.3 million sequencing reads after quality filtering. The mean read depths were 539x, and >95% of targeted bases were represented by at least 20 reads. The average number of non-synonymous mutations detected per patient was 6.1 (range 3-11). The copy number variations in which segments of the genome can be duplicated or deleted from sequencing data were also detected. Somatic mutations were found in known MM-associated genes, including *TP53* and *NRAS*. We also identified novel recurrent alterations (single nucleotide variants, indels and copy number variants). Pathway assessment has shown that somatic aberrations within MM genomes are mainly involved in several important pathways, including cell cycle regulation, RTK–MAPK-PI3K and NF-kB. We performed targeted next-generation sequencing for rapid (2 days), standardized, and cost-effective gene analysis of malignant plasma cells from patients with MM. This targeted next generation sequencing may allow clinically oriented variant screening in MM.

735F

Clonal hematopoiesis of indeterminate potential (CHIP) and pathobiology of myelodysplastic syndromes (MDS). *B.B. Ganguly.* MGM New Bombay Hospital, Navi Mumbai, Maharashtra, India.

    After the detection of *BCL2* and *BCR-ABL* rearrangements in non-malignant individuals in 1995, copy number variations were reported in 5q/11q/17p/20q in hematopoietic stem cells (HSCs) in 2012 followed by identification of point-mutations in genes involved in epigenetic mechanism in non-pathogenic apparently healthy elderly individuals in 2014-2015. The advent of genome/exome sequencing techniques has heralded the detection of landscapes of mutations in RNA-splicing/DNA-methylation/epigenetic mechanism/chromatin modification/transcription factors/cohesin complex, etc. most of which are passengers/initiating lesion. Age-related increase of such clonal mutations, mostly included *DNMT3A, TET2* and *ASXL1* with a significant prevalence of *DNMT3A*, has led to speculate that accumulation of these clonal mutations might be the cause of high morbidity and mortality in elderly with higher risk of hematologic neoplasia. These dormant mutations or clonal hematopoiesis in idiopathic nonpathogenic cytopenic patients (idiopathic cytopenia of undetermined significance, ICUS) has hypothesized the clinical entity of clonal hematopoiesis of indeterminate potential (CHIP). The presence of ≥2 mutations with significant variant allele frequencies has resulted in neoplastic transformation in ~1%/year of 10% of elderly of >70 years age detected with clonal mutations, who are also at risk for cardiovascular and psychopathic disorders. The cooperation/interaction of MDS-specific driver mutations and/or subsequent acquisition of non-random mutations through clonal expansion results in self-renewal of hematopoietic stem cells and perturbation of cellular differentiation/maturation. The knowledge gained on clinical significance of CHIP has posted an urge on monitoring elderly people visiting clinicians for ICUS with a view to lowering/preventing the onset of overt malignancy.

736W

Molecular diagnosis of lung cancers. *K. Iravathy Goud[1], M. Kavitha[1], M. Adi Maha Lakhsmi[1], V. Ravi[1], D. Sagarika[1], P. Madhuri[1], P. Vijayananad Reddy[2], G. Swarnalatha[3], P. Michelle[3], S. Meenakshi[3], M. Tejal[3], A.V.S.N. Anuradha[3].* 1) Molecular Biology & Cytogenetics Laboratory, Apollo Hospitals, Hyderabad, India, India; 2) Oncology Department, Apollo Hospitals, Hyderabad, India; 3) Histopathology Department Apollo Hospitals, Hyderabad, India.

Abstract Lung cancer is the leading cause of cancer-related deaths in the in the world. Initially molecular tools which were used as research tools are now been applied to the clinical use. These molecular tests help in identifying those patients who respond to targeted therapy; and reduce side effects of ineffective treatments. Current study focuses on the significant approach of detection of epidermal growth factor receptor (EGFR) mutations in exon 18-21 and anaplastic lymphoma kinase gene (ALK) rearrangement analysis using molecular methods such as pyrosequencing and florescence in-situ hybridization (FISH) on formalin fixed paraffin embedded (FFPE) of 267 lung cancer patients. **Methods** In this study, paraffin-embedded tissue blocks of confirmed NSCLC by two pathologists which had at least 50% of neoplastic cells per sample were collected from 267 consecutive patients (185 males and 82 females; with mean age 56 years (25-84)) during December 2009 – March 2015 and referred to the Department of Molecular Biology and Cytogenetics, Apollo Health City, Hyderabad, India, for EGFR mutation testing by pyrosequencing & ALK-1 gene rearrangement by FISH analysis. Informed consent was obtained from all the patients. Fixed samples were processed routinely and stained with hematoxylin and eosin (HE) for microscopic diagnosis. If necessary, additional IHC was performed to verify the microscopic diagnosis. Paraffin block sections of 3 μm thick were taken for FISH on frosted slides and 6 μm of 6 sections were deposited in 1.5ul eppendorff tubes for sequencing analysis and 3 **Results** Among the 267 patients 11 were positive for ALK-1 gene rearrangement. *EGFR* mutations were detected in 55 patients. The most common mutations found were exon 19 (56.36%), exon 21 (27.27%), exon 20 (3.63%) and exon 18 (9.09%). The influence of gender, non-smoking, and histological type on the *EGFR* mutations showed increase in female group. **Conclusions** Molecular techniques have been invariably used in many hematological and solid tumor diagnosis and treatment. Among such cancers, lung carcinoma is one such disease where molecular techniques are used to detect molecular alterations. The use of molecular techniques would enable in effective identification of lung cancer patients who would benefit from targeted anti-ALK therapies. Mutation testing is mandatory in lung carcinomas for the management of NSCLC. **Key words:** EGFR mutations, ALK FISH.

737T

A novel *BRCA2* mutation identified in a family with male and female breast cancer patients. *M. Islam, A. Bisgin.* Department of Medical Genetics, Balcali Clinics and Hospital, Cukurova University Faculty of Medicine, Adana, Turkey.

Introduction: Certain predisposition factors such as *BRCA1* and *BRCA2* mutations play a pivotal role in developing familial breast cancer both in males and females. More than 1,800 mutations in *BRCA2* gene have been identified till date. Many of them are directly associated with an increased risk in developing not only breast cancer but also ovarian and prostate cancers. Although all carriers do not show the symptoms for cancer, it exhibits a potential risk while transferring the mutation among generations. Genetic diagnosis to identify the mutations affecting the populations help in risk assessment of patients and individuals that undergo genetic screening. In this study, a novel disease-causing *BRCA2* mutation was reported in a Turkish family linked to breast cancer in both women and man. **Methods:** A 41-year-old Turkish female diagnosed with breast cancer was referred to our Medical Genetics Department. Genetic testing for *BRCA1* and *BRCA2* was performed by using next-generation sequencing (Illumina MiSeq). Then the variant in the sequence was characterized by in-silico analysis (Mutation Taster, SIFT, and PolyPhen-2). **Results:** The clinical history of the patient lead us to consider the possibility of inherited *BRCA1* and *BRCA2* mutations among two sisters, father, mother, two aunts and a cousin diagnosed with breast cancer. All the breast cancer patients in the family presented a novel heterozygous *BRCA2*.p.P26L(c.77C>T) mutation. **Conclusion:** To the best of our knowledge, this mutation has not been previously reported in Breast Cancer Information Core (BIC) database. It is, therefore, critical to identify and report this family with male and female breast cancer patients to offer appropriate clinical assessment and to analyze the mutation-associated risks.

738F

Germline mutations in 94 cancer predisposition genes among large epithelial ovarian cancer cohort. *R. Janavicius[1,2], V. Rudaitis[1,2], N. Archipova[1], O. Mickeviciute[1], L. Griskevicius[1].* 1) Vilnius University Hospital Santariški%1; Clinics, Vilnius, Lithuania; 2) State Research Institute, Innovative Medicine Center, Vilnius, Lithuania.

   INTRODUCTION. Recent advances in massive parallel DNA sequencing led to the generation and application of various cancer predisposition genes (CPG) panels, improving the stratification of genetic cancer subtypes for better management and targeted therapy. To determine the utility of multigene testing approach we assessed the frequency of pathogenic variants in 94 pan-CPGs in a large epithelial ovarian cancer (EOC) cohort from a homogenous population of Lithuania. METHODS. Patients with EOC (n=574) w/wo family history (FH) were recruited through single hospital center (VUHSK HOTC) and germline DNA was sequenced and analysed using Illumina Miseq multigene panels of 94 CPGs, including *BRCA1* and *BRCA2*. RESULTS. *BRCA1/2* mutations were identified in 158 patients (47%) - *BRCA1* – 141, *BRCA2* – 17 – and were present in 38,2% EOC cases w FH and 22,4% wo FH. Additional deleterious mutations were identified in 9 clinically actionable genes [*MLH1, PMS2, TP53, CHEK2, ATM, RAD51C, BRIP1, MUTYH, NBN*] (n=12; 6,6%). PT mutations were identified in 4 genes with no clear clinical actionability and association with EOC [*WRN, FANCC, FANCF, FANC*] (n=4; 2,8%). In patients w/o FH pathogenic BRCA1/2 mutations were found in 18,3% and 4,2% harboured mutations in other genes. CONCLUSION. The heritable component of EOC due to mutations in other genes than *BRCA1/2* genes is less prevalent and not affected by FH. Clinical utility for newer genes remains to be established. This work was suported by Lithuanian Research Council grant SEN18/2015.

739W

Assessment of the real-time PCR method for detection of *EGFR* mutation by using malignant pleural effusion samples from non-small-cell lung cancer patients. *J. Kim[1], S. Shin[2], Y. Kim[1], G. Yoo[1], K. Lee[3].* 1) Lab Med, Yonsei University Wonju College of Medicine, Wonju, Wonju, South Korea; 2) Lab Med, Hallym University Kangnam Sacred Heart Hospital, Seoul, Korea; 3) Lab Med, Yonsei University College of Medicine, Seoul Korea.

   Background: *EGFR* mutation is emerging biomarker for treatment selection in non-small-cell lung cancer (NSCLC) patients. However, optimal mutation detection is hindered by complications associated with the biopsy procedure, tumor heterogeneity, and limited sensitivity of the test methodology. In this study, we evaluated the diagnostic utility of real-time PCR, using malignant pleural effusion samples. **Methods:** A total of 77 pleural fluid samples from 77 NSCLC patients were tested using cobas *EGFR* mutation test (Roche Molecular Systems). The pleural fluid was centrifuged, and the separated cell pellets and supernatants were tested in parallel. The results were compared with Sanger sequencing and/or peptide nucleic acid (PNA)-mediated PCR clamping of matched tumor tissue or pleural fluid samples. **Results:** All samples showed valid real-time PCR results in one or more DNA samples extracted from the cell pellets and supernatants. Compared with other molecular methods, the sensitivity of the real-time PCR method was 100%. Concordance rate of real-time PCR and Sanger sequencing plus PNA-mediated PCR clamping was 98.7% **Conclusions:** We confirmed that real-time PCR using pleural fluid had a high concordance rate with the conventional methods, with no failed samples. The *EGFR* real-time PCR method was fast, sensitive, and robust that it can be applied to routine diagnostics in the clinical laboratory.

740T

Cooperative mutation study in Korean patients with *MLL*-rearranged acute myeloid leukemia using targeted next generation sequencing. *S. Kim[1], J. Yang[2], T. Park[2], Y. Cho[3], S. Jang[3], C. Park[3].* 1) LabGenomics Clinical Laboratory, Seongnam, Korea; 2) Department of Laboratory Medicine, School of Medicine, Kyung Hee University, Seoul, Korea; 3) Department of Laboratory Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea.

Background: Recent advancements of technologies including next generation sequencing (NGS) have led to discovery of molecular pathogenesis of malignant diseases including hematologic malignancies, and these discoveries are now enabling the beginning of molecular targeted therapy. The *MLL*-rearranged fusion gene is among the main leukemogenic mutation which is found in both acute myeloid leukemia (AML) and acute lymphoblastic leukemia with a frequency of about 5-10%, and is associated with a poor clinical prognosis. Cooperative mutation study among *MLL*-rearranged leukemia based on NGS technology has mainly been conducted in Caucasian or western population, and has yet been carried out in Asian ethnicity including Korean population. **Methods:** This study includes total of 24 *MLL*-rearranged AML patients who visited two separate tertiary care hospitals between the period of January 2009 and May 2014. The number of each *MLL* fusion genes are as follows; *MLL/MLLT3* n=12; *MLL/MLLT4* n=6; *MLL/ELL* n=2; other *MLL* fusion genes n=4. Mutation profile study for 19 candidate genes for cooperative mutation (*TET2, DNMT3A, IDH1, IDH2, NPM1, FLT3, CEBPA, ASXL1, BRAF, CBL, KIT, KRAS, NRAS, PTPN11, RUNX1, TP53, WT1, SETD2, JAK2*) was carried out using Miseq sequencing equipment (Illumina, San Diego, CA). **Results:** Among the twenty-four *MLL*-rearranged patients, 7 patients (29.2%) had positive results for more than 1 gene mutations which were analyzed for. Positive gene mutations found were in the order of frequency; *ASXL1* (n=4), *FLT3* (n=2), *CEBPA* (n=2), *KRAS* (n=1), *NRAS* (n=1) and *PTPN11* (n=1). Interestingly, 4 out of 6 patients harboring *MLL/MLLT4* were found to have additional gene mutations. Two patients had multiple gene mutations while one patient had four gene mutations concurrently detected. **Conclusion:** This is the first cooperative mutation study on *MLL*-rearranged AML patients of Asian ethnicity using targeted NGS technology. Despite small number of cases, a higher incidence of gene mutation among *MLL/MLLT4* patient group (4/6) and relatively higher incidence of *ASXL1* mutation (4/24) being found could be carefully suspected as an ethnical difference in the disease of AML. Further study requires a larger number of *MLL*-rearranged *MLL* patients and additional mutation profile study with ethnical comparison.

741F

Taming concerns regarding the accuracy of laboratory clinical data: A detailed comparative analysis. *H. LaDuca, R. McFarland, S. Gutierrez, A. Yussuf, N. Ho, P. Reineke, T. Cain, K. Blanco, J.S. Dolinsky.* Department of Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA.

Purpose: Genetic testing laboratories accumulate large amounts of clinical data from clinicians via test requisition forms (TRFs) and other clinical documents such as pedigrees and/or detailed clinic notes ('clinicals'). When curated, phenotype data can be used in variant assessment, genetic disease characterization, and clinical research; however, the completeness and accuracy of phenotype data from laboratory-based cohorts is often questioned. The purpose of this study is to evaluate the completeness and accuracy of TRF data for patients undergoing hereditary cancer multigene panel testing (MGPT). **Methods:** Ten percent of MGPT cases were randomly selected and reviewed from a cohort tested between January and June 2015 at a single clinical laboratory. For cases where clinicals were submitted *after* the test order date, TRF-reported cancer types and ages at diagnosis (dx) for probands and relatives were evaluated for accuracy and completeness using clinicals as the comparison standard. **Results:** Of 2,893 MGPT cases reviewed, 43.7% (n=1,263) submitted a pedigree and/or clinic note. For 31.3% (n=395) of cases, clinicals were submitted *after* the test order date and thus further evaluated. Cancer type and age at dx (+/- 2 years) were accurately reported for 99.4% and 97.8% of proband cancers, respectively. Proband cancer type was incomplete for 5.1% of cases and age at dx was incomplete for 7.4% of cases. Compared to probands, cancer type and age at dx for relatives were also highly accurate (type: 98.9%, p=0.553; age at dx: 95.8%, p=0.130) but significantly more likely to be incomplete (type: 29.7%; p=3E-28; age at dx: 26.4%; p=3E-16). When compared to 1st and 2nd degree relatives, incomplete data was more frequently observed among 3rd degree relatives and beyond (p=6E-8 and p=3E-5 for cancer types and ages, respectively). No significant differences were observed between maternal and paternal family history data. **Conclusions:** In this laboratory cohort, cancer type and age provided on the TRF is highly accurate for probands and relatives. Proband cancer histories exhibited a high level of completeness; however, this diminishes somewhat among relatives, particularly as relationship to the proband becomes more distant. It is imperative for clinicians to provide complete data not only on probands but also on relatives to aid in accurate molecular results interpretation, disease characterization and interpreting the overall health implications of genomic data.

742W

Multiplex TaqMan assays for rare mutation analysis using digital PCR. *M. Laig, B. Moy, F. Chan, L. Lac, T. Straub, K. Varma, D. Keys.* Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA.

   Introduction Detection of rare mutations for research purposes in tumor tissue and cell free DNA (cfDNA) allows for monitoring of tumor progression and regression. cfDNA isolated from plasma combined with a sensitive detection method like digital PCR is non-invasive and enables earlier detection compared to conventional imaging techniques. Building on the TaqMan based Rare Mutation assay set for detection of rare mutations using digital PCR on the QuantStudio 3D Digital PCR System, we are now developing multiplex assays for simultaneous detection of several mutations. We selected relevant mutations in the EGFR and KRAS genes for our initial multiplex application: EGFR G719, EGFR exon 19 deletions, and KRAS G12/G13. These mutations may have implications for potential future targeted therapy.   Methods Primers and probes of singleplex Rare Mutation Assays were reformulated to generate multiplex assays detecting the EGFR and KRAS mutations. All multiplex assays were tested on template composed of wild-type genomic DNA background mixed with mutant plasmid reflecting each of the mutations detected by the multiplex assays.   Summary Initial experimental results were successful and showed excellent signal intensity and clear cluster separation when analyzed with the QuantStudio 3D AnalysisSuite™ Cloud Software. The EGFR G719 mutations (COSM6239, COSM6253, COSM6252) were detected using a 3plex assay, EGFR exon 19 deletions (COSM12383, COSM12422, COSM12678, COSM6223, COSM6254, COSM6255) were detected using a 6plex assay, and KRAS G12/G13 mutations are underway.   Conclusion Multiplexing assays for three relevant mutation loci proved feasible and presents an efficient way to assess the presence and the percentage of mutations.

743T

Association of plasma *AR* copy number with response to pre-chemotherapy abiraterone acetate in metastatic castrate resistant prostate cancer. *J. Li[1,2], M.J. Du[1], D. Hillman[3], L.G. Wang[3], L.W. Wang[3], H.J. Zhang[1,4], P. Zhang[1], L. Wang[1], M. Kohli[3].* 1) Medical College of Wisconsin, Milwaukee, WI; 2) People's Hospital of Zhengzhou University, Zhengzhou, China; 3) Mayo Clinic, Rochester, MN; 4) Zhengzhou 7th People's Hospital, Zhengzhou, China.

   Background. We determined *AR* genomic changes in plasma cell-free DNA (cfDNA) and evaluated their associations with response to abiraterone acetate/prednisone (AA/P) and overall survival in a prospective clinical trial conducted in the pre chemotherapy metastatic castrate resistant prostate cancer stage. Copy number variation (CNV) analysis in *AR* using digital PCR technology was analyzed for determining association of *AR* CNVs with clinical outcomes. **Materials and Methods.** Metastatic sites of CRPC stage patients initiating pre-chemo AA/P were biopsied prior to (pre-AA/P) and after 12 weeks of treatment (post) and plasma was collected at the same time point on this prospective trial (NCT# 01953640). Composite response to AA/P at 12 weeks (primary endpoint) was evaluated with PSA, RECIST, bone scan and symptoms (per PCWG2). Plasma based biomarker analyses presented here are part of the correlatives included in the study protocol. CNVs were tested using Taqman CN assays in QuantStudio3D digital PCR system (dPCR). Statistical analyses were performed to test association of *AR* amplification with clinical outcomes (progression-free survival or PFS and overall survival or OS). **Results.** Between 6/2013 and 8/2015, 92 patients were enrolled of which 70 patients had plasma samples at both time points (58 patients with bone metastases; 27 with nodal metastases; 38/70 had high volume and 32/70 had low volume metastatic disease at the time of trial enrollment). At 12 weeks, 31/70 patients showed the composite progression; 21 patients had died at the time of this analysis (median/range follow up: 15.73/7.9-25.9 months). CNV analysis by dPCR showed 34/70 patients with AR amplification and 26/70 without amplification in pre-AA/P plasma. When AR CNV differences between pre- and post-treatment samples (log 2 ratio of pre- vs. post-treatment CNVs) were evaluated, the copy number increase after treatment was associated with 3-month progression (p=0.0013). A significant association of increased AR copy number with both PFS (HR= 1.23 and p=0.0019, Cox-regression) and OS (p<0.05, Kaplan-Meier analysis) was observed. **Conclusions.** *AR* amplification status in cfDNA was associated with disease progression and clinical outcomes in metastatic CRPC patients treated with AA/P in this study. Additional validation in prospective patient cohorts are needed to further test its clinical utility.

744F

An unusual translocation t(5;7) in a case with T-ALL with a complex karyotype. *L. Mendoza-Maldonado[1,2], L. Bobadilla-Morales[1,3], G. Serafín-Saucedo[3], E. Velazquez-Rivera[3], A. Vázquez-Reyes[1,2], F. Sanchez-Zubieta[5], J. Corona-Rivera[4], A. Corona-Rivera[1,3].* 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo, Instituto de Genética Humana "Dr. Enrique Corona Rivera", CUCS, Universidad de Guadalajara; 2) Doctorado en Ciencias en Biología Molecular en Medicina, CUCS, Universidad de Guadalajara; 3) Unidad de Citogenética, Servicio de Hemato-Oncología Pediátrica, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca"; 4) Servicio de Genética, División de Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca"; 5) Servicio de Hemato-Oncología Pediátrica, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca".

   T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive neoplasm of the bone marrow (Litzow and Ferrando, 2015). The acquisition of genetic abnormalities results in the proliferation of malignant lymphoid hematopoietic progenitors and could contribute to a poor prognosis (Mrózek *et al.*, 2009). Actually, only exist one report where found the presence of t(5;7)(q32;q22) which was related to cancer (Maubach *et al.*, 1985). We present a case with T-ALL with a complex karyotype including t(5;7)(q32;q22). Case report. Eight years old male with persistent low back pain and petechiae. He presented hyperleukocytosis, bicytopenia, tumor lysis syndrome, mediastinal mass, acute kidney injury, occipital lobe mass with midline shift and compression of the right lateral ventricle. On Juanuary 7, 2016, bone marrow aspirate was performed. The morphology analysis showed 82% cerebriform, large blasts. The immunophenotype result was CD2+, CD3+, CD7+, CD19+, CD22+, CD4+, CD24+, CD34+, CD38+, CD45+, HLA DR+. The patient was diagnosed with T-ALL with aberrant markers and primary CNS infiltration. Chemotherapy was initiated with doxorubicin, vincristine, L-asparaginase on January 12, 2017. He dies on 31 January, 2016. The diagnosis of defunsion was pulmonary hemorrhage, disseminated intravascular coagulation, septic shock and T-ALL. The result of karyotype was 48,XY,del(3)(q23),t(5;7)(q32;q22),+7,+8[12]/48,idem,add()14(q32)[2]/46,XY[1]. The pacient showed genetic abnormalities implicated in development of T-ALL in karyotype. Raimondi (2007) described the 14q32 region contains *TCL1, IGH, BCL11B* genes which are activated in T-cell neoplasms. Independently, trisomy 7 and 8 are common abnormalities as part of hyperdiploid karyotype in ALL and AML. However, there is only a case report by Maubach *et al.* (1985) of an adult with AML where both trisomies are presented like ours. The t(5;7)(q32;q22) is a chromosomal abnormality has been reported once in a case of recurrent glioma by Pruchon *et al.* (1994). Hoffman (2013) described *PDGFRB* gene it's found at 5q32 region and it has been identified with 22 different fusion partners in chronic myeloproliferative neoplasms although they are indeed very rare. This case report provides knowledge about a new chromosomal translocation in T-ALL which in combination with other genetic abnormalities provides a poor prognosis for the patient.

745W

A sensitive method for circulating DNA library preparation. *J.D. Murdoch[1], K. Krishnan[1], L. Apone[1], P. Liu[1], C.E. Massie[2], C.G. Smith[2], V. Panchapakesa[1], K. Duggan[1], D. Rodriguez[1], E. Dimalanta[1], T. Davis[1].* 1) New England Biolabs, Ipswich, MA 01938; 2) Cancer Research UK, University of Cambridge, Cambridge, UK CB2 ORE.

   Next-generation sequencing has enabled large-scale studies of cancer genomes to identify mutations. These tumor-specific somatic changes offer great potential as biomarkers to inform clinical practice, including detection, personalized medicine and disease monitoring. The ability to detect the mutational landscape of tumor cells using non-invasive approaches could enable rapid, personalized variant screening and efficient monitoring of therapy. Unfortunately, limitations of current standard technologies include lack of sensitivity and the scalability for a rapid analysis, especially in the case of small amounts of DNA from blood plasma samples. We have developed a DNA library preparation workflow for next-generation sequencing, which allows multiplexing and produces high quality, reproducible data using small amounts of healthy circulating plasma DNA. With as little as 500 pg of DNA we were able to generate sequencing libraries for whole genome (WGS) and whole exome (WES) with good GC coverage. We also achieved sufficient library complexity for high-confidence copy number detection and variant recovery, and we accurately determined the ratio of sex chromosomes in a mixed sample down to 5% of Y chromosome input. Exome target enrichment allowed recovery of known variants (>80%), showed good correlation with expected allele frequencies, and detection was consistent and reproducible across different starting concentrations of DNA. The simple and rapid workflow of this new kit allows a quick and sensitive approach to detect genomic variations. This method could facilitate biomarker identification to supplement diagnostic, prognostic and therapeutic applications.

746T

Amplicon panels for the detection of germline and somatic *BRCA1* **and** *BRCA2* **mutations.** *D. Obermoeller, R. Hrdlickova, J. Nehyba, C. Firmani, J. Kinman, D. Fox, A. Lewis, M. Toloue.* Bioo Scientific, Austin, TX.

The *BRCA1* and *BRCA2* tumor suppressor genes are associated with a high risk of breast, ovarian, and other types of cancer. Any germline, heterozygous, loss-of-function mutations in these DNA repair proteins lead to a strong predisposition to malignant disease. Moreover, the role of somatic mutations in these genes during the development of various cancers was recently described. Next generation sequencing allows a rapid and sensitive detection of mutations in these two genes. However, the available library preparation assays for their detection does not cover all of the needs of various applications of the analysis of *BRCA1* and *BRCA2* mutations. Bioo Scientific developed four panels for rapid detection of these mutations. For users seeking to identify germline mutations from fresh and frozen tissues, panels with 130 amplicons in two pools that can use as little as 20 ng of input DNA are available. The performance of these panels has been validated by positive identification of five described germline mutations in 20 ng DNA samples from Corriel Institute. With 100% uniformity these panels allow a large number of libraries to be multiplexed during sequencing. For 100x coverage of every targeted base, over 300 libraries could be sequenced on one Illumina MiSeq instrument making it the most economical assay on the market. For identification of somatic mutations from partially degraded DNA, panels of 260 smaller amplicons with a higher redundancy are available in four pools. These assays were fully validated with several FFPE samples reaching 99% uniformity at 20% coverage and 100% uniformity at 10% coverage. Only 40 ng of DNA extracted from the FFPE samples is necessary for the preparation of one library. These panels for the detection of somatic mutations in partially degraded DNA are available for both the Illumina and Ion Torrent platform. All of the *BRCA1* and *BRCA2* panels cover 20.4 kb pairs comprising 48 coding *BRCA1* and *BRCA2* exons and limited flanking intron regions (4.3 kb). These panels offer a versatile and economical tool for diagnostic and further research of the role of these tumor suppressor genes in human cancer filling in the current gap in the market. .

747F

Quantitative measurement of burden of Calreticulin using pyrosequencing in patients with myeloproliferative neoplasm. *Y. Oh[1], Y. Park[1], H. Kim[1], J. Lim[1], Q. Choi[2], G. Kwon[1], S. Koo[1], S. Kim[1,3].* 1) Department of Laboratory Medicine, Chungnam National University Hospital, Daejeon, South Korea; 2) Department of Laboratory Medicine, Dankook University Hospital,Cheonan, South Korea; 3) Cancer Research Institute, Chungnam National University School of Medicine, Daejeon, Republic of Korea.

Frame-shift mutations in the calreticulin (*CALR*) gene were recently identified as key pathogenic mutations of myeloproliferative neoplasms (MPNs). We developed novel method to quantitate *CALR* mutations burden using pyrosequencing technique. We validated the performance of this technique in synthesized DNA bearing type 1 and type 2 mutations, and evaluated the clinical usefulness in the MPN patients. To determine analytical performance of the *CALR* quantitation, we cloned plasmid vector bearing mutant DNA fragment bearing type 1 and type 2 *CALR* mutations. Triplets' tests for 6 serial equal-volume dilutions were performed to detect the sensitivity of the assay. Genomic DNA of bone marrow and peripheral blood samples from 14 patients diagnosed as MPN without *Jak2* or *MPL* mutation were collected. Each primer sets were designed to detect type 1 and type 2 *CALR* mutations and DNA fragment including *CALR* was amplified by polymerase chain reaction followed by Pyrosequencing (Qiagen, Hilden, Germany). Pyrograms were generated, which is then converted to peak height. The data was processed manually to quantitate the mutation burden. All the mutations were reassessed by Sanger sequencing. The assay showed high precision, linearity and reproducibility. From the sensitivity and specificity study, mutation detection level was detected as about 1%. Among the 14 patients; there were 3 type 1 mutations, 1 deletion mutation, 2 insertion mutations (GTGTC and TTTGTC). All 14 samples showed 100% concordance compared to Sanger sequencing. In conclusion, pyrosequencing can accurately measure the burden of type 1 and type 2 *CALR* mutations and can also detect other types of *CALR* indels. Therefore, pyrosequencing is considered as a rapid, convenient and reliable method for detecting type 1 and type 2 *CALR* mutations.

748W

Consensus molecular classification of colorectal cancer and association with the colonic microbiome. *R. Purcell[1], J. Pearson[2], S. Schmeier[3], F. Frizelle[1].* 1) Department of Surgery, University of Otago, New Zealand, Christchurch, Canterbury, New Zealand; 2) Department of Biostatistics and Computational Biology Unit, University of Otago, New Zealand, Christchurch, Canterbury, New Zealand; 3) Institute of Natural and Mathematical Sciences, Massey University, Private Bag 102-904, North Shore City, New Zealand.

   Colorectal cancer (CRC) is the second most common cancer worldwide, and shows heterogeneous response to treatment and outcomes. Most CRCs are sporadic, and follow a pattern one would expect from a yet unidentified environmental source. This study brings together two key emerging research avenues, namely the new clinical frontier of applying information on the gut microbiome to treatment and management of CRC, and the use of advanced molecular classification systems to better investigate the development of CRC. Recent studies have classified CRC based on *BRAF* and *KRAS* mutation status, microsatellite instability, CpG island methylator phenotype, and activation of various molecular pathway, but discrepancies exist between these classification systems. In 2015, a large international consortium published a classification system based on gene expression data. These published classifiers stratify CRC into four consensus molecular subtypes (CMS): CMS1 (MSI immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal). We have stratified 34 tumors from our CRC cohort into CMS subtypes using RNA-Seq data. Extensive clinical and follow-up data available on our cohort allowed us to investigate the associations of the molecular subtypes with clinical and histological features, and with response to therapy. The human colon plays host to a vast and complex microbial community of $< 10^{12}$ microorganisms, and gut dysbiosis is believed to play a role in the development of CRC. Given the high number of different bacterial species found in the gut ($< 10^3$), and the wide interpersonal variation in CRC microbiomes, redundancy of function is quite likely. Therefore, examining changes in meta-communities, clusters of orthologous groups, and putative virulence factors, represents a more promising means for deciphering the contribution of microbial dysbiosis to the genesis of CRC. We have used metagenomic analysis to map the CRC microbiomes of our cohort, and will examine the association between differeing microbiome patterns and CMS subtypes of our tumour cohort. We anticipate that this will reveal particular species of bacteria or bacterial virulence factors involved the development of a particular CRC subtype. This would represent an important step in understanding the interplay between microbial diversity in the gut and the molecular mechanisms involved in the development of CRC.

749T

Variants of uncertain significance in *BRCA*, a challenge yet to be solved by clinical geneticists. *V.E. Radoi[1,2], R.I. Ursu[1,2], G. Chelu[2], C. Arsene[2], R.E Bohiltea[1], L.C. Bohiltea[1].* 1) "Carol Davila" University of Medicine and Pharmacy, Faculty of General Medicine, Bucharest, Romania; 2) SYNEVO Romania, Central Laboratory, Department of Medical Genetics, Buchatrest, Romania.

   Background: BRCA1 and BRCA2 gene mutations account for 20-25% of hereditary breast cancer and about 5-10% of all breast cancers. In addition, mutations in these genes account for about 15% of all ovarian cancers. The BRCA 1 and BRCA2 have been the most studied and reported genes in human genome over the past decades. **Purpose**: The using of cuttingedge genetic analysis technologies imposes the existance of large comprehensive genomic databases for an accurate clinical interpretation. At the same time, these modern diagnostic techniques lead to the identification of novel variants with unknown or controversial clinical significance which raise great challenges for the clinicians in the analysis. **Material. Methods.** A review of the main BRCA databases (1000 Genomes,ClinVar,NCBI,SNPedia,PharmGK-B,OMIM,COSMIC,UCSC Genome Browser, BioGPS, etc.) and professional workgroups` guidelines and recommendations has been performed for a comprehensive outline of the current knowledge on BRCA variants associated with hereditary breast and ovarian cancer. Different classification categories proposed by leading national and international professional societies have been taken into consideration. **Results.** By analyzing the abovementioned databases and variant classes we observed slight differences between the variant clinical correlation classification criteria, number of classes and especially terminology used to define these classes. These discrepancies tend to produce confusion when in terms of analysis and clinical interpretation, some variants with uncertain significance being extremely difficult to categorize, with consequences in managing the patients. **Conclusions:** In conclusion, establishing an unified BRCA database would be of great value in the assessment of the multitude of data provided by modern genetic technology. The existance of a large database containing all the current BRCA knowledge would not only help physicians in accurately interpreting the implications of these variants in cancer etiopathogeny, but would also lead to the correct identification of novel pathogenic variants.

750F

Detection of monoclonal *IGH* rearrangements in circulating cells from general population from Western Mexico. *S. Rodríguez[1,2], M.T. Magaña-Torres[2], A.R. Jaloma-Cruz[2], P. Barros-Núñez[2].* 1) CENTRO UNIVERSITARIO DE CIENCIAS DE LA SALUD, UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, JALISCO, MEXICO; 2) CENTRO DE INVESTIGACIÓN BIOMÉDICA DE OCIDENTE, CMNO, IMSS, GUADALAJARA, JALISCO, MEXICO.

The monoclonal B-cell lymphocytosis (MBL) is a condition characterized by the presence of small populations of monoclonal lymphocytes B circulating in peripheral blood ($<5\times10^9$/L), in the absence of chronic lymphocytic leukemia (CLL) or other B-cell lymphoproliferative disorder, infections or autoimmune diseases. MBL frequency in the general population (asymptomatic individuals, unrelated, without neoplastic history) is significantly different in diverse studied groups, ranging from 0.6% in the USA population to 12% in the Spanish population. Approximately 1-5% of individuals carrying MBL evolve annually to CLL, while virtually all the CLL cases are preceded by MBL. In this study, we estimated the frequency of monoclonal B-cell lymphocytosis in the general population from the western of Mexico. DNA extracted from a sample of peripheral blood by CTAB/DTAB method was used to amplify the rearrangements of the immunoglobulin heavy chains by multiplex PCR, using 20 specific primers for the complete rearrangements (VDJH), and seven primers for the incomplete rearrangements (DJH) of the *IGH* gene. After, the products were separated by capillary electrophoresis and statistical analysis was performed. We studied the *IGH* rearrangements in 188 individuals of the general population from Western of Mexico. In 24 (12.8%) of them, monoclonality was detected in at least one of the regions amplified by PCR. The MBL frequency increases with the age and all individuals showing monoclonality were over 40 years. A higher frequency of MBL was observed in men, with a higher frequency of monoclonal incomplete rearrangements. In conclusion, the MBL incidence in the general population of the Western of México is significantly major to those previously described for other populations but similar to those reported in the Spanish population. Detection of monoclonal rearrangement in individuals may help to identify cases requiring continuous clinical monitoring in order to anticipate disease progression.

751W

Next-generation sequencing of the *BRCA1* and *BRCA2* genes in 700 patients for the genetic diagnostics of hereditary breast and/or ovarian cancer. *D. Trujillano[1], B. Jori[1], K. Kumar Kandaswamy[1], M. Weiss[1], S. Lehnert[1], A. Marais[1], R. Schröder[1], C. Voight[1], J. Köster[1], N. Nahavandi[1], O. Brandau[1], R. Abou Jamra[1,2], P. Bauer[1], A. Rolfs[1,3].* 1) Centogene AG, Rostock, Germany; 2) Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany; 3) Albrecht-Kossel-Institute for Neuroregeneration, Medical University Rostock, 18147 Rostock, Germany.

Genetic testing for hereditary breast and/or ovarian cancer (HBOC) mostly relies on laborious molecular tools that use Sanger sequencing to scan for mutations in the *BRCA1* and *BRCA2* genes. We introduced a more efficient strategy based on next-generation sequencing in 700 (HBOC) patients. We first validated this approach in a cohort of 115 samples with previously known *BRCA1* and *BRCA2* pathogenic variants and polymorphisms. The entire coding region of *BRCA1* and *BRCA2* including 10bp of intronic flanking sequences was amplified and sequenced on the IonTorrent PGM sequencer (Life Technologies) in the genomic DNA. The combination of robust bioinformatics tools allowed us to detect all previously known pathogenic variants and polymorphisms in the 115 samples, without detecting spurious pathogenic calls. The assay achieved a sensitivity of 100% (95% CI: 99.71% to 100%), with a specificity of detecting non-variant sites from the reference sequence of 99.99% (95% CI: 99.99% to 100%), a positive predictive value of 91.17% (95% CI: 89.72% to 92.62%), and a negative predictive value of 100% (95% CI: 100% to 100%). The same assay has been used in a discovery cohort of 585 uncharacterized HBOC patients for *BRCA1* and *BRCA2*. In addition, we describe the allelic frequencies across 700 HBOC patients of 97 unique definitely and likely pathogenic and 92 uncertain *BRCA1* and *BRCA2* variants, some of them not previously annotated in the public databases. Targeted NGS is ready to substitute classical molecular methods to perform genetic testing on the *BRCA1* and *BRCA2* genes, and provides a greater opportunity for more comprehensive testing for at-risk patients.

752T

Clonal hematopoiesis confounding genetic epidemiology research in a prospective cohort of individuals undergoing genetic cancer risk assessment. *J.N. Weitzel[1,2], S.R. Sand[1,2], D. Castillo[2], R. Mejia[1], J. Herzog[2], S.L. Neuhausen[1,2], T.P. Slavin[1,2], Clinical Cancer Genomics Community Research Network.* 1) City of Hope Comprehensive Cancer Center, Duarte, CA; 2) Beckman Research Institute of City of Hope, Duarte, CA.

We previously demonstrated apparent somatic interference in germline multigene panel test (MGPT) results from blood and sputum samples, wherein pathogenic *TP53* variants with mean allele fractions (MAF) < 30% were detected in patients without phenotypic manifestations of Li-Fraumeni syndrome. Absence of the pathogenic variants in alternate sources of constitutional or at-risk tissue suggested that clonal hematopoiesis, with or without incipient hematologic abnormalities, was likely the source of the variant in the majority of cases. The recently described entity of clonal hematopoiesis of indeterminate potential (CHIP) is defined as somatic myeloid pathogenic gene variants involving ≥2% of an individual's blood, without that individual meeting other diagnostic criteria for hematologic malignancy. For this study, we evaluated CHIP using an exon capture panel containing *ATM, CHEK2, KIT, NF1, NRAS, RUNX1, SETBP1, SF3B1, TP53,* and *PPMD1*. We conducted next generation sequencing (NGS) on an Illumina 2500HT on germline (blood) DNA from 251 women and 30 men, with uninformative commercial genetic test results, from the City of Hope Clinical Cancer Genomics Community Research Network registry. The mean age at blood sample collection was 51.7 years (range 17-90 years); prior cancer history included 183 breast or ovarian cancers, 84 colorectal cancers and other cancer diagnoses among 263 affected participants; 18 participants were unaffected. Four individuals (1.4%) had potential CHIP pathogenic gene variants: *TP53*:c.736A>G, *PPM1D*:c.1448delC, and *ATM*: c.1236-2A>G and c.1236-2A>T. Read depths ranged from 29 to 157 and allele fractions ranged from 18.5-38.3%. All 4 individuals had prior chemotherapy for breast cancer and/or were > 50 years old at the time of enrollment and sample collection (both features seen in other populations with clonal hematopoiesis), and none had any evidence of hematologic neoplasia. It is possible that more cases with CHIP would have been identified with inclusion of other genes more frequently associated with CHIP, such as DNMT3A. Thus, CHIP is detectable with conventional NGS methods and may confound cancer genetic epidemiology studies and MGPT. The clinical implications of CHIP with regard to prospective development of hematologic malignancies deserve further study.

753F

Targeted resequencing of *BRCA1* and *BRCA2* in familial breast cancer. *M. Wong-Brown[1], R. Scott[1,2].* 1) University of Newcastle, Newcastle, New Lambton Heights, Australia; 2) Pathology North, Newcastle, Australia.

Inherited loss-of-function mutations in *BRCA1* and *BRCA2* predispose to high risk of breast cancer. Since the discovery of breast cancer susceptibility genes *BRCA1* and *BRCA2* two decades ago, there have not been any other genes identified that play a significant role in predisposition to inherited breast cancer. A large proportion of individuals with inherited breast cancer are negative for *BRCA* mutations and despite numerous research efforts, further breast cancer susceptibility genes still remain elusive. We hypothesize that potentially deleterious mutations may reside in the less-researched non-coding sequences. This study aims were to identify genetic anomalies in *BRCA1* and *BRCA2* by completely re-sequencing 200 kilobases surrounding *BRCA1* and *BRCA2* using next-generation sequencing. For this study, DNA was used from 10 individuals referred for genetic testing after meeting the criteria for inherited breast cancer, and had been screened for *BRCA1* and *BRCA2* mutations by the Hunter Area Pathology Service (Newcastle, NSW, Australia). All individuals used for this study did not harbour causative genetic changes in the coding regions of *BRCA1* or *BRCA2*. Targeted sequencing of the entire *BRCA1* and *BRCA2* genes was performed. Common SNPs were removed from further analyses. Single nucleotide variants (SNVs) and insertions/deletions (indels) were identified in most individuals tested in regions that had previously remained unexplored, such as the non-coding regions of BRCA1 and BRCA2, 5'-UTR, 3'-UTR and promoter sites. One patient in particular displayed a large number of BRCA1 3'-UTR variants. The PITA miRNA target prediction algorithm showed that target miRNAs and their alignment scores were altered due to variants in this region. An overview of the alignment scores show changes of binding strengths in several miRNAs, including those that have been shown to target BRCA1. Multiple BRCA1 3'-UTR variants could potentially change miRNA target prediction profile. The relative expression of BRCA1 mRNA in this patient's cell lines was also shown to be significantly lower compared to several control cell lines, suggesting that miRNAs may be regulating BRCA1 gene expression The aim of this study is the increase in current knowledge of the genetic variations that results in the development and/or progression of inherited breast cancer, and aid in the management of individuals with breast cancers by providing a more specific diagnosis of disease risk.

754W

Comparison of genetic heterogeneity between tumor tissues and circulating cell-free DNA in hepatocellular carcinoma. *X. Zhao[1], A. Huang[2], F. Li[1], X. Yang[2], K. Wu[1], J. Fan[2], J. Zhou[2], Y. Hou[1].* 1) BGI-Shenzhen, Shenzhen 518083, China; 2) Liver Surgery Department, Liver Cancer Institute, Zhongshan Hospital, Fudan University, China.

Intratumor heterogeneity presents a major challenge for precise evaluation of hepatocellular carcinoma (HCC). Recent evidence suggests that circulating cell-free DNA (cfDNA) could reveal the profiles of mutational landscape in solid tumors. However, it remains unknown whether the intratumor heterogeneity in HCC could be reflected in cfDNA and to what extent the information obtained from cfDNA could reveal the genetic intratumor heterogeneity for clinical utility in HCC. Here we characterized the genomic architectures and clonal evolution in fifty-two tumor tissues and ten cfDNAs from five HCC patients prior to and one week after surgical removals of tumors by whole exome sequencing and targeted deep sequencing. The whole exome sequencing identified intratumor heterogeneity in tumor tissues of all cases. By targeted deep sequencing preoperative cfDNA (Pre-cfDNA), we detected approximately 82.7% (72%~96%) of somatic mutations that were found in multiregional tumor tissues. The mutation level in pre-cfDNA have showed the diversity on clonal hierarchy. And the tumor spatial heterogeneity might be the critical factor which affected the detection rates and proportion of clonal hierarchy of tumor-derived somatic mutations in pre-cfDNA. We also found the percentage of tumor-derived somatic mutations which be detected in postoperative cfDNA (Post-cfDNA) were reduced in eighty percent of HCC patients. The detection rate of ubiquitous mutations, which were detected in all the tumor regions of each patient, in post-cfDNA of high heterogeneity patients could be used as a prognostic indicator for recurrence after surgical treatment in HCC. Our findings could provide a different insight into the utility of cfDNA as liquid biopsy in clinical application of HCC for diagnosing and monitoring.

755T

Somatic driver mutations enhance survival prediction in familial chronic lymphoid leukemia. *W. Zhou[1,2], L. Goldin[2], M. Wang[1,2], M. McMaster[2], S. Chanock[2], M. Dean[2], M. Yeager[1,2], N. Caporaso[2].* 1) Cancer Genomics Research Laboratory, Leidos Biomedical Research/NCI, Bethesda, MD 20892; 2) Division of Cancer Epidemiology and Genetics (DCEG), NCI, 9609 Medical Center Drive MSC 9776 Bethesda, MD 20892.

Chronic lymphoid leukemia (CLL) is the most common type of leukemia in adults. Traditionally, four common chromosomal aberrations (deletion of 11q, 13q, 17p, duplication of 12) as well as the mutational status of the immunoglobulin heavy chain variable (IGHV) gene have been shown to stratify CLL patients into distinct prognostic groups. In the last few years, several tumor sequencing studies have revealed a number of recurrent mutations, some of which also contribute to prognosis. In this study, we performed whole exome-sequencing of whole blood samples from 98 CLL patients (and 95 subjects with SNP microarray data) from 40 CLL families. Log R ratio (LRR) and B allele frequency (BAF) were used to assess copy number alterations as well as to calculate the proportion of cancer cells. Overall, 142 detectable somatic copy number alterations (SCNAs) from 54 subjects were detected (57%), including the four most common CLL-associated cytogenetic aberrations: losses of chromosome 13q (28.9%), 11q (5.6%), and 17p (2.1%); gain of chromosome 12 (4.2%). We found many chromosomal alterations were observed only in a portion of the cells (mosaicism). We used a somatic pipeline caller to detect somatic mutations and used other filtering methods to eliminate likely germline mutations. We focused on a panel of 52 genes that previously reported as putative driver alterations in mature B-cell non-Hodgkin lymphomas. Overall, 72 non-synonymous somatic mutations in 34 of these genes were detected among 47 subjects (48%). The most frequently mutated gene was *TP53* (7.1%). We found that patients having both a point mutation in a CLL driver gene and SCNA have worse survival (HR=3.98, 95%CI=1.25-12.64; P=0.02, ref = single 13q deletion) than patients having either a point mutation (HR = 2.19 and P = 0.02, ref = no point mutation) or SCNAs (HR = 3.13 and P = 0.05 in SCNA, ref = single 13q deletion). *TP53* mutations were associated with the poorest overall survival (HR=6.01, 95%CI=1.68-21.49; P=0.006). Subjects having a somatic mutation in *CHD2*, *MYD88*, and *KLHL6* had similar survival times to those without these mutations. Additionally, we confirmed that *TP53* mutation status has independent prognostic value and that males with CLL have poorer survival than female CLL patients. Our study demonstrates that combining somatic CNAs and mutational data can add to predicting the pathway from an initial diagnosis to outcome in CLL.

756F

Isochromosome i(7q) in two patients with acute lymphoblastic leukemia.
A. Corona-Rivera[1,2], L. Mendoza-Maldonado[1], P. Chávez-Panduro[2], G. Serafín-Saucedo[2], L. Bobadilla-Morales[1,2]. 1) Dept BMG, Inst GH, Univ Guadalajara, CUCS, Guadalajara, Jalisco, Mexico; 2) Lab. De Citogenética, Genotox. y Biomonit., CUCS, UdeG. Serv. de Hemato-Oncol. Ped., HCG "Dr. Juan I. Menchaca". Guadalajara, Jalisco, México.

INTRODUCTION The i(7q) is a rare chromosomal aberration in ALL, it may present isodicentric abnormality and breakpoints in proximal 7p. Although the association between malignancy and the presence of i(7q) suggests it plays an important role in disease progression has not been possible to elucidate its consequences in patients with ALL. OBJECTIVE To present two patients with pre-B ALL and i(7q) in karyotype. CASE REPORTS Patient 1. Male, 15 years old, refers vomiting, headache, backache and presents cervical lymphadenopathy, leukocytosis. It is performed bone marrow aspirate and karyotype. He is diagnosed with pre-B ALL in December, 2011. Chemotherapy starts and achieves remission at 14 days, radiation is added for primary CNS infiltration with CSF positive in July, 2012. He is in surveillance since November, 2014. The result of the initial karyotype was 46,XY,idic(7)(p11)[13]/46,XX[7][cp20]. Patient 2. Female, 17 years old with obesity, anemia and leukocytosis. It is performed bone marrow aspirate and cytogenetic analysis. She is diagnosed with pre-B ALL in August, 2011. Chemotherapy starts and achieves remission at 14 days. She is in surveillance since August, 2014. The result of the initial karyotype was 46,XX,idic(7)(p11)[3]/46,XX[3].ish.idic(7)(p11;p11)(D7Z1++,CUTL1++,D7S2419-D7S2624++) CONCLUSION The evolution of patients in this study does not show i(7q) as the only chromosomal aberration, confers a poor prognosis, this is according to Pui et al., 1992 where 10/83 isochromosomes cases with no additional chromosomal abnormality were considered, but its mechanism of action is unclear what this is still open to discussion.

757W

Chromosomal aberration pattern predicts uveal melanoma mutation status and prognosis. *A. de Klein[1], S. Yavuzyigitoglu[1,2], W. Drabarek[1,2], A. Obulkasim[2], T. Brands[1,2], B. Eussen[1], E. Kilic[2].* 1) Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 2) Ophthalmology,Erasmus MC, Rotterdam, Netherlands.

Rationale: In uveal melanoma (UM) non-random chromosomal aberrations occur and correspond to patients' prognosis. Mutations in UM specific genes, such as *BAP1*, *SF3B1* and *EIF1AX* are also used to predict survival. Aim of this study is to identify whether these mutations correspond to a specific chromosomal signature in UM. Methods: For 277 UM patients SNP array data (n = 214) and/or conventional karyotyping (n = 119) of the tumor was available. The mutational status was determined in 189 patients. Independent of the mutational status, SNP array and conventional karyotyping data was analyzed for recurring copy number variations (CNVs) and structural variants (SVs). Hierarchal clustering of the SNP array data was performed to construct clusters. These clusters were correlated to the mutational status. Results: *BAP1*, *SF3B1* and *EIF1AX*-mutated UMs display specific chromosomal patters with recurring CNVs. Both *BAP1*-mutated and *SF3B1*-mutated UMs are characterized by specific chromosome changes. *BAP1*-mutated tumors showed predominantly numerical changes of chromosome 3 and 8 whereas *SF3B1* mutated tumors are characterized by multiple (>3) structural chromosomal anomalies. *EIF1AX*-mutated UM are characterized by only chromosome 6p gain without any additional CNVs. Hierarchal clustering of the SNP array data revealed five clusters of which two clusters predicted *BAP1* mutations: these UM patients have a high risk of developing early (<5 y) metastatic disease. Tumors with *SF3B1* mutations also cluster together and these patients have an intermediate risk of developing, often late (7-10y), metastasis. The remaining cluster are a mix of *EIF1AX*-mutated tumors or UM without detectable mutations; these patients have in general a low risk of developing metastasis. Conclusion: UMs with either *BAP1*, *SF3B1* or *EIF1AX* mutations display a mutation-specific chromosomal pattern. Based on the chromosomal patterns a genetic mutation in one of the UM cancer genes can be predicted. .

758T

The impact of LPL and CRY1 expression as prognostic genetic markers in B-Cell Chronic Lymphocytic Leukemia of Egyptian patients. *A.E. Ibrahim[1], M. Shehab[1], A. Elgerzawy[1], K. Mohammed[2], S. Abd-Elbaset[3], M. Ahmed[3], T. Abd-Elhamid[4].* 1) Human Cytogenetics Department, National Research Centre, Cairo, Egypt; 2) Molecular Genetic Department, National Research Centre, Cairo, Egypt; 3) Faculty of Science, Al-Azhar University (Girls Branch), Cairo, Egypt; 4) Medical Oncology and Hematology, National Cancers Institute, Cairo, Egypt.

  Introduction: Chronic lymphocytic leukemia (CLL) is the most common leukemia in adults.CLL is a heterogeneous disease, some patients die within a few months of diagnosis and others live for many years**.** So, it is needed to investigate reliable prognostic factors for predicting stable or progressive forms of CLL that facilitate risk-adapted treatment strategies. Cytogenetic abnormalities and molecular prognostic markers such as lipoprotein lipase (*LPL*) and cryptochrome-1 (*CRY1*) have been evaluated. **Materials and Methods:** This study included 42 patients with CLL (25 males and 17 females) and 10 controls. Patients newly diagnosed on the basis of clinical criteria and laboratory features. Thay are subdivided to 2 groups according to their response of treatment; patients with favorable outcome and patients with unfavorable outcome. Fluorescence In Situ Hybridization (FISH) technique was done for the detection of del(13q14) and del(17p13). Quantitative real-time PCR (qRT-PCR) technique was done for the detection of gene expression of target genes *CRY1*, *LPL* and Housekeeping gene Beta-Glucuronidase (*GUS*) in patients and controls. **Results:** Patients' age fell in the range (26-76) year with a median of 61 years. Deletion 17p13 combined with del(13q14) were the most common abnormality (64.3%), followed by del(17p13) 19% and del(13q14) 14.3%. Other chromosomal abnormalities were found among CLL patients; more than one copy (polyploidy) of gene locus DLEU in chromosome 13q14 and gene locus p53 in chromosome 17p13 were found in 45.2% of cases. Polyploidy is rare in CLL and not commonly found, so this study were the first study of CLL Egyptian patients had polyploidy with this percentage. High expression of *CRY1* gene was detected in 64.3% and high expression of *LPL* gene was detected in 75.6%. Significant inversely correlation of *CRY1* and *LPL* genes expression was found between patients and controls. Significant correlation was found between *CRY1* gene expression, *LPL* gene expression, patients' outcome and deletion 17p13. **Conclusion:** Patients in early stage of CLL with high *CRY1* or *LPL* genes expression must be early treated even in the presence of favorable prognostic parameters. *CRY1* and *LPL* genes expression had a higher detection rate of prognosis than cytogenetic markers. The combination of them provided more information about the prognosis than *CRY1* gene or *LPL* gene alone. .

759F

Malignant glioma with primitive neuroectodermal tumor-like component (MG-PNET): Novel microarray findings in a pediatric patient. *J. Liu[1], M.P. Keisling[1], A. Samkari[2,3], G. Halligan[2,3], J.M. Pascasio[1,2], C.D. Katsetos[1,2,4].* 1) Department of Pathology and Laboratory Medicine, Drexel University College of Medicine, St. Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA; 2) Department of Pediatrics, Drexel University College of Medicine; 3) Sections of Oncology, St. Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA; 4) Sections of Neurology, St. Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA.

  Central nervous system (CNS) tumors exhibiting dual features of malignant glioma (MG) and primitive neuroectodermal tumor (PNET) are rare and diagnostically challenging. Previous studies have shown that MG-PNET carry *MYCN* or *MYC* gene amplifications within the PNET component concomitant with glioma-associated alterations, most commonly 10q loss, in both components (Perry et al., Brain Pathol. 2009;*19*:81-90). Here we confirm and extend the profile of molecular genetic findings in a MG-PNET involving the left frontal lobe of a 12-year-old male. Histologically, the PNET component showed features akin to anaplastic medulloblastoma highlighted by widespread immunoreactivity for βIII-tubulin (TUBB3) and non-phosphorylated neurofilament protein, and to a lesser degree, Neu-N, synaptophysin, and CD99, whereas the gliomatous component was demarcated by GFAP labeling. Interphase fluorescent in situ hybridization (FISH) study on touch preparations from frozen tumor showed amplification of *MYC*. Single nucleotide polymorphism (SNP) microarray analysis revealed that the tumor carried gains of multiple chromosomes and chromosome arms, losses of multiple chromosomes and chromosome arms, gains of multiple chromosomal segments (not limited to amplification of chromosomal segments 4q12 including *PDGFRA,* and 8q24.21 including *MYC*), and chromothripsis on chromosome 8. The molecular genetic findings support the presence of a CNS-PNET as an integral part of the tumor coupled with overlapping genetic alterations found in both adult and pediatric high-grade gliomas/glioblastoma. Collectively, microarray data point to complex underpinning genetic alterations associated with the MG-PNET tumor phenotype. .

760W

Frequency of additional cytogenetics abnormalities in chronic myeloid leukemia patients at time of diagnosis and during follow-up: A single centre experience. *S. Shan[1], M. Nadeem[2], S. Sheeba[3], A. Nida[4], S. Tahir[5].* 1) Cytogenetics, National Institute of Blood Diseases & BMT, Karachi, Sindh, Pakistan; 2) HOD, National Institute Of Blood Disease & BMT, Karachi, Sindh, Pakistan; 3) Ex Employee, Cytogenetics, National Institute of Blood Diseases & BMT, Karachi, Sindh, Pakistan; 4) Consultant, National Institute Of Blood Disease & BMT, Karachi, Sindh, Pakistan; 5) MD, National Institute Of Blood Disease & BMT, Karachi, Sindh, Pakistan.

INTRODUCTION:At the time of presentation, 90-95% cases of Chronic Myeloid Leukemia (CML) have characteristic t(9;22)(q34;q11.2) reciprocal translocation that results in Ph Chromosome. The remaining cases either have variant translocations that involve a third or even a fourth chromosome in addition to chromosomes 9 and 22 or have cryptic translocation of 9q34 and 22q11.2 that cannot be identified by routine cytogenetic analysis. The additional cytogenetic abnormalities, at times, have been associated with adverse prognostic outcome and resistance to treatment.OBJECTIVE:To study the frequency of cytogenetic abnormalities in addition to Ph Chromosome in patients of CML at the time of diagnosis.MATERIAL AND METHODS:A cross sectional study was carried out in the department of cytogenetics NIBD Karachi from May 2010 to April 2016. Cytogenetic analysis was performed on newly diagnosed patients & patient on followup of CML using un-stimulated 24 hrs and stimulated 72 hrs cultures employing trypsin giemsa banding technique.RESULTS:Bone Marrow samples of 220 Ph positive CML patients were analyzed for cytogenetic abnormalities. There were 142 (64.5%) male and 78 (36%) female patients with median age of 38 years (range 14-73 years). Out of total, 19 (07%) patients had additional cytogenetic abnormalities including Trisomy 8 (01 patients), Complex Karyotype (05 patients), -Y (1 patient), t(1;17) (1 patient), double Ph Chromosome (02 patients), Hyperploidy (03 patient), inv (9) (01 patient), Trisomy 21 (01 patient), t(7;9;22) a novel translocation (01 patient),t(4;9;22) (01 patient), del(7q) (02).CONCLUSION:Additional cytogenetic abnormalities are not infrequent in patients of CML. High degree of suspicion and thorough examination of metaphases should be performed to detect these aberrations. Presence of additional chromosomal aberrations at diagnosis or during the course of disease, at times signifies impending conversion to advance phase and/or resistance to therapy.**KEYWORDS:**Chronic Myeloid Leukemia, Ph Chromosome, Additional Cytogenetic Abnormalities.

761T

Mapping complex genomic translocations in tumor cells using Strand-Seq. *M.W. Yuen[1,2], M. Hills[2], A.D. Sanders[2], E. Falconer[2], D. Spierings[3], P. Lansdorp[2,3].* 1) University of British Columbia, Vancouver, Canada; 2) Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, Canada; 3) European Research Institute for the Biology of Aging (ERIBA), University of Groningen, The Netherlands.

Structural variations (SV) such as translocations are often difficult to study. Traditional methods lack the effective resolution to determine breakpoints or require *a priori* knowledge of the type and location of rearrangements. Next generation sequencing has improved our ability to map novel SV, but often requires deep sequencing to identify reads that can be used to detect the breakpoints. Here we explored the use Strand-seq, a single cell sequencing technique that preferentially sequences inherited parental DNA template strands, to map SV. Strand-seq differentiates Watson and Crick strands, and the changes in strand-state can be used to mark regions of chromosome rearrangements. To validate the use of Strand-seq in a global, unbiased interrogation of SV, we used an Acute Lymphocytic Leukemia cell line that was characterized by cytogenetics and found to harbour a complex 4-way translocation with 1 known fusion gene. Sequencing with just a single MiSeq lane, we identified each of the 4 translocations and were capable of resolving the breakpoint at each translocation within a ~20kb region. This is orders of magnitude better than commonly used cytogenetics approaches. We are exploring automation of the analysis pipeline to allow rapid, high-resolution breakpoint identification and the ability to resolve translocation partners for each breakpoint. Taken together, the Strand-seq single cell sequencing approach presents a new vista in cancer genomics, with potential for accurate, high throughput calling of genomic rearrangements present in a plurality of cells as well as in smaller sub-populations.

762F

Clinical application of next generation sequencing in Omani patients with gastrointestinal tumors. *A. Alsaegh[1], M. AlMondhari[2].* 1) Genetics and Developmental Medicine Clinic, Sultan Qaboos University, Muscat, Oman; 2) Sultan Qaboos University, Muscat, Oman.

    Cancer genetics service was officially established in Sultanate of Oman in August 1st 2012 at the Genetics and Developmental Medicine Clinic in Sultan Qaboos University Hospital. Individuals with diagnosed gastrointestinal tumors (GI) were referred mainly from the oncology department from Sultan Qaboos University Hospital and other tertiary hospitals in the country. From 2012 to 2016, a total of 45 patients with GI tumors underwent Next Generation Sequencing (NGS) <u>as a first line investigation</u> . 28/45 (62%) patients were found to have pathogenic mutations associated with familial cancer syndromes , out of which 18 mutations associated with lynch syndrome, 9 with familial adenomatous polyposis and 1 with PTEN mutation. This study emphasizes on the clinical applications of NGS and the high diagnostic yield in identifying individuals with high risk familial cancer syndromes associated with GI Tumors.

763W

Diagnostic detection, characterization & classification of partial single exon deletions in hereditary cancer genes. *P. Fernandes, B. Morris, K. Brown, D. Mancini-DiNardo, K.R. Bowles.* Myriad Genetics Laboratories, Inc., Salt Lake City, UT.

    Background: Hereditary cancer syndromes are caused by pathogenic genetic variants, including sequence variations and large rearrangements (LRs). Methodologies such as Chromosomal Microarray Analysis, exon-based targeted Microarray-CGH, Next Generation Sequencing (NGS) and Multiplex Ligation Probe Amplification are often used to detect large deletions and duplications. However, LRs involving partial exons require additional follow-up in order to determine the exact breakpoints for accurate characterization. Here, we describe our testing strategy for confirming LRs involving partial exons and providing appropriate clinical classifications. **Methods:** Our laboratory uses exon-based targeted Microarray-CGH and/or NGS dosage analysis to identify LRs as part of a 25-gene hereditary cancer panel. In addition, Long Range PCR (LR PCR) and subsequent Sanger sequencing analyses are utilized for breakpoint determination of large deletions and duplications involving partial exons. **Results:** To date, we have identified 43 unique LRs (41 deletions and 2 duplications) involving partial exon deletions/duplications in individuals referred to our laboratory for hereditary cancer testing. 38/43 of these LRs were classified as 'pathogenic' or 'likely pathogenic' based on disruption or loss of critical gene regions or removal of consensus splice junctions, most likely resulting in abnormal RNA splicing. The remaining 5 LRs were classified as 'uncertain' based on specific criteria. These include, but are not limited to, LRs with a deletion of in-frame non-critical gene region(s), the deletion of a region located after the last known deleterious mutation, or terminal duplications of unknown location or orientation. LR PCR was performed in 15/43 cases, as breakpoint determination could significantly impact classification. One such case was the analysis of a *BRCA2* LR ("partial del exon 11") where we confirmed the presence of an in-frame deletion of 711 base pairs within a non-critical region, leading to an 'uncertain' classification. **Conclusions:** Using a combination of exon-based targeted Microarray-CGH and/or NGS dosage analysis along with LR PCR and Sanger sequencing, we are able to resolve the precise endpoints of a subset of LR deletions and duplications involving partial single exons. This leads to appropriate classifications of these rearrangements and provides healthcare providers with accurate information to appropriately guide patient management.

764T

High-throughput characterization of *BRCA1* **and** *BRCA2* **splicing defects.** *R. Karam, S. Farber-Katz, V. Hsuan, S. Wu, H. Vuong, H.M. Lu, R. Huether, S. Lam, J. Hoo, A. Elliott. Ambry Genetics, Aliso Viejo, CA.*

Statement of purpose: Advances in DNA Next Generation Sequencing (NGS) techniques resulted in the detection of many thousands of germline unclassified variants with unknown functional impact. Variants of unknown significance (VUS), especially in clinically actionable genes such as the hereditary breast and ovarian cancer susceptibility genes *BRCA1* and *BRCA2* (OMIM 113705 and 600185 respectively), pose significant challenges to the medical community and patients. Among the variants that are frequently classified as VUS are those with unclear effects on splicing. Currently, there are no cost effective and high-throughput assays to quantify and characterize germline splicing defects in a time-frame necessary for clinical testing. Here we discuss the validation and implementation of a novel high-throughput RNA-NGS assay designed to perform *quantitative* and *qualitative* characterization of splicing VUS. Methods: We compared the gold-standard mRNA splicing assays recommended by members of the ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles) consortium, including Capillary Electrophoresis and Sanger sequencing of subcloned transcripts, to high-throughput RNA-NGS assays. Following the ENIGMA protocol we performed cDNA analysis of lymphoblastoid cell lines (LCLs) generated by the kConFab consortium from carriers of *BRCA1* or *BRCA2* variants known to be associated with splicing defects, and of control LCLs, blood samples, and breast tissue. In parallel, we performed RNA-NGS assays and compared the results derived from these analyses to evaluate sensitivity and specificity of our assays. Summary of Results: Using these techniques described above, we were able to detect and characterize the splicing aberrations described in the literature by the ENIGMA consortium. Differences in protocols allowed us to determine whether RNA-NGS assays are a reliable alternative for clinical characterization of *BRCA1* and *BRCA2* splicing alterations. RNA-NGS provides a cost effective and high-throughput alternative to the gold-standard with a reduced turnaround time, thereby improving the interpretation of splicing variants detected on clinical genomic tests.

765F

Implementation of a multi-gene panel testing for hereditary cancer (qCanceRISK) in a hereditary cancer and genetic counselling unit. *M. Palacios[1], A. Abulí[1], B. Rodríguez-Santiago[2], L. Armengol[2], R. Fábregas[3], M. Cusidó[4], X. Estivill[1,5,6]. 1) Unit of Medical Genomics, Department of Obstetrics, Gynaecology and Reproduction. Dexeus Women's Health, Barcelona, Barcelona, Spain; 2) Research and Development Department, qGenomics Laboratory, Barcelona, Spain; 3) Gynaecological Cancer Unit, Department of Obstetrics, Gynaecology and Reproduction. Dexeus Women's Health, Barcelona, Barcelona, Spain; 4) Unit of Oncological Risk.Department of Obstetrics, Gynaecology and Reproduction. Dexeus Women's Health, Barcelona, Barcelona, Spain; 5) Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology. Barcelona, Spain; 6) Experimental Genetics Division, Sidra Medical and Research Centre. Doha, Qatar.*

Introduction Advances in next-generation sequencing (NGS) technology have allowed testing multiple genes at a reasonable price and time. Several multi-gene panels have been introduced in hereditary cancer. However, their clinical utility has not been fully established. Nevertheless, multi-gene cancer panels can help to identify causative mutations in high-risk patients with clinical implications in their management and in genetic testing for at-risk relatives. **Material and Methods** Patients who underwent a multi-gene NGS panel (qCanceRISK) at Dexeus Women's Health were included in the study. Patients with personal or family history of cancer had an oncology risk assessment visit and were offered genetic testing based on risk assessment. Patients received pre and post-test genetic counselling. The panel consists of 100 genes associated with risk for breast and other types of cancer. **Results** A total of 89 individuals were included in the study. Approximately 81% (72/89) of patients had personal and 19% (17/89) had family history of cancer. Clinical criteria for hereditary breast and ovarian cancer (HBOC) were met by 61% (54/89) of families and 20% (19/89) had moderate risk. Only one patient met clinical criteria for hereditary non-polyposis colon cancer (HNPCC). We found a pathogenic mutation in 12% (11/89) and a likely pathogenic mutations in 2.2% (2/89) of cases with breast cancer and one pathogenic mutation in *MLH1* in the patient who met HNPCC criteria. Pathogenic and likely pathogenic mutations were found in *BRCA1*, *BRCA2*, *PALB2*, *PMS2* and *NBN* genes. A heterozygous pathogenic mutation was also found in *MUTYH*. Variants of unknown significance (VUS) were found in 37% (33/89) of cases. The global detection rate of qCanceRISK panel was 15% (13/89) and for high-risk patients 20% (11/55). **Conclusions** Our study identified mutations in other genes that would have been missed if classic testing of *BRCA1* and *BRCA2* would had been performed; however, we also found VUS in an important percentage of patients. Multi-gene panel testing should be performed in the context of an adequate pre and post-genetic counselling sessions, in which transmitting to the patient information regarding incidental findings and the higher detection rate of VUS are the main challenge.

766W

The *PTEN* promoter: To be or not to be included on multigene panel tests? *R. Pilarski[1], S. Li[2], R. McFarland[2], H. LaDuca[2], B. Smith[2], T. Pesaran[2].* 1) Division of Human Genetics, Comprehensive Cancer Center, Ohio State Univ, Columbus, OH; 2) Ambry Genetics Corp, Aliso Viejo, CA.

Background Pathogenic *PTEN* alterations have been associated with an increased risk of developing cancers of the breast, thyroid, endometrium, and others as part of *PTEN* hamartoma tumor syndrome. Although several alterations in the *PTEN* promoter region have been described as mutations, data are insufficient to support a classification of pathogenicity and they are currently characterized as variants of unknown significance (VUS) by clinical labs. We sought to assess the association of *PTEN* promoter variants with breast and other cancers in a cohort of 61,552 patients undergoing *PTEN* testing. Methods Clinical histories of patients who underwent comprehensive *PTEN* analysis from January 1, 2014 to March 31, 2016 as part of a multigene panel testing (MGPT) were retrospectively reviewed. Personal and family cancer histories of *PTEN* promoter variant carriers (PV; N=704) were compared against *PTEN* mutation-positive patients (MP; N=57) and also matched mutation-negative individuals (WT; N= 2,194). The frequency of MP, PV and WT were compared between patients with and without selected cancer phenotypes using the Fisher's exact test and multivariate logistic regression analysis, controlling for age at testing, MGPT ordered, ethnicity and gender. Age differences were tested using Welch t-test or Wilcoxon rank test based on the data distribution. Results *PTEN* promoter variants were observed with a frequency of 1.1% (704/61,552). PV individuals were significantly less likely than MP individuals to have a personal history of female breast cancer (p=0.003), bilateral breast cancer (p=0.014), uterine cancer (p=0.001), or kidney cancer (p=0.002). No differences were observed when comparing personal history of colon cancer or thyroid cancer, or family history of female breast cancer among first degree relatives between these two groups. No significant differences were observed when comparing PV individuals to WT individuals. MP individuals had a younger diagnosis age of female breast cancer (p=0.022; p=0.021), endometrial cancer (p=0.027; p=0.015) and thyroid cancers (p=0.042; p=0.032) compared to WT and PV individuals, respectively. Conclusions In this cohort, *PTEN* promoter variants were not associated with an increased risk of *PTEN*-related cancers. These results do not support the inclusion *PTEN* promotor sequencing on cancer panels, as the identification of these alterations contributes to increased VUS burden without increasing diagnostic yield.

767T

A broader genotype-phenotype spectrum in hereditary colorectal cancer syndromes. *A.M.H. Rohlin[1], E. Rambech[2], A. Kvist[2], T. Törngren[2], F. Eiengård[1], U. Lundstam[3], T. Zagoras[1], S. Gebre-Medhin[4,5], A. Borg[2], T. Olausson[1], J. Björk[6], M. Nilbert[2,7], M. Nordling[1].* 1) Department of Molecular and Clinical Genetics, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; 2) Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Medicon Village, SE 22381 Lund, Sweden; 3) Department of Surgery, Sahlgrenska Academy at University of Gothenburg, Sahlgrenska University Hospital/Östra, SE 416 85 Gothenburg, Sweden; 4) Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden; 5) Department of Clinical Genetics, Office for Medical Services, Division of Laboratory Medicine, Lund, Sweden; 6) The Swedish Polyposis Registry, Department of Medicine, Karolinska Institute, Stockholm, Sweden; 7) The HNPCC-register, Hvidovre University Hospital, Copenhagen University, Hvidovre, Denmark.

Hereditary colorectal cancer (CRC) can be divided into syndromes. High penetrant mutations in known CRC genes explain 5-6% of the cases. Clinical testing is usually preformed based on phenotype of the different syndromes. However, overlapping phenotypes between CRC syndromes, including both polyposis and non-polyposis, present diagnostic difficulties in the clinic. It is crucial that the patient get a correct molecular diagnosis that allows for adequate follow-up since the majority of the syndromes include predisposition for tumors also in other organs.   To address this concern we constructed a panel consisting of 19 CRC susceptibility whole gene regions including introns. Analysis of SNPs, indels and CNVs were performed on approximately 100 individuals divided into clinical subtypes based on phenotype.   After filtration by a comprehensive clinical pipeline we detected several pathogenic mutations. In patients diagnosed with classical FAP, attenuated FAP, atypical FAP and none-polyposis subgroups, we identified mutations in *BMPR1A* and *SMAD4*. These two genes are mutated in juvenile polyposis. We also detected novel CNVs in upstream regions of *SMAD4*, *MSH3*, *CTNNB1* and one deletion in an intronic region of *CDH1*   By using this gene panel we obtained increased mutation detection frequency for CRC syndromes which enable appropriate follow of patients based on the clinical feature of each syndrome. We have also found misdiagnoses for especially juvenile polyposis, which also broadens the genotype-phenotype spectrum for these genes. By using this gene panel more extensive possibilities to detect deletions and other structural rearrangements can be achieved. This lies beyond the potential of the commonly used techniques eg MLPA, and also includes a reduction in costs and work load.   .

768F

Clinical NGS Pipeline outperforms a combined approach using Sanger sequencing and MLPA in targeted gene panel analysis. *L. Schenkel[1], J. Kerkhof[2], A. Stuart[2], J. Reilly[2], C. Howlett[1,2], P. Ainsworth[1,2], B. Sadikovic[1,2].* 1) Department of Pathology and Laboratory Medicine; Western University; London, ON, Canada; 2) Molecular Genetics Laboratory; Molecular Diagnostics Division, London Health Sciences Centre, London, ON, Canada.

Recent advances in next-generation sequencing (NGS) have enabled parallel analysis of multiple genes allowing the implementation of more cost-effective, rapid and high-throughput methodologies for the molecular diagnosis of multiple genetic conditions. Clinical-grade NGS analysis must meet high standards of sensitivity and specificity for both sequence mutations and structural rearrangements including copy number alterations. The objective of this study was to develop a single NGS-based pipeline to replace clinical Sanger sequencing and MLPA copy number analysis, first for hereditary breast and ovarian cancer testing, and then to expand to other clinical gene panel tests. We analyzed by NGS 402 retrospective patients, with previous Sanger and MLPA confirmation of HBOC, and 240 clinical prospective patients. Custom sequence capture probes were designed using the SeqCap EZ Choice Library system and included enrichment for 46 exons, including *BRCA1* exons 2, 3, 5-24 and *BRCA2* exons 2-27, with 20 nucleotides of intronic regions both 5' and 3' of each exon. The DNA sequencing was conducted using an Illumina MiSeq platform with 24 different patient samples multiplexed per run. Sequence analysis for variant identification, alignment and coverage distribution was performed with NextGene software v2.4.1, and copy number variant analysis was performed using single nucleotide resolution depth of coverage and a quantile normalization algorithm. Within those patients, 183 unique variants, including sequence and copy number variants were detected in the retrospective (95) and prospective (88) cohorts. This standardized NGS pipeline demonstrated 100% sensitivity and 100% specificity, in addition to uniformity and high depth nucleotide coverage per sample (approximately 7000 reads per nucleotide). Using the same approach, this NGS pipeline was applied to the analysis of larger gene panels, including Charcot Marie Tooth, hereditary cancer syndromes, mitochondrial genome, epilepsy syndromes, lysosomal and urea cycle disorders, all of which have shown similar uniformity, sample to sample reproducibility in coverage distribution, and sensitivity and specificity for detection of sequence and copy number variants. In conclusion, we have validated a clinical-grade NGS pipeline which outperforms the "gold standard" Sanger sequencing and MLPA for sensitive and specific detection of sequence and copy number alterations in a single test.

769W

Standardized multiplex genetic testing in a population-based hereditary cancer clinic. *K.A. Schrader[1,2], T.R. Docking[1], I. Bosdet[1], S. Young[1], T. Tucker[1], S. van den Berg[2], M. He[1], Z. Lohn[1], J. Nuk[1], C. Cremin[1], J. Scott[1], A. Mindlin[1], C. Portigal-Todd[1], K. Turner[1], J. Thompson[1], A. Bedard[1], L. Armstrong[2], B. McGillivray[2], G. Aubertin[1], S. Sun[1], G. Mitchell[1], HCP Cclinical Team[1], R. Moore[1], A. Mungall[1], L. Swanson[1], L.P. Yang[1], J. Slind[1], M. Balasundaram[1], J. Schein[1], A. Karsan[1,2].* 1) BC Cancer Agency, Vancouver, British Columbia, Canada; 2) Univ British Columbia, Vancouver, British Columbia, Canada.

The BC Cancer Agency's Hereditary Cancer Program (HCP) is the sole provider of cancer genetics services to the population of British Columbia and the Yukon. Publically funded testing is currently conducted at the Centre for Clinical Genetics (CCG) using a 14-gene panel for affected individuals ages 19 and over meeting testing criteria for at least one syndrome on the panel. The HCP has utilized this panel which tests for Hereditary Breast and Ovarian Cancer (*BRCA1, BRCA2*), Li Fraumeni (*TP53*), Cowden (*PTEN*), Hereditary Diffuse Gastric Cancer (*CDH1*), Peutz Jeghers (*STK11*), Lynch Syndrome (*MLH1, MSH2, MSH6, PMS2*), MYH-Associated Polyposis (*MUTYH*), Familial Adenomatous Polyposis (*APC*), and Juvenile Polyposis (*SMAD4, BMPR1A*) since October 2014. The uptake of multiplex genetic testing in the hereditary cancer clinic has altered the way in which patients and families are counseled regarding genetic testing, with increased attention given to the risk for identification of variants of uncertain significance and unexpected diagnoses. Interim analysis of 411 cases revealed variants of uncertain significance in 42% of individuals tested. Examination of the subpopulation of cases meeting criteria for Hereditary Breast and Ovarian Cancer revealed the rate of positive findings in genes other than *BRCA1* and *BRCA2* to be 10% (4/38). We will present aggregate results from our consecutive experience to date of over 1000 individuals who have been tested through the HCP using the 14-gene panel approach. We will present mutation detection rates for both expected and unexpected presumed pathogenic germline variants, variants of uncertain significance, and the frequency of findings per individual. Results from this population-based ascertainment with extensive phenotypic data will help provide important insights into potential germline findings using panel approaches and may provide guidance in settings that analyze panels of genes in affected individuals, such as secondary analysis of cancer genes in tumor-normal sequencing. These data demonstrate that even within relatively small testing panels, phenotypic heterogeneity is seen within well-defined syndromes, likely due to previous biases in ascertainment and testing criteria.

770T

A randomized controlled trial of exome sequencing vs. usual care in patients with colorectal cancer/polyposis syndromes: Preliminary results. *D.L. Veenstra[1], L.M. Amendola[2], P. Heagerty[3], B. Comstock[3], M. Horse-Pyne[2], F. Hisama[2], M. Li[1], A. Burt[2], R.L. Bennett[2], M.O. Dorschner[4], P. Tarczy-Hornoch[5], S.M. Fullerton[6], S.B. Trinidad[6], D.A. Regier[7], D.L. Patrick[8], D.A. Nickerson[9], W. Burke[6], G.P. Jarvik[2].* 1) Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA; 2) Medical Genetics, University of Washington, Seattle, WA; 3) Biostatistics, University of Washington, Seattle, WA; 4) Pathology, University of Washington, Seattle, WA; 5) Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 6) Bioethics and Humanities, University of Washington, Seattle, WA; 7) Canadian Centre for Applied Research in Cancer Control, BC Cancer Agency, Vancouver, BC; 8) Health Services, University of Washington, Seattle, WA; 9) Genome Sciences, University of Washington, Seattle, WA.

   Background. Next generation DNA sequencing is a potentially transformative technology in clinical care, but little evidence exists as to whether this approach improves patient outcomes or provides good economic value. This knowledge gap can be addressed by comparative effectiveness and patient-centered outcomes research. **Methods.** Patients seeking genetic testing to evaluate hereditary colorectal cancer and polyposis syndromes (CRCP) were randomized to usual care (UC) vs. UC + whole exome sequencing (WXS). Patients received findings associated with colorectal cancer and additional findings in separate visits. The primary outcome was whether the patient had any detection of pathogenic variants associated with CRCP; secondary outcomes included psychosocial impact and healthcare resource utilization. Additional findings returned were determined by a formal return of results committee, and initially consisted of incidental findings from 112 medically actionable gene-disease pairs, and have been expanded to include a limited set of pharmacogenomics variants and carrier status genes. UC was dependent on the provider, but in general evolved from smaller to larger CRCP gene panels during the course of the study. **Results.** A total of 65 patients were randomized to usual care (UC), and 68 were randomized to UC+WXS. The proportion of patients with a CRCP variant that was pathogenic, likely pathogenic, or of unknown significance was 24.6% (16/65) in the UC group vs. 32.4% (22/68) in the UC+WXS group, (p=0.3); the proportion with a pathogenic variant was 6.2% vs. 7.4%, respectively. Additional findings in the 52 patients in the WXS+UC arm who had been sequenced included a disease finding (strong genetic risk) in 5.8% of patients and a pharmacogenomic finding in 73.1% of patients. Patient-reported anxiety, depression, and test result-specific psychosocial impact did not differ between groups. Healthcare resource utilization was similar between groups, although a higher proportion of patients in the WXS+UC group thought about making life insurance changes. **Conclusions.** Our preliminary results indicate exome sequencing led to a 1.3 fold increased genetic diagnosis for CRCP cases that was not statistically significant, return of clinically important additional findings, and no differences in average patient quality of life or healthcare costs. Larger studies are needed to achieve power to confirm these findings.

771F

To panel or not to panel: Do genetic professionals differ in NGS testing preferences? *G.L. Wiesner[1,2], S. Lewis[1], K. Taylor[1], K. McReynolds[1], S.K. Rao[1].* 1) Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN.

   Introduction: The advent of NGS testing has ushered in a new era in genomics where genetic testing has become easily accessible and yet more complicated to interpret. Literature is lacking about current genetic panel test ordering preferences amongst genetics professionals. The Hereditary Cancer Genetics Clinic at Vanderbilt University Medical Center is uniquely placed in having credentialed genetics providers of various backgrounds; 3 genetic counselors (GC), 1 genetics advance practice nurse (NP) and 2 geneticists (MD); that see patients at high risk for Hereditary Breast/ Ovarian Cancer (HBOC) syndrome. We report on our own experiences in cancer genetic testing. **Methods:** A retrospective analysis was done of the Hereditary Cancer Registry of patients seen during a 30 month period between July 2012- Dec 2014. Study data were collected and managed using RedCap electronic data capture tools hosted at Vanderbilt University (Harris et al; *J Biomed Inform* 2009). Parameters analyzed were test type (panel vs. single gene; single site excluded), and provider type (GCs vs MD/NPs) stratified by breast/ovarian cancer affected and unaffected probands, over first and second half of time period described. Summary statistics and chi square analyses were done. **Results:** For breast/ovarian patients (n=535); 48.6% (260/535) tests were ordered by MD/NPs and 51.4% (275/535) by GCs. Of these, 69% (179/260) of MD/NP orders were panel tests versus 55% (151/275) by GCs (p= 0.0004). For unaffected probands with a family history of breast/ovarian cancer (n=188); MD/NPs (80.7%) and GC (72.5%) patterns were similar (p= 0.1786). An analysis comparing the first and second 15 month periods showed that NGS panel orders rose overall from 34.6% to 82.3% (GCs 17% to 76.5%; MD/NPs 48.8% to 88.8%;p= 0.0001). **Conclusion:** This data indicates a significant difference in test ordering profiles between genetic professionals. Genetic counselors appear to be more conservative in NGS panel testing compared to MD/NPs geneticists. The time progression analysis shows an increased uptake overall with the clinician providers adopting panel testing earlier and the counselor group making a more gradual transition. Some parameters not analyzed in this study were patient's insurance coverage, or surgical decision constraints which might affect NGS panel test ordering decisions. Nonetheless, these data will provide a foundation for further studies into the uptake of new technologies into clinical care.

772W

Low cost liquid biopsy by selective sequencing of mutant DNA. *M.J. Wiggin[1], L. Mai[1], S. Walsh[1], D. Broemeling[1], A. Marziali[1,2]. 1) Boreal Genomics, Vancouver, BC, Canada; 2) University of British Columbia Dept. of Physics & Astronomy, Vancouver BC, Canada.*

Clinical circulating tumor DNA (ctDNA) assays (liquid biopsy) require high sensitivity and specificity, multiplexing over multiple mutations, and importantly, low cost. Achieving all of these requirements simultaneously is very challenging, and the most common approaches to liquid biopsy assays are limited by either very low multiplexing (droplet digital PCR, ddPCR), or very high cost (barcoded sequencing), making many clinical applications impractical. In sequencing-based liquid biopsy assays, majority of reads are wasted re-sequencing wild-type cell free DNA (cfDNA) fragments that deliver no useful information about the patient. Therefore, a powerful solution to the problems of cost, sensitivity, and specificity is to separate mutant cfDNA from wild-type cfDNA so that only tumor molecules are sequenced. Given that in many clinical cases the ctDNA represents only 0.1% - 0.01% of the total cfDNA, this approach has the potential to reduce the required sequencing by 1,000 – 10,000 fold, dramatically reducing sequencing cost. By narrowing the sample to contain primarily mutant ctDNA, both the sequencing cost and false positive sequencing errors are greatly reduced. We have developed such a technology, which can selectively enrich for hundreds of different mutations simultaneously from a single sample, which we call the OnTarget assay. We present the technical details of the OnTarget assay, including the underlying method used to enrich samples for mutations, and results from an analytical validation study using controlled reference samples. These results demonstrate limit of detection measurements on each mutation in the assay, demonstrating that OnTarget performance is similar or better than ddPCR on individual mutations, while testing many more mutations in parallel, at a fraction of the cost of barcoded sequencing assays. We also present data from clinical samples in studies that include both cancer patients and healthy individuals, as a precursor to larger clinical studies for population-wide screening and early cancer detection. In summary, the addition of wild-type DNA depletion to NGS assays enables lower cost, high sensitivity, and high specificity liquid biopsy tests, enabling commercialization of this powerful diagnostic in applications with limited reimbursement, potentially including early cancer detection.

773T

What do public databases really tell us about classification of variants in *BRCA1* and *BRCA2*? *R. Nussbaum[1], S. Lincoln[1], S. Yang[1], M. Cline[2], C. Zhang[2], Y. Kobayashi[1], S. Topper[1], D. Haussler[2], B. Paten[2]. 1) Invitae, San Francisco, CA; 2) University of California, Santa Cruz, CA.*

The ClinVar database has collected over 170,000 de-identified records from commercial and academic clinical laboratories. However, certain laboratories not only decline to submit data, but they also prominently highlight disagreements in various public databases, suggesting that these differences "preclude their wider use in clinical practice". While disagreements are real, this proposition is inconsistent with our experience. For example, in a published study of 975 patients, we observed 99.8% concordance of reports in BRCA1 and BRCA2, comparing actionable vs. not-actionable classifications (a) based on publicly available data, the literature, and recent guidelines, with (b) classifications from another lab who also used a large proprietary database. This result is intuitive because experienced and responsible lab directors never simply copy variant classifications from any public database. Instead, they critically evaluate underlying evidence and report classifications following established guidelines. In this study we sought additional data to address this question. We gathered ClinVar submissions from established clinical laboratories, most from the last 5 years, including Myriad Genetics' classifications submitted to ClinVar by the Sharing Clinical Reports Project (SCRP). For variants in BRCA1/2, concordance between pairs of labs is high (98.5%) in a data set that represents test results from roughly 20,000 patients. Importantly, all of the discordant interpretations were seen in rare variants that, by definition, are present in very few patients. Moreover most rare variants (98.4%) have fully concordant classifications. We used prevalence data and population allele frequencies to estimate that 99.8% of patients would be expected to receive concordant reports from two labs for BRCA1/2, a similar finding to our previous study's results. While substantial disagreements in BRCA1 and BRCA2 are infrequent, they are of course important to patients and their clinicians. It is essential that we as a community work to resolve these differences collaboratively in order to deliver the best patient care, as is done in other areas of medicine. It is also essential that we establish standards for how concordance is measured and reported in order to avoid misleading clinicians.

774F

Targeted next-generation sequencing for the identification of genomic BCR-ABL1 fusion junctions to quantify residual disease in CML patients in CMR. *M. Alikian[1,2], P. Ellery[2], M. Forbes[1], D. Kasperaviciute[3], A. Sozinski[3], L. Foroni[1], J. Apperley[1].* 1) Centre for Haematology, Imperial College London, London, United Kingdom; 2) Imperial College London NHS trust, London, United Kingdom; 3) Clinical Sciences Centre and bioinformatics facility, Imperial College London, London, UK.

Recent studies indicate that 40% of CML patients who achieve complete molecular remission (CMR) on imatinib remain disease-free after drug discontinuation, raising the possibility of an "operational cure". However, the safe introduction of a TKI withdrawal policy would require a reliable and cost effective method of identifying patients with the lowest likelihood of relapse, which is likely to be related to presence residual disease. Preliminary data suggest that PCR of genomic DNA might be more sensitive for the detection of residual disease than one that relies on cDNA and may therefore help to predict outcome post-withdrawal. However, the former method is arduous since it requires a customised patient-specific assay. Here we describe a method based on targeted-next-generation sequencing allowing identification of BCR-ABL1 breakpoints from enriched genomic BCR and ABL1 DNA followed by rapid generation of DNA-based qPCR assays. The location of the BCR-ABL1 fusion junction was mapped in disease samples from 36 CML patients using Illumina's MiSeq platform. A custom TruSeq DNA target enrichment kit (Illumina) was used to enrich for the BCR and ABL1 genes. Subsequent mapping of t(9;22) translocation junctions was performed via a custom designed bioinformatics algorithm. All breakpoints were successfully mapped. DNA qPCR assays were designed and validated for 18 patients. In clinical samples from patients in complete molecular remission, the RT-qPCR assays detected residual disease in 6 out of 18 patients (33.3%), demonstrating that DNA-qPCR can detect residual disease in patient samples in which CML cells persist below the detection threshold of RT-qPCR. Furthermore, we investigated disease status in 48 CMR samples coming from 6 CML patients (3 positive and 3 negative by DNA-qPCR) using digital PCR (dPCR)(Fludigm BioMark HD platform) which allows single DNA molecule detection. We found that all 6 patients had detectable disease signals and that the level of positivity was reducing over time despite being negative by DNA-qPCR. In Conclusion, NGS-facilitated DNA-dPCR may therefore prove valuable for the stratification of patients with low levels of residual disease and, therefore, in the identification of patients for whom TKI therapy could be safely reduced or stopped.

775W

Accurate detection of copy-number changes in the highly conserved regions of *PMS2* using NGS. *V. Ramey, M. Kennemer, D. Kvitek, K. Ouyang, R. Harte, S. Lincoln, S. McCalmon, N. Patil, S. Aradhya.* Invitae, San Francisco, CA.

Background: Lynch syndrome is caused by inherited mutations in PMS2 in 4-11% of cases. However, testing for PMS2 is hampered by the presence of a pseudogene, PMS2CL, which is nearly identical to PMS2 in both intronic and exonic regions of exons 12–15. Ordinary NGS reads cannot be unambiguously aligned to PMS2 or PMS2CL, and gene conversion between PMS2 and PMS2CL further complicates this issue.   Methods: We developed a unique process to test PMS2 for copy number variants (CNVs). The first step is a bioinformatics screen in which NGS reads are analyzed for potential CNVs using a modified, PMS2CL masked reference sequence. We then use multiplex ligation-dependent probe amplification (MLPA) [2] to confirm CNV calls made by the screen. Finally we use long-range PCR (LR-PCR) and sequencing of the MLPA targeted paralogous sequence variant (PSV) or an alternative NGS-detected variant to disambiguate the CNV location.   Results: A sensitivity of 100% and specificity of 100% for this method has been demonstrated using specimens tested using traditional methods. Clinical cases will be shown where our NGS methods detect variants that the MLPA based approach would still find ambiguous.   Conclusions: Detecting CNVs by NGS in general can be challenging, although methods are available that achieve high clinical performance standards [2]. Regions with high homology however remain challenging. Our approach addresses this issue and can help reduce the cost while improving accuracy of comprehensive genetic tests.   [1] Vaughn et al., Human Mutation, 2011 [2] Jacobs et al., CSHL 2013.

776T

Development of highly multiplexed reference materials for inherited cancer testing. *R. Santhanam[1], C. Huang[2], J. Dickens[2], D. Yuzuki[2], R. Garlick[1], B. Anekella[2].* 1) SeraCare Life Sciences, Milford, MA; 2) SeraCare Life Sciences, Gaithersburg, MD.

   Introduction: Next Generation Sequencing (NGS) has improved the diagnostic yield of genetic testing, but has presented new challenges for test development, validation, and quality control management. Genetic tests for increased cancer risk are challenging due to variability of NGS workflow and difficulty in mapping large or complex variants. Reference materials are fundamental for monitoring NGS assays and evaluating both wet-lab procedures and software analysis pipelines. We sought to develop a reference material that assesses detection of variants that are challenging to detect including large indels (>10 bp), insertion of Alu elements, tandem duplications and combined deletion/insertion elements.   **Methods:** Human genomic DNA was extracted from the well-characterized GM24385 cell line and blended with biosynthetic constructs bearing human heritable mutations. Blending was performed to achieve 50% allele frequency and allele specific digital PCR (dPCR) verified mixing accuracy. A pilot study used a biosynthetic construct with nine (9) SNVs and one (1) large (25 bp) deletion. Testing on two different capture-based NGS assays in different laboratories suggested that a reference material that focused on the more challenging variants, with a few simple SNVs that represented common pathogenic variants would be most useful. **Results:** Preliminary testing of biosynthetics showed SNVs were detected with the expected allele balance at 50%. Testing of the 25 bp deletion was detected at ~30% frequency by both NGS assays, even though the variant frequency in the reference material was verified at 50% by dPCR; the difference was attributable to challenges mapping variants of this size. The results from the pilot study were used to design a new reference material containing portions of seven (7) human genes including MSH2, MSH6, MLH1, PMS2, CDKN2A, BRCA1 and BRCA2. Each gene segment contains one or more variants of clinical significance such that there are eleven (11) large/complex variants and six (6) common pathogenic SNVs and small indels.   **Conclusions**: Seraseq™ Inherited Cancer Reference Material is a comprehensive reference material designed to challenge NGS based genetic tests for inherited cancer susceptibility and fulfill unmet needs for accurate, multiplexed quality controls. These materials challenge variant detection pipelines, aid in assay optimization, and provide assurance in the ability to detect complex mutations on an ongoing basis.

777F

Single-cell genetic analysis reveals insights into clonal development of cervical cancer and confirms gain of *TERC* **as an early and dominant aberration.** *K. Heselmeyer-Haddad[1], A. Bradley[1], L. Hernandez[1], I. Torres[1], S. Andersson[2], T. Gaiser[3], D. Hirsch[3], P. Ströbel[4], S.A. Chowdhury[5], E.M. Gertz[6], R. Schwartz[7], A.A. Schäffer[6], T. Ried[1].* 1) Genetics Branch, CCR, NCI, Bethesda, MD; 2) Karolinska University Hospital Solna, Department of Women and Children's Health, Karolinska Institutet, Stockholm Sweden; 3) Institute of Pathology, University Medical Center Mannheim, Germany; 4) Institute of Pathology, University Medical Center Göttingen, Germany; 5) Simons Center for Data Analysis, Simons Foundation, New York, NY and Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ; 6) Computational Biology Branch, National Center for Biotechnology Information, NIH, MD; 7) Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA.

   We developed a Multiplex Fluorescence in situ Hybridization (FISH) approach allowing enumeration of 20 different gene loci within individual nuclei from archival histological patient samples to assess tumor heterogeneity and clonal evolution. We have previously shown that this assay allowed us to gain new insights into breast tumor evolution analyzing synchronous DCIS and invasive breast cancer.   Our current project aims to shed light on cervical tumor evolution. The assay targets candidate genes in locations that are frequently gained in cervical cancer, including *COX2* (1q), *TERC* (3q, most commonly gained), *TERT* (5p), *MYC* (8q), *ZNF217* (20q) or lost, including *ING5* (2q), *FHIT* (3p) and *CHEK1* (11q).   Lesions representative of cervical intraepithelial neoplasia 3 (CIN3) and invasive cervical carcinoma (ICC) were macrodissected from 50 μm sections of FFPE tissue blocks. A subset of these lesions were synchronous CIN3 and ICC. Sequential hybridizations of multicolor FISH cocktails were analyzed with a custom-made automated spot counting software. Data were analyzed with algorithms developed to deduce phylogenetic trees.   Nineteen out of 20 lesions (11/11 CIN3, 8/9 ICC) showed a major clonal population with a gain of *TERC* confirming the dominant role of this aberration. Nine out of the 20 lesions showed a concomitant loss of *FHIT*, suggesting an isochromosome 3q formation. Most lesions were dominated by a major clone comprising around 30-40% of all cells analyzed. Major clones could be either explained by a single event, like an isochromosome 3q formation, or were more complex involving aberrations for several or most of the genes analyzed. The only lesion without a *TERC* gain was defined by a *MYC* gain in 90% of the cells. Interestingly, gains of *MYC* were more prevalent in ICC compared to CIN3 (78% vs. 27%), indicating that *MYC* might play a role in progression. Of note, the synchronous CIN3 and ICC analyzed did not show similar major clones, indicating independent evolution (multifocal disease). However, their only commonality was a *TERC* gain, confirming the importance of this aberration in cervical carcinogenesis. While overall levels of heterogeneity were indistinguishable between CIN3 and ICC, phylogenetic models of single tumors suggested greater average numbers of aberrations per cell in ICC versus CIN3. Our study will improve the understanding of clonal development during cervical carcinogenesis and genome dynamics at early stages of carcinogenesis.

778W

Functional analysis of variants of uncertain clinical significance in components of the TORC1 and RAS signaling pathways. *M. Nellist, M. Hoogeveen-Westerveld, M. Kroos-de Haan, P. Elfferich, J. Boonman, G. Mancini, Y. van Bever, R. van Minkelen, F. Verheijen.* Functional Unit, Department of Clinical Genetics, Erasmus MC, Wytemaweg 80, 3015CN, Rotterdam, The Netherlands.

   One of the major problems in clinical genetic diagnostic is the interpretation of variants of uncertain clinical significance (VUS) and novel approaches to implement effective functional screening of genomic variants are required. We will present two examples that have been validated for use in our diagnostic laboratory.   1. The mechanistic target of rapamycin (mTOR) complex 1 (TORC1) is an essential protein kinase complex that controls cell growth and metabolism. Mutations in different components of the TORC1 signaling pathway are associated with a broad spectrum of inherited and somatic diseases, including cancer. We have developed simple cell-based assays to investigate the effects of VUS identified in genes encoding components of the TORC1 signaling pathway on TORC1 activity. To date we have analysed >300 variants in *AKT3*, *DEPDC5*, TBC1D7, *TSC1* and *TSC2*.   2. *NF1* encodes neurofibromin, a GTPase activating protein (GAP) for RAS and inactivating mutations in *NF1* result in increased levels of active RAS-GTP, increased RAS signaling and cause neurofibromatosis type 1 (NF1). We have performed RAS-GTP pull-down assays to assess the RAS-GAP activity of *NF1* VUS identified in our NF1 patient population.   Our work has provided insight into the likely pathogenicity of VUS in *AKT3*, *DEPDC5*, *TBC1D7*, *TSC1* and *TSC2*, the genetic risks in the families segregating the tested variants, genotype-phenotype correlations and structure-function relationships. Investigating the effects of VUS on TORC1 and RAS signaling is a useful adjunct to standard genetic testing and can provide essential information for appropriate clinical management.

779T

Variants of hereditary breast and ovarian cancer associated genes in Czech patients: The first experience with targeted NGS sequencing. *M. Koudová[1], V. Krutílková[1], F. Lhota[1], L. Černá[1], M. Bittóová[1], B. Honysová[1], F. Zembol[1], J. Libich[1], Z. Kleibl[2], D. Stejskal[1].* 1) GENNET s.r.o., Prague, Czech Republic; 2) Inst. of Biochemistry and Exp. Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic.

   Introduction: Next-generation sequencing (NGS) allows to concurrently analyze germline variants of multiple genes associated with hereditary breast and ovarian cancer (HBOC). **Methods:** 283 DNA samples of patients (148) and healthy relatives (135) fulfilling HBOC criteria according to Czech guidelines (similar as the National Comprehensive Cancer Network guidelines) were analyzed using NGS panel CZECANCA (**CZE**ch **CA**ncer pa**N**el for **C**linical **A**pplication) designed at Inst. of Biochemistry and Exp. Oncology, First Faculty of Medicine, Charles University in Prague. This panel targets 212 cancer susceptibility and candidate genes based on the genetic variability of Czech cancer patients. We focused on analysis of 13 HBOC predisposing genes (*BRCA1, BRCA2, TP53, CDH1, PTEN, STK11, CHEK2, PALB2, BRIP1, ATM, RAD51C, RAD51D, NBN)*. NGS was complemented by the MLPA screening for large genomic rearrangements of *BRCA1, BRCA2 PALB2, RAD51C, RAD51D* and deletion of exons 9-10 of *CHEK2* gene. Positive results were validated by direct sequencing. We have developed a bioinformatic pipeline using local installation of Ensembl genomic database for annotation and SQL server variant database for data handling and clinical reporting. **Results:** In total 48/283 samples (17%) a deleterious variant was detected- in 31/148 patient samples (20,9%) and in 17/135 healthy relatives samples (12,6%). Out of these, 3/283 samples (1,1%) digenic heterozygous variants were found. Variants of *BRCA1/2* genes constituted 47,1% of findings, *CHEK2* gene variants 31,3% and other gene variants 21,6% of findings. **Conclusion:** These results are consistent with the published findings that *BRCA1/2* gene mutations are responsible for up to half of the heritable mutations in HBOC and highlight the importance of extending the examination to other susceptibility genes and testing of all first degree relatives at risk if affected family member is not available.

780F

ERASE-seq: Implementation of super-accurate mutation detection at low allelic frequency and liquid biopsy applications. *N. Kamps-Hughes, C. Ionescu-Zanetti.* Research and Development, Fluxion Biosciences, South San Francisco, CA.

Purpose: Cancer genotyping experiments are increasingly utilizing blood samples to non-invasively track the mutational profile over time, requiring variant detection with increased sensitivity and specificity. We present validation of ERASE-Seq, a method for accurate detection of low frequency DNA variants that yields significant performance improvements with respect to molecular barcoding methods. Our approach differs from previous methods by using a custom replicate amplification step coupled to calling variants based on statistically significant differences in allele counts between replicate sample amplification reactions and a previously determined locus specific noise model, thereby eliminating false positives resulting from both jackpot mutations and error-prone loci. **Methods:** Analytical samples were created by spiking DNA from a number of different cell lines into human reference DNA samples down to expected variant allele frequencies of 0.1%. The ERASE-Seq assay was applied to these samples in order to determine analytical sensitivity and specificity. The clinical utility of ERASE-Seq was demonstrated on a set of circulating tumor cell samples from colorectal and pancreatic cancer patients. **Results:** ERASE-Seq low frequency variant detection in spiked DNA mixtures demonstrates perfect sensitivity and specificity down to 0.3% allele frequency across multiple pan-cancer amplicon panels. At 0.1%, the specificity remains perfect and the sensitivity remains above 70%. WGA pre-amplified samples maintain perfect specificity down to 0.3% allele frequency with ERASE-Seq but show a small number of false positives when calling variants to 0.1% (specificity = 0.9999). For clinical samples, we obtained an average of 3 high confidence calls per sample (>0.3% allele frequency), including alterations commonly found in solid tumor tissue for the relevant indications. In biological replicates from two CRC patients we have found a correlation of 80% between the draws. **Conclusions:** We present validation of ERASE-Seq, a highly sensitive solution for the detection of somatic variants from a blood draw using standard amplicon technology. The method's performance has been validated to the 0.1% allele frequency range where robust sensitivity and specificity have been demonstrated across DNA sample types. The clinical applicability of this performance was then demonstrated with robust and reproducible detection of somatic mutations in patient liquid biopsies.

781W

Characterization of *TP53* sequencing variants initially detected in peripheral blood using next generation sequencing analysis. *D. Mancini-DiNardo, B. Coffee, H.C. Cox, K. Brown, S. Manley, K.R. Bowles, B.B. Roa.* Myriad Genetics Laboratories, Inc, Salt Lake City, UT.

Background: Individuals with germline pathogenic variants (PVs) in *TP53* have Li-Fraumeni Syndrome (LFS), which is associated with a high cancer risk and early age of diagnosis. Approximately 40% of *TP53* PVs were detected at Next-Generation Sequencing (NGS) allele frequencies between 10-30% and are suspected to be somatic mosaic variants. However, internal evidence demonstrates that NGS read frequencies for somatic PVs can increase over time to overlap with those observed for true germline heterozygotes (30%-70%). Given the severe clinical implications of germline PVs in *TP53* and the relatively recent recognition of the prevalence of somatic *TP53* PVs, it has now become critical that apparent germline PVs be confirmed to enable appropriate medical management. Here, we present findings from a commercial testing laboratory program that offers confirmatory analysis to all individuals with an apparent germline *TP53* PV. **Methods:** Individuals tested with a 25-gene hereditary cancer panel from September 2013-present who were found to have an apparent germline *TP53* PV (n>100) were assessed and offered confirmatory testing via genetic testing of a family member, testing of fibroblast cells, or repeat testing on a blood sample from the proband. **Results:** At this time, 24 individuals with *TP53* PVs have participated in additional testing (13 fibroblast testing, 9 family testing, 2 repeat blood testing). These studies have provided additional evidence regarding the PV origin (germline or somatic) for 19 (79.2%) individuals thus far. Somatic *TP53* PVs were confirmed for 4 (16.7%) individuals. Germline transmission was confirmed in 2 (8.3%) cases where a relative was found to carry the same *TP53* PV. Evidence that was consistent with, though not conclusive for, germline origin was found in 13 (54.2%) individuals. Negative genetic test results in relatives were uninformative for the remaining 5 individuals (20.8%). **Discussion:** In this ongoing study, we have demonstrated that *TP53* PVs detected with NGS read frequencies consistent with an inherited PV can be either germline or somatic in origin. The confirmation of germline *TP53* PVs in patients with a clinical presentation inconsistent with LFS may have a significant impact on medical management decisions, while confirmation of a somatic PV may prevent inappropriate patient care. Overall, this demonstrates the value of confirmatory testing in individuals with apparent germline *TP53* PVs.

782T

Validation of a droplet PCR-based targeted gene panel for the detection of sequence variants in myeloid neoplasms. *A. Sahota[1], L. Liang[1], N. Sanapareddy[1], F. Patel[1], Q. Tan[1], W. Hussain[1], H. Gaspard[1], L. Goodell[2], S. Bhagavathi[2], G. Riedlinger[2], J. Rosenfeld[3], D. Foran[3], L. Rodriguez[3], R.M. Rattray[4], S.M. Gandhi[5], D. Qin[5], G.J. Tsongalis[6], S.A. Turner[6], F.B. de Abreu[6], H.B. Steinmetz[6], A. Brooks[1], J.A. Tischfield[1].* 1) Dept Genetics, Rutgers Univ, Piscataway, NJ; 2) Dept Path Lab Med, Rutgers Robert Wood Johnson Med Sch, New Brunswick, NJ; 3) Cancer Inst New Jersey, New Brunswick, NJ; 4) Molec Genomics Lab, Providence Health and Services, Portland, OR; 5) Molec Diagnostics Lab, Moffitt Cancer Ctr, Tampa, FL; 6) Dept Path Lab Med, Dartmouth Hitchcock Med Ctr, Lebanon, NH.

   Introduction: Next generation sequencing of cancer tissue is becoming a mainstream technique in clinical laboratories because of its potential to contribute to the selection of patient-specific therapies. Here, we describe the validation of the ThunderBolts (TB) Myeloid Panel from RainDance Technologies for the detection of sequence variants in DNA isolated from bone marrow or peripheral blood samples from patients with a variety of myeloid neoplasms. **Methods:** We received 72 previously tested DNA samples from three CLIA-certified and CAP-accredited laboratories (located in OR, FL, NH). All three labs used the TruSight Myeloid Panel and the MiSeq instrument from Illumina for variant analysis. We used 40 ng DNA for the TB panel and pooled PCR libraries from up to 12 samples and sequenced them on the MiSeq. DNA from three well-characterized human cell lines was also used to validate the TB panel. Sequence data were analyzed using the NextGENe software from SoftGenetics. **Results:** We observed an average of 4,886 reads per amplicon and over 98.3% of the amplicons had a read depth of ≥500X. The analytical sensitivity and specificity of the assay, tested using DNA from three cell lines, was 98.1%. The reproducibility of the assay, determined by analyzing two patient samples in three runs was 92.8%. We determined the lower limit of detection for sequence variants by mixing cell line DNAs in different proportions. We were able to detect sequence variants at the 2% level in these mixtures. Sequence data were available for 46 samples from the three external labs. These labs did not report any pathogenic variants or variants of unknown significance in 12 of these samples and variants in 5 samples could only be detected by the TruSight panel. The corresponding genes were not part of the TB panel. The external labs reported 50 sequence variants in the remaining 29 samples. The TB panel detected 48 of these variants, giving a concordance of 96%. Our software did not detect 2 variants, but one variant was in the region of a gene (CEBPA) where the coverage depth was low (281X). When we examined this region in more detail, we were able to detect this variant. This gave an overall concordance of 98%, suggesting excellent agreement between the TB and the TruSight panels. **Conclusions:** We have validated the TB myeloid panel for the detection of sequence variants in myeloid neoplasms. The assay has high accuracy and it can readily detect minor allele frequencies below the 5%.

783F

Multi-center assessment of a next-generation sequencing assay for detection of germline and somatic *BRCA1* and *BRCA2* gene variants from formalin fixed paraffin embedded samples. *C. Scafe[1], R. Qi[1], G. Nistala[3], G. Bee[3], N. Garg[5], M. Manivannan[5], A. Broomer[6], V. Mittal[7], P. Williams[7], D. Brinza[1], F. Hyland[1], J. Bishop[6], S. Sadis[7], B. Passkiewicz[4], J. Sherlock[2].* 1) Sequencing Applications, Thermo Fisher Scientific, South San Francisco, CA; 2) Life Sciences Solutions, Thermo Fisher Scientific, Paisley, Scotland, UK; 3) Clinical Sequencing Division, Thermo Fisher Scientific, Carlsbad, CA; 4) Life Sciences Solutions, Thermo Fisher Scientific, Pleasanton, CA; 5) Clinical Sequencing Division, Thermo Fisher Scientific, South San Francisco, CA; 6) Life Sciences Solutions, Thermo Fisher Scientific, Carlsbad, CA; 7) Life Sciences Solutions, Thermo Fisher Scientific, Ann Arbor, MI.

   We designed a new highly multiplexed amplification-based assay which allows detection of *BRCA1* and *BRCA2* gene variants (SNVs, in/dels and exon/gene deletions) in somatic and germline samples through semiconductor-based DNA sequencing. This study comprises the findings of over 20 research centers in 12 countries that were given early access to the new assay to evaluate its performance on a variety of sample types, including formalin fixed paraffin embedded (FFPE) tissues, under realistic conditions. *BRCA1* and *BRCA2* gene target regions, including all coding exons, were enriched via highly multiplexed amplification reactions. Libraries were constructed using manual and automated preparation methods, followed by semiconductor-based sequencing on multiple instrument platforms. Single nucleotide and insertion or deletion variants were identified using commercially available software solutions. Participants reported the concordance between the variants detected by the new assay and expected variants, identified by orthologous techniques. A novel software application for detection of exon-level and whole gene deletions in *BRCA1* and *BRCA2* was developed by analyzing known deletion samples. This coverage-based method implements an algorithm which first identifies large gene-level deletion candidates and then executes a fine-grained detection of exon-level variants. Automated and graphical data quality evaluation that was developed has been shown to be particularly useful for FFPE samples with mixtures of normal and tumor tissue. A high degree of coverage uniformity, 98%, was achieved on both germline and somatic samples. Sensitivity of detection of SNV and short insertion or deletion variants was 97.8%, with positive predictive value of 94.3%. Exon deletion detection on simulated variants in germline cell-line samples showed high call rates and sensitivity, with low rates of false positive calls. The single exon, multiple-exon and whole-gene deletions in the multi-center early access study data were also called with high sensitivity, high call rates, and minimal false positive call rates. Performance of the new *BRCA1* and *BRCA2* sequencing assay in this collaborative early access study indicates the new research use only panel, planned for official release in late 2016, will enable *BRCA1* and *BRCA2* gene researchers to interrogate a broader spectrum of variant types in both blood and FFPE samples.

784W

Chromosomal abnormalities in meningiomas. *B. Ouled Amar Bencheikh, C. Mirarchi, P.A Dion, G.A Rouleau.* Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada.

Background: Meningioma are deemed to be benign tumors of the central nervous system. However, a subset of these tumours are actually malignant and present an aggressive behavior that is refractory to treatments combining surgery and radiotherapy. Loss of a chromosome 22q region (containing the *NF2* gene) is the most frequently observed alteration for these tumors (>70% of meningioma); additional chromosomal abnormalities were also reported (e.g. 1p, 1p, 6q, 7, 9q, 9q, 10q, 14q, 15q 17q, 18q, 19q, 20q). We are now using data from a high density SNP chip microarray to further complete a profile of chromosomal abnormalities in sporadic meningioma and hopefully identify novel genetic drivers that will be specific to these particular tumors.   **METHODS:** A cytogenetics microarray (Cytoscan® HD, Affymetrix, Santa Clara, CA, USA) was used to examine 112 sporadic meningioma samples obtained from unrelated individuals. Previous studies conducted based on microsatellite markers excluded chromosome 22q abnormalities across these 112 samples. **RESULTS:** Chromosomal abnormalities were detected in 73 meningioma (65%). The remaining tumors appeared to be normal using the Cytoscan HD array. The nature of the abnormalities observed included loss of heterozygosity and/or complete deletion or duplication. Interestingly, we identified a loss of chromosome 22q in 45 (40%) meningioma with abnormalities. This last observation warrants for the increased sensitivity of the array used here as 22q abnormalities went unnoticed from the earlier microsatellites genotyping efforts conducted in previous years. Additional chromosomal alterations were identified with or without loss of 22q, these alterations were both new and known abnormalities (LOH, deletion and duplication) in all autosomal and X chromosomes.   **CONCLUSION:** Our findings confirm and extend previous observations in regard to the high frequency of 22q loss in meningioma. Our observation of previously unreported abnormalities across chromosomes also support the notion that there is a great extent of genetic heterogeneity across these tumours (with or without 22q loss). A better understanding of meningioma subtypes may help to better define the more aggressive tumors and in conjunction with additional clinical information help to deploy more personalized treatments for those affected.

785T

Significant effect of polymorphisms in *CYP2D6* **on response to Tamoxifen therapy for breast cancer: A prospective multicenter study.** *H. Zembutsu[1], S. Nakamura[2], S. Akashi-Tanaka[2], T. Kuwayama[2], C. Watanabe[2], T. Takamaru[2], H. Takei[3], T. Ishikawa[4], K. Miyahara[4], H. Matsumoto[5], Y. Hasegawa[6], G. Kutomi[7], H. Shima[7], F. Satomi[7], M. Okazaki[8], H. Zaha[9], M. Onomura[9], A. Matsukata[10], Y. Sagara[10], S. Baba[10], A. Yamada[11], K. Shimada[11], D. Shimizu[12], K. Tsugawa[13], A. Shimo[13], T. Yu[14], M. Hartman[15], C. Wan[15], S. Lee[16], Y. Nakamura[17].* 1) Research Institute, National Cancer Center, Tokyo, Tokyo, Japan; 2) Showa University, Tokyo, Japan; 3) Nippon Medical School, Tokyo , Japan; 4) Tokyo Medical University, Tokyo , Japan; 5) Saitama Cancer Center, Saitama , Japan; 6) Hirosaki Municipal Hospital, , Japan; 7) Sapporo Medical University, Sapporo , Japan; 8) Sapporo Breast Surgical Clinic, Sapporo , Japan; 9) Nakagami Hospital, Okinawa , Japan; 10) Sagara Hospital, Kagoshima , Japan; 11) Yokohama City University Medical Center, Yokohama , Japan; 12) Yokohama Minato Red Cross Hospital, Yokohama , Japan; 13) St. Marianna University School of Medicine Hospital, Kawasaki , Japan; 14) Tan Tock Seng Hospital, Singapore; 15) Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore; 16) National University Cancer Institute, National University Health System, Singapore; 17) Department of Medicine, The University of Chicago, Chicago, USA.

Purpose CYP2D6 is the key enzyme responsible for the generation of the potent active metabolite of tamoxifen, "endoxifen". We previously reported that reduced- or null-function alleles of *CYP2D6* were significantly associated with poor clinical outcome of breast cancer patients treated with tamoxifen. However, there are still controversial reports questioning the association between *CYP2D6* genotypes and tamoxifen efficacy. Hence, we carried out a prospective multicenter study to evaluate the clinical effect of *CYP2D6* genotype on tamoxifen therapy. **Patients and Methods** We enrolled patients with hormone receptor–positive and human epidermal growth factor receptor 2-negative, invasive breast cancer receiving pre-operative tamoxifen monotherapy for 14 - 28 days. Ki-67 response in breast cancer tissues after tamoxifen therapy was used as a surrogate marker for response to tamoxifen. We prospectively investigated the effects of allelic variants of *CYP2D6* on Ki-67 response, pathological response and hot flushes. **Results** Ki-67 labeling index in breast cancer tissues significantly decreased after preoperative tamoxifen monotherapy (P = .0000000000000013). Moreover, proportion and Allred scores of estrogen receptor positive cells in breast cancer tissues were significantly associated with Ki-67 response. Although *CYP2D6* variants were not associated with pathological response nor hot flushes, they showed significant association with Ki-67 response after pre-operative tamoxifen therapy. **Conclusion** This is the first prospective study evaluating the relationship between *CYP2D6* variants and Ki-67 response after tamoxifen therapy. Our results suggest that genetic variation in *CYP2D6* is a key predictor for the response to tamoxifen in patients with breast cancer.

786F

Development of a cloud-based comprehensive cancer transcriptome profiling pipeline for clinical diagnostics. *H.A. Costa, C. Kunder, J.L. Zehnder, C.D. Bustamante.* Stanford University School of Medicine, Stanford, CA, USA.

The application of DNA sequencing in clinical cancer genomics has proven to be an invaluable tool to characterize acquired tumor mutational load, inform therapeutic routes and to predict treatment response in oncology patients. Despite these powerful benefits, this approach lacks the ability to quantitatively characterize downstream gene expression levels or capture non-genomic transcript variation—two important avenues currently under characterized in next generation sequencing based clinical diagnostics. RNA sequencing (RNAseq) addresses these issues by quantitatively capturing the entirely of the expressed transcripts in a tissue, while additionally detecting all of the expressed genomic alterations as seen in DNA sequencing. One of the barriers to its implementation in a clinical setting is the availability of computational pipelines and bioinformatics tools to comprehensively characterize the broad range of variation present in transcriptome data. To this end, we have developed a cloud-based RNAseq computational workflow to detect expressed variation including SNVs, INDELs, structural variants, gene fusions, differential gene expression, allele specific expression, and novel alternative allele isoforms in a single pipeline. We apply our method to characterize the transcriptomes of pediatric sarcoma and lung adenocarcinoma to identify novel variation. Further, we demonstrate that our approach compares favorably to DNA based sequencing methods when performing variant calling on somatic patient samples.

787W

Genetic testing of 248 Chinese aortopathy patients using a panel assay. *Y. Ma[1], H. Yang[1], M. Luo[2], Y. Fu[1], Y. Cao[3], K. Yin[1], C. Meng[1], K. Zhao[1], J. Zhang[2], Y. Fan[4], C. Shu[2], Q. Chang[2], Z. Zhou[1].* 1) State Key Laboratory of Cardiovascular Disease, Beijing Key Laboratory for Molecular Diagnostics of Cardiovascular Diseases, Diagnostic Laboratory Service, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences an; 2) State Key Laboratory of Cardiovascular Disease, Center of Vascular Surgery, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 3) Analyses Technologies, Beijing, China; 4) John Welsh Cardiovascular Diagnostic Laboratory, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA.

Inherited aortopathy, which is characterized by a high risk of fatal aortic aneurysms/dissections, can occur secondarily to several syndromes. To identify genetic mutations and help make a precise diagnosis, we designed a gene panel containing 15 genes responsible for inherited aortopathy and tested 248 probands with aortic disease or Marfan syndrome. The results showed that 92 individuals (37.1%) tested positive for a (likely) pathogenic mutation, 70 (28.2%) had a variant of uncertain significance (VUS). Most of the (likely) pathogenic mutations were located in the *FBN1* gene. Also, We found that patients with a *FBN1* truncating or splicing mutation were more prone to have a severe aortic disease or valvular disease. Among patients with a *FBN1* truncating or splicing mutation, 15 suffered a life-threatening aortic dissection, 5 had a severe valvular disease, while 9 had an aortic aneurysm and therefore were performed a prophylactic surgery. Additionally, in the aneurysm group, patients with a *FBN1* truncating or splicing mutation took a prophylactic surgery at a younger age (25.6y vs. 33.4y) than those with a missense mutation. To date, this is the largest reported cohort of Chinese patients with aortic disease who have undergone genetic testing. As such, it can serve as a considerable dataset of next generation sequencing data analysis of the Chinese population with inherited aortopathy. Moreover, in our study, we developed an optimized pipeline named iAorta that can realize automatic analysis. Additionally, according to the accumulated data, we optimized the analysis pipeline by adding quality control steps to assess the sequencing quality and to indicate possible false-negative variants, remove frequent false-positive mutations based our existing data and drop the low confidence indel variants to reduce the false-positive rate. In summary, our data further expands the *FBN1* mutation spectrum and offer evidence for the genotype-phenotype correlation between *FBN1* mutation type and aortic events. The aortopathy panel assay undoubtedly presents a highly valuable clinical tool and lays the foundation for further study. We are dedicated to constructing the largest Chinese aortopathy genetic database and continually improving our testing quality.

788T

Uncovering copy number variants (CNVs) in degraded post-mortem samples of carries of congenital heart disease (CHD). *F.A.R. Madia[1], A.T. Dias[1], M. Rocha[2], E.A. Zanardo[1], T.V.M.M. Costa[1], J.G. Damasceno[1], C. Milani[1,3], A.F. Amorim[1], A.M. Nascimento[1,3], G.M. Novo-Filho[1,3], M.M. Montenegro[1,3], Y.G. Oliveira[1,3], F.T. Gonçalves[4], C. Fridman[4], C.A. Kim[3], R. Schultz[5], L.D. Kulikowski[1,2,3].* 1) Laboratório de Citogenômica, LIM 03, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 2) Centro de Reprodução e Genética, Departamento de Saúde Coletiva, Faculdade de Medicina do ABC, Santo André-SP, Brazil; 3) Unidade de Genética do Instituto da Criança, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 4) Departamento de Medicina Legal, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 5) Serviço de Anatomia Patológica , Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil.

Congenital heart disease (CHD) is the most prevalent group of congenital malformations affecting approximately 8-1,000 live births and is a significant cause of childhood morbidity and mortality. Recently, the literature demonstrated the increased frequency of pathogenic copy number variants (CNVs) in congenital heart disease. However, there are few post-mortem genomic studies in congenital heart disease cases. Several technical issues can make post-mortem genetic testing challenging. Thus, the aim of the study was identify the CNVs in post-mortem degraded samples of CHD. We investigate samples from skin and heart tissues of 15 cases of stillbirth and new-born from Serviço de Verificação de Óbitos, HC-FMUSP. The molecular analysis was performed by genotyping of seven autosomal STR loci (D13S317, D7S820, D2S1338, D21S11, D16S539, D18S51, CSF1PO and FGA), and the results allowed us to disclose different CNVs. Thus, the results showed pathogenic CNVs associated with syndromes in 20 % (3/15) of stillbirth and new-born cases, including Edwards Syndrome, Down Syndrome and one case of Turner Syndrome with Y chromosome mosaicism. The genomic analysis showed efficiency for identifying pathogenic CNVs in post-mortem degraded samples of CHD and also help to provide the adequate genetic counseling to families. Financial support: FAPESP: 09/53105-9 and FINEP-CT INFRA 0160/12 SP.

789F

Next-generation sequencing diagnostics of inherited arrhythmogenic cardiac disorders. *C. Marschall, K. Wittkowski, H.G. Klein.* Center for Human Genetics, 82152 Martinsried, Germany.

A custom NGS 72 gene panel approach was performed on a collective of 100 unrelated patients referred for genetic testing for hereditary arrhythmogenic cardiac disorders such as long QT syndrome (LQTS), catecholaminergic polymorphic ventricular tachycardia (CPVT), Brugada syndrome (BrS), hypertrophic cardiomyopathy (HCM), dilatative cardiomyopathy (DCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC). All these diseases are associated with significant morbidity and mortality and are a known risk factor for sudden cardiac death. The presymptomatic diagnosis is essential as for most of the cases a therapy is available. Mutation detection in the index cases followed by cascade screening of the family members could be life saving. As some cardiomyopathies may present with only little structural symptoms, differential diagnosis of arrhythmogenic disorders may be difficult. Moreover, some arrhythmogenic disorders are clinically and genetically extremely heterogeneous. An Agilent SureSelect Target Enrichment assay was designed for the capture of coding regions including splice sites of 72 genes known to be associated with arrhythmogenic disorders. Target enrichment followed by re-sequencing on the Illumina NextSeq platform was used for mutation detection. Data analysis was performed with the CLC Genomics Workbench and custom developed Perl scripts. Potentially disease causing variants and regions with an insufficient coverage were re-analyzed with Sanger sequencing. Pathogenic or potentially pathogenic variants were detected 20-50% of cases. Diagnostic sensitivity was compared with results from Sanger sequencing of more than 5,000 patients. The diagnostic yield was higher with NGS in CPVT, BrS, ARVC and DCM. Additional 5-50% of cases showed variants of unclassified significance.

790W

Diagnostic yield of three years next generation sequencing in clinical genetic diagnostics of cardiomyopathies. *R.C. Niessen, P.A. van der Zwaag, L.G. Boven, Y.M. Hoedemaekers, R.J. Sinke, J.D.H. Jongbloed.* Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, Netherlands.

The strength of next generation sequencing (NGS) in both research and diagnostics is becoming increasingly evident. NGS is successfully applied to find causal mutations and thus confirm the clinical diagnosis in patients suspected of genetic cardiomyopathies. In 2012 NGS analysis was introduced in the diagnostic clinical genetic laboratory in the University Medical Center Groningen. Since then, we analyzed DNA of 1153 patients with several types of cardiomyopathies using an enrichment kit targeting 55 genes associated with hereditary cardiomyopathies. Two classification trees were used to classify variants, taking in account population frequencies, in silico predictions, type and location of mutation, conservation scores, clinical data and information from the literature. Additional haplotype and cosegregation analyses were performed to further support pathogenicity of novel potentially causal mutations that had been identified multiple times. Our approach results in a significant increase in diagnostic yield: in 14% (157/1153), of patients a pathogenic mutation was identified, explaining the clinical phenotype. In 10% (114/1153), one or more likely pathogenic mutations were identified. In 593 patients (51%) 885 variants of unknown clinical significance were found, including the patients who also carried a (likely) pathogenic mutation. Taken together, in 23% of patients the clinical phenotype was (likely) explained by either a pathogenic or a likely pathogenic mutation. Analysis of the different cardiomyopathy subtypes is ongoing and these results will be presented. The contribution of the variants of unknown significance remains to be elucidated. In the near future, we hope to further characterize these variants by sharing our data with other laboratories in the Netherlands and abroad, by performing segregation studies and functional studies. In conclusion, the results from this large cohort show that NGS provides a genetic diagnosis in a large proportion of cardiomyopathy cases and that datasharing is essential to further classify the many variants with unknown significance.

791T

Glycine missense variants in the *COL3A1* **triple helix domain: How to assess functional domain data during clinical variant interpretation.** *D. Beltran, J. Thusberg, M. Anderson, Y. Kobayashi, J. Garcia, T. Winder, S. Topper, K. Nykamp.* Invitae, San Francisco, CA, U.S.A.

Clinical variant classification should incorporate gene-specific knowledge about critical residues or essential protein domains. Missense variants that affect these residues or regions are more likely to be deleterious and pathogenic than variants that affect other residues or domains. However, clear guidelines for identifying a residue or region as essential do not exist. Without such guidelines, variant classification becomes subjective and capricious. To address this shortcoming, we developed a statistical approach for the identification of critical residues that takes into account background variation, the distribution of rare, clinically observed variants, and knowledge of the three-dimensional structure of proteins. We selected *COL3A1* as the paradigm for protein domain evaluation. *COL3A1* is a well-characterized gene associated with vascular Ehlers-Danlos syndrome (vEDS). Most of the protein is made of a triple-helix (TH) domain encoded by 343 Gly-X-Y (GXY) repetitions. Gly residues therein are essential for protein structure, and missense variants affecting TH Gly residues are over-represented in vEDS patients (97%) but under-represented in ExAC (5%). ExAC contains five TH G*XY missense variants, each of which is present three or fewer times. As a general rule, structural residues in TH domains can be considered essential, as there is a substantial statistical difference between the variation seen in patients and that in a control data set such as ExAC. Applying this analysis to other collagen genes, we can easily distinguish between those for which this logic is warranted (*COL3A1*) and those such as *COL5A1* and *COL5A2* for which it is not because the number of TH G*XY missense variants described in patients is low or the frequency of some of the missense variants at these residues in ExAC exceeds that expected for pathogenic variants. Detailed analysis also suggests a correlation between the exactness of the GXY repetition in TH and the severity of the associated phenotype. A similar approach is applied to cysteine missense variants in EGF-like domains in genes such as *FBN1*, *FBN2*, and *LDLR*.

792F

Use of exome sequencing in cases of sudden cardiac death. *E. Smith[1,2], S.B. Seidelmann[1,2,3], A. Mani[1,2,3,4].* 1) Yale University School of Medicine, New Haven, CT; 2) Yale Program for Cardiovascular Genetics, Yale School of Medicine, New Haven, CT; 3) Division of Cardiovascular Medicine and Department of Radiology, Division of Cardiac Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Department of Genetics, Yale School of Medicine, New Haven, CT.

Sudden cardiac death (SCD) caused by ventricular arrhythmia in subjects with or without normal structural heart disease can be caused by single gene mutations. This condition incorporates a spectrum of inherited disorders from Long QT (MIM 192500) and Brugada syndromes (MIM 601144), to hypertrophic (MIM 192600) and dilated cardiomyopathies (MIM 115200). Clinical information obtained in these cases, including automatic external defibrillator rhythm recordings, post-resuscitation electrocardiograms or echocardiograms, and cardiac MRI often does not provide a definitive clinical diagnosis. In such circumstances, genetic testing may be greatly useful for diagnosis, risk stratification and management of the patients and their close relatives. With the advent of exome sequencing, one may be able to obtain a genetic diagnosis in cases in which the clinical data has remained inconclusive. Twenty-eight adult patients with a presenting diagnosis of SCD underwent proband-only exome sequencing as a first line test. These individuals are part of larger cohort of adults with inherited cardiovascular diseases who underwent exome sequencing at our center. All genetic diagnostic reports were made based on American College of Medical Genetics, as well as College of American Pathology and CLIA guidelines. When possible, segregation analysis was carried out to further elucidate the potential underlying genetic causes. Variants were called pathogenic or likely pathogenic for 9 patients (32%). Another 9 patients had variants of uncertain significance in related genes. There were seven subjects in whom more than one variant was reported. In this group of 28, a total of 64% of SCD cases received either definitive or presumptive genetic diagnosis. Identified variants were all in genes involved either in excitation or contraction of cardiomyocytes or both. Most common variants identified were radical *TTN* mutations. In conclusion, exome sequencing was effective in determining the potential genetic etiology of SCD in two-third of cases where clinical presentation was inconclusive.

793W

Prenatal genetic testing following abnormal fetal ultrasound reveals 6% of cases with pathogenic variants implicated in Noonan spectrum disorders. *N. Cody, G. Cai, J. Liao, G. Mendiratta-Vij, S. Sperber, L. Edelmann, R. Kornreich.* Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York City, NY.

Noonan spectrum disorders (NSD) are an autosomal dominant, multi-systemic group of disorders caused by mutations in the RAS-MAPK pathway and have an estimated prevalence of 1 in 1000 to 1 in 2500 live births. Abnormal ultrasound findings that may be indicative of NSD in fetuses include increased nuchal translucency or nuchal fold, cystic hygroma, hydrops, pleural effusion and cardiac or renal anomalies. We conducted a retrospective study of 195 prenatal cases with indication for NSD testing. Molecular tests were performed on either direct or cultured chorionic villi or amniocytes. We reviewed the indication for testing along with the number of pathogenic variants detected using a next generation sequencing (NGS) panel. The panel currently tests 14 well-established genes implicated in NSD: *PTPN11, SOS1, RAF1, KRAS, NRAS, BRAF, MAP2K1, MAP2K2, CBL, HRAS, NF1, SHOC2, RIT1,* and *SPRED1.* Increased nuchal translucency or nuchal fold (126 of 195), followed by cystic hygroma (57 of 195), were the most common fetal ultrasound findings leading to NSD testing. In addition, seven fetuses had cardiac defects on ultrasound, two had hydrops fetalis and four had a previous family history of Noonan syndrome. Following genetic testing, 13 of 195 patients (6%) were determined to carry a pathogenic or likely pathogenic variant. The most frequently mutated genes were *PTPN11* (5/13) and *RIT1* (4/13). *KRAS, BRAF, SOS1,* and *RAF1* were each found to be causative in one fetus. Of the seven cases in which parental testing could be performed, 4 carried de novo variants, two were maternally inherited and one was paternally inherited. While increased nuchal translucency was the most common indication for testing, it had a very low positive predictive value with 1 of 126 (0.8%) fetuses testing positive for NSD. The presence of a cystic hygroma on ultrasound resulted in a positive rate of 16% (9/57) and is perhaps the most predictive indicator for prenatal testing for NSD. These findings are consistent with other studies and support the use of a multi-gene panel for genetic testing for NSD.

794T

Exploiting whole genome sequencing for finding new genetic causes of familial dilated cardiomyopathy. *A.E. Minoche[1], C. Horvat[2], M.E. Dinger[1], D. Fatkin[2], M.J. Cowley[1].* 1) Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst NSW 2010 Australia; 2) Victor Chang Cardiac Research Institute, 405 Liverpool St, Darlinghurst NSW 2010 Australia.

   The genetic cause of familial dilated cardiomyopathy (DCM) is often unknown. The genetically heterogeneous nature of the disease and ongoing research requires testing for genetic variation in constantly growing lists of genes. Whereas gene panel sequencing struggles to keep up with this challenge, whole genome sequencing overcomes it and has the potential to advance the discovery of new variant disease associations.   We show that WGS not only covers more genes, it even covers disease genes captured by targeted sequencing more comprehensively. For a set of 67 DCM genes targeted by panel sequencing, WGS covered 7.8% to 21.1% more gene positions >=15x (depending on the selected isoforms). On median WGS covered 99% of exon bases with this read depth.   All causative SNVs detected in 21 of 42 patients using panel sequencing were also found in WGS. Beyond that WGS enabled the detection of 10 additional predicted high and medium impact variants per patient within the same set of target genes, structural variation including copy number variation and variants in regulatory regions. Further through WGS we identified damaging variants in additional genes not targeted by the panel, including 15 recently identified DCM genes, 274 genes specifically expressed or over expressed in heart and 47 genes recommended for reporting incidental findings (ACMG genes).   Of these additional variants one was found causative (Nkx2.5). In the remaining patients, we have identified additional 12 candidate variants that are rare in healthy individuals, and predicted to be damaging, which are currently tested for co-segregation. Seven of these are in known DCM genes and missed by targeted sequencing; two are in newly discovered DCM genes (GATA4, DSC2), two in genes expressed in the heart (TRIM63 and PDE4DIP), and two caused by structural variants (DSC2 and BAG3). In only seven patients we didn't find any new candidates. With WGS the genetic cause was found for 52% of the patient cohort, which is already slightly higher than with panel sequencing. Overall it is expected that the yield for WGS will further increase, thanks to candidate variants in 33% of undiagnosed patients.

795F

Targeted sequencing for genetic diagnosis of cardiovascular disorders. *S. Mikaeeli[1], N. Mahdieh[2], B. Rabbani[2], S. Mikaeeli[3].* 1) Islamic Azad University of Medical Branches, Tehran, Iran; 2) Genetic Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran; 3) Islamic Azad University of North Tehran Branches, Tehran, Iran.

   During last few years next generation sequencing (NGS) technique has provided the vast majority of applications both in basic and clinical research. This technique is applied, in order to identify rare variants in genetic disorders more especially in heterogeneous diseases such as cardiomyopathies. Hence, targeted sequencing technique could be affordable and rapid technique to overwhelm diagnostic problems. Here, we present families whom their children suffer from Marfan Syndrome (MFS [MIM 154700]) and Dilated cardiomyopathy (DCM[MIM 115200]), candidate genes were selected using targeted sequencing technique for genetic testing followed by Sanger sequencing test for confirming recognized variants. In one family, 24 different variants were obtained out of 14 candidate genes which Sanger test has revealed that only one mutation at c.7783G>A (Gly259Ser) position which is located at *FBN* gene has pathogenic effect. In another family 65 variants out of 23 applicant genes were discovered and one pathogenic mutation identified at c.3373G>A (Val1125Met) position in *MYBPC3* gene. Of course in silico analysis predicted the pathogenicity of the mutations. This investigation could demonstrate the pivotal role of targeted sequencing in diagnostic genetic testing.

796W

Next generation sequencing (NGS) evaluation of *CYP11B1* and *CYP11B2* single nucleotide variants (SNP) for a correlation in hypertension and Cushing Syndrome. *W. Shi, C. Chin, A. Beams, J. Lameh, R. Pollner.* Bio-Pharma, Genoptix Medical Laboratory, A Novartis company, Carlsbad, CA.

   Introduction: Members of the CYP11B subfamily participate in the biosynthesis of important steroid hormones. Among them, CYP11B1 and CYP11B2 control the synthesis of cortisol and aldosterone, respectively. Cortisol is a steroidal hormone essential for homeostasis by regulating changes in the body in response to stress. Prolonged exposure to excess cortisol leads to Cushing syndrome. Aldosterone has been implicated in the pathogenesis of hypertension, heart failure, and renal disease. A new potential method to treat Cushing syndrome or hypertension is controlling plasma cortisol and aldosterone levels by CYP11B1 and CYP11B2 inhibition. We developed a NGS amplicon-based assay for sequencing *CYP11B2* and *CYP11B1* gene regions, and are evaluating the correlation of SNPs with clinical presentations among different populations. **Methods**: We designed custom primers to cover all *CYP11B1* introns and exons as well as 500 bp upstream and 500 bp downstream. Further upstream *CYP11B1* promoter was excluded due to the presence of repetitive sequences. Primers for *CYP11B2* were designed to cover all exons and introns, a 2 kb promoter region, and 500 bp downstream of the strop codon. A total of 78 amplicons, ranging in sizes from 125 to 375 bp, were generated in three pools to enable maximum coverage of both genes and to minimize gaps. DNA from whole blood was isolated. Library preparations using Hi-Q Library reagents followed the Ion AmpliSeq 2.0™ procedure. Subsequent sequencing was performed using Thermo Fisher's Hi-Q sequencing reagents with 318 v2 chips on the Ion Chef™ and Personal Genome Machine™ (PGM) was all performed. Germline SNPs and indels were identified using Thermo Fisher's Ion Reporter™ Software version. **Results**: The NGS amplicon assay developed for evaluation of *CYP11B1* and *CYP11B2* SNPs based on the Thermo Fisher PGM™ platform shows sufficient coverage among all 78 custom-designed amplicons. We found that the assay performance was accurate by evaluating the Illumina Platinum Genome, NA12878. Multiple-sequenced NA12878 showed high concordance with each other and to the reference Illumina Platinum Genome for the regions of interest. Clinically relevant SNPs were identified in a subset of subjects and tested for an association with hypertension or Cushing syndrome. **Conclusions**: Reliable SNPs reported for amplicons covering both *CYP11B1* and *CYP11B2* genes can potentially provide valuable information for developing novel therapeutics and guiding treatment decisions.

797T

***CTNS* molecular genetics profile in a Persian nephropathic cystinosis population.** *B. Behnam[1,2,3], F. Ghazi[1], R. Hosseini[2], H. Otukesh[2], M. Ackouchakian[1], S. Teimourian[1], Z. Ataee[1], F. Vali[1], W. Gahl[3].* 1) National Human Genome Research Institute, National Institute of Health, Bethesda, MD, USA; 2) Department of Medical Genetics and Molecular Biology, Faculty of Medicine, Iran University of Medical Sciences (IUMS), Tehran, Iran; 3) Department of Pediatrcics, Faculty of Medicine, Ali Asghar Children Hospital, Iran University of Medical Sciences (IUMS), Tehran, Iran.

   Purpose: In this report, we document the *CTNS* gene mutations of 28 Iranian patients with nephropathic cystinosis age 1-17 years. All presented initially with severe failure to thrive, polyuria, and polydipsia. **Methods:** Cystinosis was primarily diagnosed by a pediatric nephrologist and then referred to the Iran University of Medical Sciences genetics clinic for consultation and molecular analysis, which involved polymerase chain reaction (PCR) amplification to determine the presence or absence of the 57-kb founder deletion in *CTNS*, followed by direct sequencing of the coding exons of *CTNS*. **Results:** The common 57-kb deletion was not observed in any of the 28 Iranian patients. In fourteen of 28 patients (50%), mutations were observed in exons 6 and 7. No mutation was detected in exon 5, and only one (3.6%) patient with cystinosis showed a previously reported 4-bp deletion in exon 3 of *CTNS*. Four patients (14.3%) had a previously reported mutation (c.969C>A; p.N323K) in exon 11, and five (18%) had novel homozygous deletions in exon 6 leading to premature truncation of the protein. These deletions included c.323delA; p.Q108RfsX10) in three individuals and c.257-258delCT; p.S86FfsX37 in two cases. Other frame-shift mutations were all novel homozygous single base pair deletion/ insertions including one in *CTNS* exon 9 (c.661insT; p.V221CfsX6), and four (14.3%) in exon 4, i.e., c.92insG; p.V31GfsX28 in two and c.120delC; p.T40TfsX10 in two. In total, we identified eight previously reported mutations and eight novel mutations in our patients. The only detected splice site mutation (IVS3-2A>C) was associated with the insertion mutation in the exon 9. **Conclusion:** This study, the first molecular genetic analysis of non-ethnic-specific Iranian nephropathic cystinosis patients, may provide guidance for molecular diagnostics of cystinosis in Iran.

798F

Mutation spectrum of *ABCD1* in 17 Vietnamese patients with X-linked adrenoleukodystrophy. *K.N. Nguyen[1], D.C. Vu[1], H.T. Nguyen[1], T.P. Bui[1], N.B.T. Can[1], H.T. Le[1], N. Shimozawa[2].* 1) Medical Genetics and Metabolism, Vietnam Children's Hospital, Hanoi, Viet Nam; 2) Division of Genomics Research, Life Science Research Center, Gifu University, Department of Pediatrics, Gifu University School of Medicine, Gifu, Japan.

   Background: X-linked adrenoleukodystrophy (X-ALD) is caused by a defect in the gene *ABCD1*, which maps to Xq28 and codes for a peroxisomal membrane protein that is a member of the ATP-binding cassette transporter superfamily. This disease characterized by progressive neurologic dysfunction, occasionally associated with adrenal insufficiency. **Objective:** to identify mutations of gene *ABCD1* in Vietnamese patients with X-ALD. **Method:** Genomic DNA from 17 Vietnamese patients from 15 unrelated families was extracted using standard procedures from the peripheral blood leukocytes. Mutation analysis of *ABCD1* was performed using Polymerase chain reaction (PCR) and DNA direct sequencing. **Results:** We identified 14 different mutations of *ABCD1* in 17 patients including missense mutations (9/14), deletion (4/14) and splice site mutation (1/14). Of which, six novel mutations including c.1202G>T (p.Arg401Trp); c.1208T>A (p.Met403Lys); IVS8+28-551bp del; c.1668G>C (p.Q556H); c.292_296delTCGGC (p.S98RfsX95); and the extent of deletion included between IVS1+505 and IVS2+1501, containing whole the exon 2 (4243bp), plus insertion of 79bp from BAP31 and 8bp from unknown origin in this deleted region were identified in six unrelated patients. Eight reported mutations including c.796G>A (p.Gly266Arg); c.1628C>T (p.Pro543Leu); c.1553G>A (p.Arg518Gln); c.1552 C>T (p.Arg518Trp); c.854G>C (p.R285P); c.1825G>A (p.E609K); c.1415_1416delAG (p.Q472RfsX83) and c.46-53del insG were identified in 10 patients from 8 families. **Conclusion:** Mutation analysis of *ABCD1* helped confirmation of diagnosis of X-ALD, genetic counselling and prenatal diagnosis.

799W

The Personalized Diabetes Medicine Program identifies multiple individuals with treatment-altering diagnoses of monogenic diabetes. *J.W. Kleinberger[1], K.A. Maloney[1], L.J.B. Jeng[1], E.A. Streeten[1], S.A. Stein[2], Y. Mohtasebi[1], Y. Guan[1], C.L. Overby[3], T.J. Mathias[1], N. Ambulos[1], D. Sewell[1], K. Tanner[1], C.M. Damcott[1], M.J. Nicholson[1], K. Palmer[1], T.D. O'Connor[1], R.B. Horenstein[1], K.D. Silver[1], A.R. Shuldiner[1], T.I. Pollin[1].* 1) University of Maryland School of Medicine, Baltimore, MD; 2) DC Endocrine, Washington, DC; 3) The Johns Hopkins University School of Medicine, Baltimore, MD.

   Monogenic diabetes accounts for approximately 1-2% of all diabetes mellitus and is often misdiagnosed due to factors such as lack of awareness, clinical similarity to common forms of diabetes, limited availability/high cost of genetic testing and complexity of result interpretation. A diagnosis of certain forms of monogenic diabetes can lead to replacement of insulin with oral medications or even no treatment at all while maintaining or improving glycemic control and allowing diagnosis of family members. The Personalized Diabetes Medicine Program (PDMP) is a project in the NIH supported Implementing Genomics in Practice (IGNITE) Network to implement and evaluate a sustainable approach to detection, diagnosis and individualized treatment for patients with monogenic diabetes. Patients being treated for diabetes in endocrinology clinics in an urban academic medical center, a Veterans Affairs medical center, an integrated healthcare system and a suburban private practice are screened using a simple questionnaire and review of standard bloodwork and/or referral by clinicians noting unusual phenotypes. Those suspected to have monogenic diabetes are enrolled, and the patient's medical and family history is collected upon consent. Genetic testing, when indicated, is performed using a next generation sequencing custom gene panel covering coding and flanking splice regions of 40 monogenic diabetes genes. Rare protein-altering variants are manually reviewed and classified according to ACMG/AMP guidelines for interpretation of sequence variants. Pathogenic or likely pathogenic variants are confirmed in a CLIA/CAP accredited laboratory and results are communicated to the patient by a genetic counselor and endocrinologist/clinical geneticist, placed in the electronic health record, and routed to the patient's diabetes care provider. To date, 4 patients with pathogenic variants and 2 with likely pathogenic variants were found in 65 individuals tested. These variants in *HNF1A*, *GCK*, and *INS* genes generated specific treatment recommendations. Our discovery rate of 9% indicates that our screening approach is enriching the test sample for mutation-positive individuals. Additionally, 54 variants of uncertain significance (VUS) were discovered in 34 patients, including multiple suspicious variants in the most common monogenic diabetes genes. We are currently evaluating the feasibility of incorporating *in vitro* and animal model studies to aid the variant interpretation process.

800T

The first Iranian case of mucopolysaccharidosis IIIC: Use of homozygosity mapping in a consanguineous pedigree. *N. Hosseini[1], A. Eslahi[1], F. Ashrafzadeh[2], K. Hasanpour[3], M. Mojarrad[1,4].* 1) Department of Medical Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, IRAN; 2) 2Department of Pediatric Neurology, Ghaem Medical Center, Mashhad University of Medical Sciences, Mashhad, Iran; 3) Departmentof Clinical Sciences, Sabzevar University of Medical Sciences, Sabzevar, Iran; 4) Medical Genetics Research Center, School of Medicine, Mashhad University of Medical.

Mucopolysaccharidosis type IIIC (MPSIIIC) is a rare subtype of mucopolysaccharidosis disorder family caused by mutations in *HGSNAT* gene. Mucopolysaccharidosis type IIIC consist of four subtype which have overlapping features and are indistinguishable in clinical level. In population with high consanguineous marriage rate like Iran homozygosity mapping can be a good choice to finding of disease locus.Herein we report a female patient with a novel mutation in *HGSNAT* gene. Homozygosity mapping was performed using SNP-array technology to detect candidate locus in this patient. All coding exons of *HGSNAT* gene were scanned by direct DNA sequencing. We found a novel ins/del mutation as c.1357TA>C in *HGSNAT* gene. This mutation is a frameshift mutation which eventually leads to premature protein truncation. To the Best of our knowledge this is the first case report of sanfilippo type C in Iranian population. This result supports applicability of homozygosity mapping to diagnosis of sanfilippo subtypes.

801F

QuantStudio carrier screening panel for Ashkenazi Jewish diseases. *M. Procter[1], R. Mao[1,2].* 1) ARUP Institute for Clinical & Experimental Pathology, Salt Lake City, UT; 2) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT.

Background In an effort to expand our current Ashkenazi Jewish (AJ) mutation panel, we investigated the QuantStudio platform, capable of targeting up to 64 variants simultaneously. The QuantStudio platform is able to handle varying assay complexities, from small multiplex panels, to the Ashkenazi Jewish mutation panel of 53 targets, to larger panels in varying formats. We evaluated the feasibility of using a multi-sample array wherein loci are amplified and analyzed independent of each other, eliminating PCR competition. **Materials and Methods** Fifty-three loci on 17 genes on the Ashkenazi Jewish panel were amplified in micro assay wells called "through-holes" using TaqMan chemistry on a QuantStudio™ 12K Flex OpenArray®. Each array chip contains 48 subarrays, each of which is capable of amplifying 64 loci in separate through-holes. Thus, a single OpenArray chip can test 48 patients simultaneously. We chose to test patient samples in duplicate, therefore 24 patient samples and controls are tested on a single chip. Four OpenArray chips can be amplified in a single run, the entire work flow from set up to results takes approximately 4-5 hours to complete. Forty-five previously genotyped, positive, de-identified DNA specimens were tested on the OpenArray chips to define assay performance and turn-around time. Additionally, six plasmid pools and three individual plasmid stocks including all loci detected in this assay were used to test the specificity of primers and probes. **Results** Accuracy of both the AJ assay was found to be greater than 99%. The failure rate of total targets is relatively low at less than 5%. However these sporadically failed targets occur over multiple samples, making the percentage of patient repeats higher than 5%. Plasmid targets verified oligo specificity and complemented the positive sample pool for variants for which we did not have positive genomic DNA. **Discussion** Sporadic call ambiguity and amplification failure of single loci necessitates running AJ samples in duplicate. Multi-sample array provides rapid turn-around time for our expected AJ sample volumes and is considerably more cost effective than alternative methodologies such as NGS. The QuantStudio format is a high quality, low cost method for carrier screening.

802W

The application of targeted next-generation DNA sequencing to newborn screening for inherited metabolic diseases. *R. Gao[1], J.L. Ren[1], C. Liu[1], W.X. Zhen[1], Z.H.W. Huang[1], J. Zhu[2,3], H. Zhang[4].* 1) FirStories science and technology Co., LTD; 2) The National Office for Maternal and Child Health Surveillance of China; 3) West China Second University Hospital, Sichuan University; 4) PrecisionMDX Co., LTD.

Purpose: In Mainland China, LC-MS/MS newborn screening (NBS), have succeeded in reducing morbidity and mortality for approximately 13 years. However, several issue of LC-MS/MS method, such as relative low recall rate, instability of metabolite and complex cutoff value setting, have limited its application in NBS program. DNA sequencing, on the other hand, won't be influenced by those issues. Furthermore, using trace-level DNA as start-up material, genetic testing could be performed by applying only one dried blood spot. This kind of genetic testing could be introduced as a second-tier method in conventional LC-MS/MS NBS program, and facilitate the confirmation process. Methods: A total of 63,024 dried blood samples (DBSs) of newborns have been screened for 48 IMDs by using LC-MS/MS from Jan 2014 to Dec 2015. Two hundred and sixty subjects have an abnormal result from the initial test and 178 (68.5%) were recalled successful for further confirmation. The other 82 subjects failed to be recalled and nobody knows whether they have diseases or not at the moment. We had access to collect 50 samples from the 82 subjects and analyzed them on TNGS (target-capturing next generation sequencing) of 48 inherited metabolic diseases (IMDs). The metabolites' concentration of all the 50 samples was far out of the cutoff threshold in first-time LC-MS/MS screening. The TNGS test was performed on the residual DBS after the first-time screening. Result: From the conventional NBS confirmation process, 15 of the 178 recalled subjects were finally diagnosed with certain diseases. It is noticeable that miss detection of 82 failed recall subjects could inevitably contribute to false negative in the conventional NBS program. For the 50 subjects conducted with genetic testing, 9 new subjects were detected with disease-causing mutations in the candidate gene. Taken the abnormal LC-MS/MS result into consideration, we infer these 9 samples are with high-risk of the correspondence diseases. Nevertheless, all of them were ignored by the conventional NBS program due to failed recall for further diagnosis. This could be the main source of false negative in the conventional NBS program. We thereby believe that the application of TNGS to LC-MS/MS NBS programs for all the positive subjects after first-time screening is beneficial. Especially, the process doesn't need repeat sample collection and all the subjects could benefit from the early, presymptomatic diagnosis.

803T

Defining *KIR* and *HLA class I* genotypes at highest resolution using high-throughput sequencing. *P.J. Norman[1], N. Nemat-Gorgani[1], W.M. Marin[2], S.J. Norberg[3], M. Ronaghi[3], J.A. Hollenbach[2], P. Parham[1].* 1) Structural Biology, Stanford, Stanford, CA; 2) Neurology, University of California San Francisco School of Medicine, San Francisco, CA; 3) 3.Illumina Inc., 5200 Illumina Way, San Diego, CA.

The physiological functions of natural killer (NK) cells in human immunity and reproduction depend upon diverse interactions between killer cell immunoglobulin-like receptors (KIR) and their HLA class I ligands: HLA-A, -B and -C. Extensive allelic and structural variation of *KIR* genes directly affects NK cell activity by diversifying the expression, signaling capacity or HLA binding properties of KIR allotypes. The genomic regions containing the *KIR* and *HLA class I* genes are unlinked, structurally complex, highly polymorphic, and co-evolving. They are also strongly associated with a wide spectrum of disease, including infections, autoimmunities, cancers, and pregnancy disorders, as well as the efficacy of transplantation and other immunotherapies. To facilitate study of these extraordinary genes, we developed a method that captures, sequences and analyzes the 13 *KIR* genes (MIM: 604936-7, 604945-7, 604952-6, 605305, 610095, 610604) and the *HLA-A*, -B and -C genes (MIM: 142800, 142830 and 142840) from genomic DNA. We also devised a bioinformatics pipeline to call high-resolution genotypes for each of the *KIR* genes. PING (Pushing Immunogenetics to the Next Generation) uses a novel combination of virtual sequence probes and filtering to attribute sequencing reads to specific *KIR* genes and determine their alleles and copy number. The complete method was validated using DNA from well-characterized cell lines by comparison with established methods of *HLA* and *KIR* genotyping. We also determined *KIR* genotypes from the 1000 Genomes Project sequence data using PING. This identified 116 novel *KIR* alleles, which were all demonstrated authentic by sequencing them from source DNA using standard methods. Analysis of just two *KIR*, showed that 22% of the 1000 Genomes individuals have a novel allele or a structural variant. In analyzing over 1000 Europeans with the complete method we identified one in ten individuals to harbor a previously undiscovered *KIR* variant. These findings indicate heterozygosity at the *KIR* locus likely rivals that of the *HLA class I* genes, which have more than 10,000 alleles identified following sequencing millions of potential bone marrow donors worldwide. The method we describe is suited to the large-scale analyses that are needed for characterizing human populations and defining the precise *HLA* and *KIR* factors associated with disease. The methods are applicable to other highly polymorphic genes. PING is an open-source R program available on Github.

804F

Well-groomed participants: Eyebrow plucks as surrogates for biomarker samples and a viable source of constitutional DNA. *D. Castillo[1], J. Herzog[1], S. Sand[1], T. O'Connor[2], C. Clark[2], J.N. Weitzel[1,2].* 1) Division of Clinical Cancer Genetics, City of Hope National Medical Center, Duarte, CA; 2) Beckman Research Institure, City of Hope, Duarte, CA.

Purpose: The human hair follicle is a potential source of proliferating epithelial cells and can be used in multiple clinical applications. This study tested the feasibility of using plucked eyebrow hairs as biomarker samples in pharmacodynamics (PD) studies and as quality surrogates for DNA sources. Previous reports of eyebrow plucks as DNA damage indicators following drug treatment have suggested that they would be useful as PD markers for poly(ADP-ribose) polymerase (PARP) inhibitor and other drug-related DNA damage. **Methods**: For PD studies on NCT01149083, a PARP inhibitor trial, plucked follicles were stored in saline and processed within 24 hours; control samples were treated with H_2O_2 and fluorescent antibodies, using confocal microscopy to reveal YH2AX foci as an indicator of double strand breaks. After creating a standard operating procedure (SOP) for collecting and processing eyebrow plucks to extract DNA, samples with 10-15 plucks were obtained from subjects with suspected somatic mosaic *TP53* mutations or hematologic malignancies who enrolled in an IRB-approved registry. **Results**: For the PD study YH2AX foci were detected in the control samples, but not in patient samples following veliparib and carboplatin treatment . Eyebrow pluck uniformity was limited due to variability inherent in the fact that hair follicles have 5 stages of maturation. Eyebrow plucks were obtained from 11 patients for the alternate source of DNA study; 5 to establish whether a previously found *TP53* mutation was mosaic or constitutional, and 6 taken from allogeneic transplant patients seen for genetic testing. Of the 11 samples, 9 were clinical quality (>1 μg yield with 260/280 OD=1.8-2.0), and 3 of these were successfully used in commercial *TP53* testing to establish clonal hematopoiesis in mosaic *TP53* mutations found in the blood that were not present in the eyebrow DNA. The 2 samples not meeting standards had fewer quality hair follicle bulbs and had longer times from pluck isolation to DNA extraction. **Conclusions**: While DNA yield is limiting, collection of eyebrow hair follicles is less invasive and faster than skin biopsy. We obtained quality DNA as long as the sample was handled in a timely manner and in accordance with our SOP. Thus, applications may include assessment of DNA for patients with bone marrow transplant or hematological disorders. The variability of follicle maturation must also be considered when designing PD studies for this biospecimen.

805W

Evaluation of the gene *FANCA* DEL/DUP through MLPA in a cohort of Fanconi anemia patients. *L. Bobadilla-Morales[1,2], A. Vázquez Reyes[1], E. Zapata-Aldana[2], E. Torres-Anguiano[1], H.J. Pimentel-Gutiérrez[2], C. Ortega-de-la-Torre[2], M. Ortíz[2], R. Silva-Cruz[1], L. Corona-Bobadilla[1,2], F. Sánchez-Zubieta[2], J.R. Corona-Rivera[1,2], A. Corona-Rivera[1,2].* 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo, Instituto de Genética Humana, CUCS, Universidad de Guadalajara; 2) Servicios de Genética, 2Unidad de Citogenética y 3Hematoncología, División de Pediatría, Hospital Civil de Guadalajara Dr. Juan I. Menchaca.

Introduction: The Fanconi anemia (FA) is an inherited disorder from the DNA reparation system characterized with progressive pancytopenia and impaired bone marrow function, various congenital malformations and predisposition to develop solid or hematologic tumors (OMIM: 227650). Herein are characteristic, growth delay, short stature and hematologic, musculo-sceletal, renal, cardiovascular, endocrinologic, dermatologic and central nervous system disorders. The FA has an incidence rate of 1 in 350,000. It has been established 15 genetic subtypes, but 60% of the patients present the gen FANCA (16q24.3) mutation. The MLPA technique is an effective method to detect an abnormal number of copies through multiple PCR. In the present study we evaluated a cohort of patients with FA as studied in the Hospital Civil de Guadalajara "Dr. Juan I. Menchaca" (HCGJIM) aiming to detect the presence of DEL/DUP of the gene FANCA applying the MLPA technique. **Methods.** Descriptive study accomplished during the period January 2009 to May 2016 in a cohort of patients attended by the Pediatric and Genetic Hemato-Oncology services in the HCGJIM with clinical diagnosis of FA. All the patients included in such study had cytogenetical corroboration of the choromosomal instability evaluated by means of mitomycin C as inductor. Prior informed consent it was obtained DNA from peripheral blood. For the MLPA assay it was deployed the MLPA kit® SALSA p031/p032 *FANCA* (MCR-Holland, Amsterdam, The Nehterlands). Data analysis of the data was performed by the software Coffalyser®. **Results.** We evaluated 14 patients presenting positive choromosomal instability before time of exposure to mitomycin C. Morover, one of the 14 patients presented homozygotic deletion of the exon 11 to 18, whose parents presented heterozygotic state; furthermore, another patient and his mother showed gains of the 42 exon. **Conclusions.** Our results show a frequency of deletions lower to the reported, prospectively due to the fact that we are a reference center, thus we do not represent the general population, all of our patients therefore show hematologic disorders; on the other hand, the gain of the 42 exon in a heterozygotic way should be established as a Fanconi Anemia coincidence. Before anemia development, the MLPA method could be applied as an initial screening in order to detect deletions of the gene FANCA, and as an aid to search for the gene carriers.

806T

Fanconi anemia: Five years experience for patients referred to the Human Cytogenetics Department National Research Center - Egypt.
M.M. Eid[1], G.Y. El Kamah[2]. 1) Human Cytogenetic , National Research Center, Cairo, Egypt, Cairo, Dokki, Egypt; 2) Clinical Genetics , National Research Center, Cairo, Egypt, Cairo, Dokki, Egypt.

Fanconi anemia (FA) is one of a well described chromosomal instability disorder. It is inherited usually as autosomal recessive disorder except complementation subtype B inherited as X -linked disorder. To date 19 complementation subtype have been reported in the literature. Bone marrow failure leading to aplastic anemia is usually the presenting symptoms, unless the patient has an apparent congenital malformation such as thumb or radial rays anomaly. About 10% of the patients the presenting symptom is the myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML). The diagnosis of FA depends on the clinical assessment and induction of chromosomal breakage by diepoxy butane (DEB).   In this work we report on the patients referred to our department in the last 5 years for DEB test and various studies carried out for them. A total of 238 patients were referred to perform DEB test for, the common presenting feature is pancytopenia. The other causes for the referral is either being a sib for patient with Fanconi anemia, or having one of common features of FA such as radial anomaly with or without thumb affection and/or skin pigmentation abnormalities and microcephaly. Out of the 238 patients, FA was diagnosed in 70 patients, 32 were normal sibs for the FA patients they were referred either to exclude FA or the possibility to be a proper bone marrow donor for their affected sib and 136 were diagnosed as having acquired aplastic anemia. The male to female ratio of the FA patients was 1.25:1. Six families show 2 affected sibs. Usually the second sib diagnosed accidentally through the family investigation. MDS and acute myeloid leukemia was the presenting symptoms in 3 patients. The sensitivity for the DEB is variable between the patients. 8 patients show mosaicism in their response to DEB. These group of patients have normal cell line in 40% up to 90% of their studied metaphases. The patient that has 90% normal cell line she has near normal blood picture and her age was 18yrs.   Immunoglobulin evaluation shows low level of all immunoglobulin groups in one FA patient. Anti oxidant and pro oxidant status studies were done for FA patients which revealed a significant response to various antioxidant.   Although our research for FA have covered different aspects still we did not achieve much as regard molecular aspects. However mutation screening using MLPA is currently under study and we expect to reach valuable results from it.

807F

HLA-DQ1 alpha and beta genotypes associated with non-celiac gluten sensitivity. *M. Maki, D. Caporale.* Biology, University of Wisconsin Stevens Point, Stevens Point, WI.

Non-Celiac Gluten Sensitivity (NCGS) is a gastrointestinal disease occurring in 3-6% of the population, while celiac disease (CD) and wheat allergy (WA) are only found in 2-4%. This indicates that 17 million Americans may have NCGS. Only by eliminating a diagnosis of CD and WA through medical testing and the elimination of foods from their diet can a patient be diagnosed with NCGS. Some reports have shown that proteins embedded in the membranes of leukocytes known as human leukocyte antigens (HLAs) are linked to comparable gastric diseases like CD and WA. It is possible that certain HLA genotypes may also be associated with NCGS. To test this hypothesis, we screened a cohort of 50 NCGS and 50 non-NCGS participants for their HLA DQ alpha and beta genotypes, using AllSet+ Gold DQ alpha 1 and beta 1 high resolution allele-specific PCR kits. DNA fingerprints were generated by electrophoresing the PCR products on large 3% agarose gels and photodocumented for allele identification. Since the HLADQ surface protein consists of one alpha polypeptide and one beta polypeptide and both subunits are in association with the antigen molecule, it seemed most appropriate to combine the alpha and beta genotypes for each individual and observe allele combinations within and between each cohort. For simplicity, alpha and beta allele types were numbered α1 through 5 and β1 through 6, respectively. So far, the most common genotype combinations within the gluten-sensitive cohort were A1-3/B2-6 and A1-5/B2-6. The most common genotype combinations for the control group were A1-3/B3-5 and A1-3/B3-6; both were also unique to the control group. The genotype combination A1-3/B2-6 was only found within the gluten-sensitive cohort. Interestingly, allele A5 (DQA1:05) and allele B2 (DQB1:02) are both higher in frequency within the gluten-sensitive group than within the control group. Previous studies have shown high association between these two allele types and Celiac disease. We are continuing to screen more participants within each cohort, to test the statistical significance of our findings. If the A5 and B2 allele frequencies continue to be significantly higher in the gluten-sensitive cohort, then screening for these alleles may be used as a more efficient diagnosis tool in the future.

808W

Spectrum of alpha and beta thalassemia mutations in Singapore. *K. Tan[1], P.T. Huan[1], L. Chiu[1], S.S. Chong[1,2], E.S.C. Koay[1,3].* 1) Laboratory Medicine, National University Hospital, Singapore, Singapore; 2) Department of Paediatrics, Yong Loo Lin School of Medicine, National University Singapore, Singapore; 3) Department of Pathology, Yong Loo Lin School of Medicine, National University Singapore, Singapore.

Background Thalassemia is a common inherited gene disorder and is prevalent in South East Asia. Molecular testing of the alpha and beta globin genes plays an important role in the diagnosis of thalassemia. Methods The Molecular Diagnosis Centre in the National University Hospital in Singapore performs molecular analysis for thalassemia which includes DNA extraction and mutation analysis for both alpha and beta globin genes. The alpha globin gene cluster is examined for seven deletional mutations (SEA, -3.7, -4.2, FIL, THAI, -20.5 and MED) by gap-PCR, and six point mutations within the alpha2-globin (*HBA2*) gene (codon 30 / Hb G, codon 59 / Hb Adana, Hb Quong Sze, Hb Constant Spring, Hb Pakse and polyA signal), known to be common to individuals in this region, by PCR and sequencing. The beta globin gene was examined by PCR followed by sequencing of all three exons. Results Over a period of 10 years from 2006 to 2016, peripheral blood from 970 individuals were analyzed. There were no mutations found in 363 patients. The most common alpha globin deletion mutations were the -3.7 and SEA deletions, which were each present in 20% of the individuals tested. Forty eight individuals were homozygous for the -3.7 deletion, and 13 had both the -3.7 and the SEA deletion together, while only 2 were homozygous for the SEA deletion. The -4.2 deletion was present in 17 cases, while the FIL and THAI deletions were present in 6 and 2 cases respectively. The -20.5 and MED deletions were not detected. Among the beta globin mutations, the HbE mutation was the most common, and was present in 6% of all cases. The next most common beta globin mutations were the codons 41/42 (-TCTT) mutation (26 cases), the IVS1 nt5 mutation (22 cases) and the IVS2 nt654 mutation (21 cases). Five percent of all cases (49 individuals) had both alpha globin mutations and beta globin mutations together. Conclusion Molecular analysis is valuable in clarifying the diagnosis of thalassemia and both the alpha and beta globin genes should be examined. .

809T

Whole exome sequencing for identification of genetic alteration of light chain amyloidosis. *S.Y. Kim[1,3], Y. Park[1], H. Kim[1], Y. Oh[1], J. Lim[1], Q. Choi[2], G.C. Kwon[1], S.H. Koo[1].* 1) Department of Laboratory Medicine, Chungnam National University School of Medicine, Daejeon, South Korea; 2) Department of Laboratory Medicine, Dankook University Hospital, Cheonan, Republic of Korea; 3) Cancer Research Institute, Chungnam National University School of Medicine, Daejeon, Republic of Korea.

Light chain (AL) amyloidosis is a plasma cell disorder characterized by a clonal population of plasma cells that produce monoclonal immunoglobulin chains that are deposited in tissues, resulting in associated symptoms. In AL amyloidosis patients, immunoglobulin is deposited in tissues before a large tumor burden develops; therefore, AL amyloidosis patients typically do not present with overt multiple myeloma (MM) at the time of diagnosis. Because of rarity of disease and lower burden of clonal plasma cells, few data on genomic alterations of clonal plasma cells in AL amyloidosis have been reported. In this study, we performed whole exome sequencing (WES) of in bone marrow (BM) and peripheral blood (PB) cells collected from 5 AL amyloidosis patients and 2 MM patients. Four AL amyloidosis patients presented monoclonal gammopathy related to light chain lambda type, and one AL amyloidosis patient presented kappa light chain deposition. The median plasma cell burden in BM aspirate samples of AL amyloidosis patients was 10% (range, 2%-41%). After extraction of DNA from total nucleated cells of BM and PB samples, WES was performed using Illumina HiSeq 2000. Putative mutations in BM samples of each patient were analyzed in comparison to PB samples. No previously reported hotspot mutations in hematologic malignancies were found in 5 AL amyloidosis patients. We intensely investigated for the presence of 95 variants of AL amyloidosis that was recently reported by Paiva et al. (2016), however, none of these mutations were found. In our patients, non-recurrent variants were found in each case in genes including *GBP4*, *SH2D4A*, and *AR*. In addition, variants in the intergenic region including SLC25A46-TSLP were also found. Our results was concordant with previous studies of WES of AL amyloidosis in that unifying mutation by WES is absent in AL amyloidosis. To elucidate pathogenesis of AL amyloidosis, deeper and broader genetic investigations beyond exome may be needed.

810F

The frequency *F8* gene intron 22 inversion in patients with hemophilia A from Russian Federation. *T. Beskorovainaya, T. Milovidova, A. Polyakov.* Federal State Budgetary Institution "Research Centre for Medical Genetics", Moscow, Russian Federation.

Introduction: Hemophilia A (HA) is a frequent X-linked recessive bleeding disorder resulted from deficiency or dysfunction of coagulation factor VIII (FVIII). It affects 1 of 5000 males. HA is caused by mutations in *F8* gene located on chromosome Xq28 and consisted of 26 exons. More than 3000 mutations were described in the gene. The most common mutation is the intron 22 inversion (Inv22). It was found in 40-45% patients with severe HA (coagulation activity of FVIII <1%). Materials and Methods: In this study we searched Inv22 in unrelated probands with HA (38 patients) and relatives of probands whose material was not available (33 persons) from 71 Russian families using inverse-PCR method (Rossetti L.C. et al., 2005). Results: Inv22 was found in 32.4% cases. Mothers of probands with detected Inv22 were carriers of this mutation in all cases. Four probands with clinical diagnosis "hemophilia B" (HB) were also analyzed for Inv22 which was found in one case. Moreover, in two families large rearrangements of the *F8* gene not due to inversion intron 22 were identified by the presence of abnormal length fragment on the electrophoresis. Further these fragments were analyzed by direct sequencing. Conclusions: Results of this study are consistent with literature data. It was suggested that coagulation test allowing to distinguish HA and HB wasn't made for all patients. This technology has wider diagnostic capabilities than expected and can detect some genetic rearrangements besides the *F8* gene inversion intron 22 (long deletions or insertions up to exon 21).

811W

Comparative analysis of whole exome sequencing and targeted next-generation sequencing in patients with primary immunodeficiencies. *J. Chou, W. Bainter, B. Cangemi, M. Massaad, R. Geha.* Division of Immunology, Boston Children's Hospital and Harvard Medical School, Boston, MA.

Primary immunodeficiencies (PIDs) are typically monogenic disorders characterized by recurrent infections, autoimmunity, and/or malignancy. Identifying the genetic etiology of PIDs is challenging due to genetic diversity and phenotypic heterogeneity. Targeted next-generation sequencing (NGS) panels and whole exome sequencing (WES) have been used to identify genetic causes of PIDs, but there are no published studies comparing these two approaches. A total of 148 individuals with clinical and/or laboratory data indicative of a PID were sequenced. WES was performed on the first 101 study subjects using the Illumina HiSeq 2000 system with an average coverage of 150x. Downstream processing and variant calling were performed using SAMtools, Picard Tools, and the Genome Analysis Toolkit. The average analysis time for each exome was two hours. Targeted NGS was performed on the next 47 study subjects enrolled, using the Ion S5™ to sequence 264 genes associated with immune function. Average coverage was 406x. Variant calling and annotation were performed with IonReporter™. The average analysis time for each subject's targeted NGS panel was 25 minutes. WES identified causative variants in 56 of 101 individuals (55%), of which 43 had a pathogenic variant in a gene known to cause PID. Nine of the 56 individuals had a deleterious variant in a gene not previously associated with human disease, resulting in a disease-causing effect confirmed by functional studies. Four individuals had causative mutations in genes unrelated to immune function, resulting in a clinical phenotype resembling a PID. Nineteen (19%) of the 101 individuals had candidate variants that remain under investigation, and 26 (26%) had no identified genetic etiology. Targeted NGS sequencing identified causative genetic variants in 38 of 47 individuals (81%). Three of the 47 individuals (6%) had candidate variants that require additional investigation; no causative variant was found in the remaining 6 individuals (13%). Targeted NGS had a shorter analysis time than WES and identified a causative variant in a higher percentage of individuals. This difference may be due to the higher depth of coverage and greater ease of identifying large exonic deletions. These findings support the use of targeted NGS for initial genetic screening of patients with PIDs, with WES reserved as a second-tier approach for identifying novel causes of human disease.

812T

TACI, *ICOS* **and** *BAFFR* **mutation analysis in 449 Common Variable Immunodeficiency (CVID) patients.** *E. Pariltay[1], A. Aykut[1], A. Durmaz[1], F. Hazan[2], N. Gulez[3], N. Karaca[4], H. Onay[1], O. Ardeniz[5], G. Aksu[4], F. Genel[3], N. Kutukculer[4], F. Ozkinay[1].* 1) Ege University, Faculty of Medicine, Department of Medical Genetics Izmir, Turkey; 2) Dr. Behcet Uz Children's Hospital Department of Medical Genetics Izmir, Turkey; 3) Dr. Behcet Uz Children's Hospital Department of Pediatric Immunology Izmir, Turkey; 4) Ege University, Faculty of Medicine, Department of Pediatric Immunology Izmir, Turkey; 5) Ege University, Faculty of Medicine, Department of Immunology Izmir, Turkey.

Common Variable Immunodeficiency (CVID) is an immune system disorder resulting recurrent infections of respiratory and gastrointestinal tract. CVID is the most common clinically manifested primary immunodeficiency. Bacterial infections are significand and also CVID patients may develop autoimmune and lymphoproliferative complications. Human immune system is very complicated and mandatory for defensing against viral and bacterial infestations. Molecular mechanism is not well understood yet. In CVID patients at least two isotypes of immunoglobulin defect lead hypogammaglobulinemia. Primary defect is not clear but CVID is related with some gene defects such as "Transmembrane activator and CALM interactor (TACI or TNFRSF13B)", "Inducible co-stimulator (ICOS)" and "B-cell-activating factor receptor (BAFFR or TNFRSF13C)". Tumor necrosis factor receptor superfamily (TNFRSF) member TACI is most frequent gene defect that almost %10 of the CVID patients have at least one mutation, but unaffected heterozygous mutation carriers also reported that suggests unknown regulatory elements are needed for clinical manifestation. Here we present mutation analysis of 449 individuals who are clinically diagnosed as CVID. TACI gene is sequenced in all of the cases and in 36 of the patients we found at least one mutation. p.C104R (n=8), p.C172Y (n=5) and R202H (n=4) were the most common mutations. We found two novel mutations with unknown clinical significance. At 7 of the cases ICOS gene is sequenced and at the 64 of the cases both ICOS and BAFFR genes are sequenced. All sequenced ICOS genes (73) were normal. We found c.475 C>T (p. H159Y) / c.62 C>G (p.P21R) mutation at BAFFR gene in two unrelated cases. Parental controls suggested these mutations to be cis position. These changes are known as polymorphism but also reported to be disease causing at cis position. Beyond these tree genes CVID molecular mechanism is still need to be investigated. Wide definition of clinical diagnosis and un-clarified molecular regulatory mechanisms complicate genetic diagnosis. We found mutation in almost 9% of the patients which is consistent with literature.

813F

Complete androgen insensitivity syndrome due to the silent p.S510S mutation causing a premature stop codon in androgen receptor gene. *A.D.S. Rodrigues, R.L. Batista, M.Y. Nishi, E.M.F. Costa, S. Domenice, B.B. Mendonca.* Unidade do Desenvolvimento, Laboratorio de Hormonios e Genetica Molecular, Hospital da Clinicas-Faculdade de Medicina da Universidade de Sao Paulo, Sao Paulo, Brazil.

Background: Androgen insensitivity syndrome (AIS) is the most common cause of 46,XY disorder of sex development (DSD). The clinical spectrum of this disorder ranges from partial to complete form (CAIS). Most CAIS patients have defects in androgen receptor (*AR*) gene. The existence of a shortened *AR*-transcript variant in addition to the full-length *AR*- messenger RNA (mRNA) has been reported in and this seems to be the result of a variable splicing event. The association between this type of mutation with complete form of AIS has not been reported. **Objective**: To describe a silent mutation altering the splicing site in *AR* in two families with CAIS. **Patients:** three 46,XY affected members from one family and a single 46,XY patient from another family, all of them with normal female external genitalia, with a history of primary amenorrhea, inguinal hernia and spontaneous breast development. **Method**: The whole coding region of the *AR* gene, including exon/intron boundaries, were PCR-amplified and submitted to direct automated sequencing in ABI PRISM 3130XL. *AR* cDNA was obtained from reverse transcription of mRNA extracted from testicular tissue of one affected patient (family 1) and amplified by PCR. The size of the fragment amplified was compared with a commercial pool of testes mRNA. **Results:** Sanger sequencing identified the silent mutation c.1530C>T (p.S510S) in exon 1. Mutation taster and Human Splicing Finder sites predicted the variant as deleterious due to splicing site changing. Sequencing of partial cDNA, flanking the mutation region, showed that the patient's *AR* mRNA lost around 90 bp at the end of exon 1 leading to a premature stop codon. **Conclusion**: We identified a novel mutation in the *AR* gene associated an abnormal splicing resulting in a truncate protein. This is the first description of a silent mutation associated to CAIS. Our results further expand the spectrum of mutations associated with the AIS and may contribute to the understanding of the molecular mechanisms involved in splicing defects.

814W

A novel mutation of *NR5A1* gene in 46,XYDSD with gender dysphoria. *H. Yagi[1], M. Takagi[1], A. Ohtake[2], Y. Hasegawa[1].* 1) Tokyo Metrpolitan Children's Medical Center, Tokyo, Japan; 2) Saitama Medical University, Saitama, Japan.

   Androgen exposure to the fetal brain is well known to be associated with male-typical behavior. Genetic cause, hormonal effects and psychosocial factors also influence pshycosexual development. Pshychosexual development of 46,XY disorder of sex development (DSD) is not well known, and the actual prevalence of the gender identity in DSD patients is reported variously. Here we report on a 31-year-old social female with 46,XY DSD. She was clinically diagnosed as 46,XY DSD, through the examination of inguinal hernia. When she became teenager, she felt herself as a male without knowing her karyotpye. At 31 years old, she referred to our hospital for further infromation about her disorder. Through counseling, she selected neither of the sex orientation for herself. We checked all the coding exons and flanking introns of *NR5A1* by PCR and direct sequencing, and found a novel heterozygous mutation of c.227-229delTGG (p.Val76Glyfs*15). Functional analyses revealed that this novel mutation lost transcriptional activity as a result of loosing DNA binding affinity. This is the first report to describe gender dysphoria in individuals with *NR5A1* gene mutation. Early accurate diagnosis and reccurent explanation depending on their age and long-term psychosocial suppots are necessry for DSD patients.

815T

47,XXY Klinefelter syndrome associated with partial androgen insensitivity phenotype conferred by a complete inactivating mutation in the androgen receptor gene. *R.L. Batista, A. Feitosa, A.D.S. Rodrigues, E.M.F. Costa, S. Domenice, B.B. Mendonca, M.Y. Nishi.* Laboratorio de Hormonios e Genetica Molecular LIM42, Hospital das Clinicas da FMUSP, Sao Paulo, Brazil.

   Background: Klinefelter syndrome (KS) and androgen insensitivity syndrome (AIS) are among the most common disorders of sex development and rarely, both syndromes affect the same individual. **Clinical case**: A 19 years old healthy woman sought medical attention due to primary amenorrhea and severe pain in both inguinal regions. At physical examination she was a proportional tall overweighed young female (height=173.3 cm (SD +1.85); weight: 89.4 Kg (SD +4.2), BMI: 30; Breast development was Tanner V and pubic hair was normal at Tanner IV stage. External genitalia examination disclosed an enlarged clitoris (3.5x1.5 cm), two perineal opening and bilateral inguinal gonads. Her karyotype was 47,XXY. Ultrasonography of the inguinal region disclosed a 3.6 -4.9 mL male gonads. Hormonal investigation revealed elevated basal LH (20 U/L - nv: 0.4 to 5.7) and FSH (35 U/L -nv: 1.1 to13.5) levels and testosterone levels at normal male levels (325 ng/dL- nv: 260 to 990). She has undergone bilateral orchiectomy at 20 years old and anatomopathological result was consistent with seminiferous tubules atrophy. The analysis of the androgen receptor (*AR*) gene using genomic DNA revealed a heterozygous mutation c.3070_3071insA at exon 7, which results in a frameshift, leading to a premature stop codon (p.Asn849Lysfs*32). This mutation has been described in several patients with CAIS. Microsatellite analysis using 14 X-STS markers identified the same haplotype in Xp (8 markers) and different haplotype in Xq (6 markers), indicating a uniparental dissomy of X-chromosome. KS is rarely associated with ambiguous genitalia, mainly with almost female external genitalia. This fact associated with high LH levels, normal male testosterone levels and normal breast development suggested complete androgen insensitivity syndrome. The heterozygous state of *AR* mutation and microsatellite analysis point to a maternal (Xm1Xm2Y) nondisjunction event in meiosis I, plus a recombination between maternal short arms of X chromosome. The PAIS phenotype of our patient suggests that normal *AR* is partially expressed. Probably, because of random X-inactivation in tissues our patient had a complete breast and pubic hair development. **Conclusion**: We report a female patient with KS and PAIS phenotype due to a nonsense *AR* mutation that should generate a CAIS phenotype. We emphasize the importance of considering this association in KS patients with ambiguous genitalia.

816F

Cytogenetic findings in children with ambiguous genitalia. *I.I. Qazi.* microbiology and molecular genetics, University of the Punjab, lahore, punjab, Pakistan.

Qazi I.I. Department of Microbiology and Molecular Genetics, University of the Punjab Lahore, Pakistan. In affiliation with Department of Cytogenetics, Children's Hospital and Institute of Child health, Lahore, Pakistan. **Abstract:** Sexual ambiguity is a complex and often confusing medical problem in new born. Determination of the under lying cause of ambiguity is an urgent matter especially in congenital adrenal hyperplasia (CAH). This study was aimed to identify the chromosome compliment of children presenting with ambiguous genitalia in order to describe the underlying cause of their condition and consequently to reassign gender to these patients. A total of 25 patients with ambiguous genitalia, aged between 0-18years were studied over the period of one year. Peripheral blood samples were cultured in RPMI-1640 media and harvested through a routine as well as synchronized procedure. Chromosomes at metaphase were karyotyped for chromosome compliment in addition to the identification of any gross abnormality. Hormonal profile and abdominal ultrasound were also conducted to supplement the diagnosis. In this study, before cytogenetic analysis, 60% (15/25) of the patients were assigned a male status, while female sex was assigned to 40% (10/25) due to the appearance of their external genetalia at the time of birth. After karyotyping 44% (11/25) turned out to be with 46, XY complement and 52% (13/25) had 46, XX compliment. Klinefelter's syndrome accounted for 4% (1/25). Furthermore, pseudohermaphroditism and more specifically congenital adrenal hyperplasia in majority of these patients was found to be due to 21- hydroxylase deficiency. Additionally, Male pseudohermaphrodites accounted for 40% (10/25) of the cases. True hermaphrodites, Klinefelter's syndrome accounted for 4% (1/25) cases. The current study has helped in reassigning the sex in patients with misassigned or unidentified sex at an early stage so that the social trauma be reduced. It aims further to help in diagnosis of salt losing state of 21-hydroxylase deficiency which is a life threatening condition. Hormonal profile of these patients has also helped in planning the required hormone therapy for these patients. **Key words:** Ambiguous genitalia; Congenital adrenal hyperplasia (CAH); Chromosome compliment; External genitalia; Karyotype; Klinefelter's syndrome; Pseudohermaphrodite; Peripheral blood; True hermaphrodite.

817W

Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients. *M. Lamontagne[1], N. Gaudreault[1], F. Maltais[1], W. Tan[2], J. Bourbeau[3], Y. Bossé[1,4] for the Canadian Respiratory Research Network, for the CanCOLD study group.* 1) Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Québec, Canada; 2) University of British-Columbia, Vancouver, British-Columbia, Canada; 3) McGill University, Montréal, Québec, Canada; 4) Department of Molecular Medicine, Laval University, Québec, Canada.

Background: Alpha-1 antitrypsin deficiency (AATD) is one of the most common hereditary disorders. AATD is caused by genetic mutations in the *SERPINA1* gene and is associated with increased risk of lung and liver diseases. Only a minority of AATD patients is identified and the current multi-step diagnosis algorithms are not helping to enhance recognition of this condition. **Methods**: DNA sequencing of the coding regions of *SERPINA1* was performed as a single diagnostic test to detect AATD. The frequencies of deficient alleles were evaluated in 400 Canadian patients with chronic obstructive respiratory disease (COPD). **Results**: Nineteen genetic variants were identified including 15 missense mutations. These includes three mutations known to generate normal protease inhibitor (Pi) M1 to M4 alleles; five causing deficient Pi alleles Z, S, P$_{Lowell}$, F, and I; and seven with unknown significance. Overall, at least 9.4% of alleles in this Canadian COPD cohort were deficient. Three patients were identified as AATD including two homozygotes ZZ and one compound heterozygote SZ. Seventy-one out of 400 patients (18%) were carriers of at least one deficient allele. DNA sequencing can be easily implemented, with a turnaround of less than 48 hours, and a price that is lower than the multi-step algorithms currently used to diagnose AATD. **Conclusions**: DNA sequencing of *SERPINA1* is effective as a single test strategy to detect AATD. Simplifying and improving diagnostic strategy will ultimately expedite the clinical decision-making process recommended by many national and international lung societies about targeted testing and augmentation therapy.

818T

Mutational hotspot in the *ACTG2* gene and mutation frequency in Chronic Intestinal Pseudo-Obstruction. *A. Milunsky[1,2], C. Baldwin[1], X. Zhang[1], D. Primack[1], A. Curnow[3,4], J. Milunsky[1,2].* 1) Center for Human Genetics, 840 Memorial Drive Suite 101, Cambridge, MA; 2) Department of Obstetrics and Gynecology, Tufts University School of Medicine, 860 Washington Street, Boston, MA; 3) St. Luke's Regional Medical Center, 190E Bannock Street, Boise, ID; 4) University of Washington Center for Mendelian Genomics, Seattle, WA.

Chronic Intestinal Pseudo-Obstruction [CIPO] or hollow visceral myopathy, mostly with megacystis, is a progressive mostly autosomal dominant disease due to a failure of enteric smooth muscle. This study aimed to determine the mutational spectrum in the ACTG2 gene in CIPO and the frequency of a mutation in this monogenic disorder. Sequelae include a high mortality rate, vomiting, abdominal pain and distension, and a massive non-functional bladder, frequently evident at birth, and prenatally by ultrasound. Life-long total parenteral nutrition and bladder catherization is the rule, compounded by catheter sepsis, urinary tract infection, and liver compromise. Diagnosis has depended on clinical features, manometry and imaging, and often unnecessary intestinal biopsy. Initial Whole Exome Sequencing of 4 families revealed only one with an ACTG2 mutation. Subsequent Sanger sequencing of 24 additional probands detected another 3 with mutations. Detailed clinical data of the 4 affected probands and their siblings or parents (7 in all) include the following: all 7 had CIPO and megacystis, 3 died in childhood, and 3 had a late-term prenatal diagnosis. Other features of the 7 were colectomy (3), ileostomy (3), G-tube (3), GERD (4), and lifetime TPN (4). At 38 years, after 29 years of TPN, one had total visceral transplantation with removal of intestine, stomach, liver, pancreas and spleen. We combined our 28 probands with published probands who had ACTG2 sequencing and found that 45/106 (42.5%) harbored a pathogenic variant. Moreover, a mutational hotspot in the ACTG2 gene was recognized. Genetic heterogeneity was also evident. New pathogenic variants in the ACTG2 gene are common, the rate being unknown at present. Early precise diagnosis by ACTG2 sequencing is important, helps avoid unnecessary invasive and other tests, and enables future avoidance/prevention by prenatal or pre-implantation genetic diagnosis. Clinical and genetic heterogeneity is recognized while another gene or genes are still to be determined.

819F

Molecular characteristics of 100 Vietnamese patients with congenital hyperinsulinism. *D.C. Vu[1], D.A. Dang[1], N.B.T. Can[1], K.N. Nguyen[1], T.P. Bui[1], H.T. Le[1], D.P. Nguyen[1], D.M. Tran[1], S. Flanagan[2], S. Ellard[2].* 1) Medical Genetics and Metabolism, Vietnam Children's Hospital, Hanoi, Viet Nam; 2) University of Exeter Medical School. Royal Devon & Exeter Hospital. South West NHS Genomic Medicine Centre. UK.

Background: Hyperinsulinemic hypoglycemia (HH) is a consequence of unregulated insulin secretion by pancreatic β-cells. Congenital HH is caused by mutations in genes involved in regulation of insulin secretion (*ABCC8, KCNJ11, GLUD1, CGK, HADH, SLC16A1, HNF4A* and *UCP2*). Severe forms of congenital HH are caused by inactivating mutations in *ABCC8* and *KCNJ11,* which encode the two components of the pancreatic β-cell ATP-sensitive potassium channel. **Objective and hypotheses:** Our aim is to identify mutations in the *ABCC8* and *KCNJ11, HNF4A* and *GLUD* genes, and to describe genotype and phenotype correlations of Vietnamese children with congenital hyperinsulinism. **Method:** A prospective study was conducted on 100 cases with congenital hyperinsulinism diagnosed and treated at Vietnam Children's Hospital from January 2007 to March 2016. Patients were selected by using inclusion criteria of Hussain K (2008). All exons of *ABCC8; KCNJ11, HNF4A* and *GLUD1* were amplified from genomic DNA and directly sequenced. **Results:** Mutations were identified in 51 cases (51.0%) including mutations of *ABCC8* gene (45 cases; 45.0%), Among these cases 26 with homozygous/compound heterozygous of *ABCC8* and 19 cases with one paternal/maternal mutation of *ABCC8* gene); *KCNJ11* (5 cases; 5.0%), *HNF4A* (1 case; 1.0%). 100% of cases with homozygous/compound heterozygous recessive mutations or one paternal dominant mutation of *ABCC8* gene did not respond to diazoxide treatment and required 95% pancreatectomy or octreotide injection. Other cases without identified mutations responded to diazoxide and/or glucose infusion. **Conclusion:** children with congenital hyperinsulinism should be performed mutation analysis which helps in making diagnosis and treatment decision. Families of children with congenital hyperinsulinism should be given genetic counseling. Prenatal diagnosis should be performed as well as follow - up and treatment should be given to children with congenital hyperinsulinism immediately after birth.

820W

Update of *CFTR* **recurrent mutation spectrum in Russia.** *T. Adyan, A. Stepanova, S. Krasovsky, A. Polyakov.* Research Centre for Medical Genetics, Moscow, Russian Federation.

   Cystic fibrosis (CF) is a common and generally severe autosomal recessive disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator gene (CFTR). Over 1500 putative mutations have been described in *CFTR*, which can be divided into classes according to the molecular mechanism of the genetic defect. The CFTR mutation detection rate varies by test method and ethnic background. A total of 386 CF patients from different regions of the Russian Federation were investigated. DNA of the probands was analyzed for the presence of 38 common mutations in CFTR gene by multiplex AFLP-analysis (CFTRdele2,3 (Del21kb), F508del, I507del, 1677del-TA, 2143delT, 2184insA, 394delTT, 3821delT, L138ins, 3944delGT, 604insA, 4015delA, 4022insT) and by MLPA-method (E92K, G542X, R334W, S1196X, 621+1g>t, W1282X, N1303K, 3849+10kbC>T, 3272-26A>G, W1282R, 2789+5G>A, 3272-16T>A, S466X, 1898+1G>A, 3120+1G>A, R347P, S945L, G85E, S1159P, W1310X, R553X, 3849G>A, 712-1G>T, S1159P, G551D, dup7-11). As a result of the analysis in 309 probands CFTR mutations were identified on both chromosomes and in 77 probands on a single chromosome. The most frequent mutation was F508del with allelic frequency 53,6%; the following by frequency of occurrence were CFTRdele2,3 (Del21kb) (6,5%) and E92K (5,3%) mutations. Another 10 mutations met with frequencies ranging from 1% to 3%, 17 mutations met with a frequency of less than 1%, 8 mutations did not meet at all. Total allelic frequency of 38 studied mutations in the sample of patients accounted 90%, probability of finding two mutations in proband - 81% and probability of finding at least one mutation - 99%.

821T

Assessment of genetic pathogenicity for combined pituitary hormone deficiency. *Q. Fang, A.S. George, P. Gergics, M.L. Brinkmeier, A.H. Mortensen, L.Y.M. Cheung, M.I.P. Millan, A. Daly, A. Ajmal, S.A. Camper.* Department of Human Genetics, University of Michigan, Ann Arbor, MI.

   Combined Pituitary Hormone Deficiency (CPHD), or panhypopituitarism, is defined as the deficiency of growth hormone and at least one other pituitary hormone. CPHD affects 1 in 8,000 individuals. Both genetic and environmental factors contribute to the incidence of CPHD. Since *POU1F1* was reported as the first causative gene in 1992, more than 30 genes and about 2,500 genetic variants have been identified in CPHD patients. Additional CPHD genes will be discovered by increased clinical resequencing of patient genomes. There has not been a systematic analysis to review the pathogenicity of the reported genes and variants using the criteria set in 2015 by the American College of Medical Genetics and Genomics (ACMGG) and the Association for Molecular Pathology (AMP). We searched for genetic causes of CPHD by keywords in PubMed, OMIM and HGMD databases and evaluated them at the gene-level using 4 criteria: 1) Were mutations in the gene found in different affected families? 2) Is the gene expressed in the developing head, pituitary and/or hypothalamus? 3) Is the gene involved in known biological pathways for pituitary development and physiology? 4) Is there a convincing functional study to support the role of the gene in hypothalamus or pituitary gland? Only 22 genes passed the gene-level evaluation by meeting at least 2 out of 4 criteria. Using the 2015 guidelines, we classified the pathogenicity of each individual variant in those 22 genes by the inheritance and segregation pattern, the allele frequency in the population, the computational and predictive annotations, and *in vitro* and *in vivo* functional data. The recessive variants, such as those in *PROP1* and *POU1F1*, meet the criteria for authentic pathogenicity consistently. Dominant variants with incomplete penetrance, such as those in *GLI2* and *OTX2*, are difficult to classify without strong evidences from functional studies, and some of these cases may involve digenic or oligogenic diseases. In an international collaboration, we sequenced the exomes of 64 CPHD patients and unaffected relatives from 26 different families. We found mutations in several known genes and identified several candidate genes for functional studies. The high degree of genetic heterogeneity for CPHD makes it difficult to achieve all four criteria for causality amongst novel candidate genes. Our systematic analysis of CPHD genetics provides a framework for clinical geneticists and endocrinologists to evaluate future cases. .

822F

A case of hypopituitarism with 9q34.3 subtelomeric deletion syndrome. *S. Higuchi, M. Takagi, Y. Hasegawa.* Endocrinology and metabolism, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

   Introduction; While reports of short stature, micropenis or cryptorchidism with 9q34.3 subtelomeric deletion syndrome have been reported, hypopituitarism with the syndrome has not been reported. We here present a 6-year-old Japanese male with hypopituitarism developed in 9q34.3 deletion. **Case report;** He was born uneventfully by spontaneous cephalic delivery. At the age of 1 year, he was referred to our department for genital anomalies. He had developmental delay, hypotonia, atrial septal defect, micropenis and cryptorchidism without hypoglycemia and polyuria. Microarray was performed in the inspection of the multiple malformations. 9q34.3 deletion was detected. The parents did not have the deletion. Height velocity began to decrease around the age of 2 years. At the age of 3 years, we diagnosed him as having hypothyroidism by thyroid function test. We started l-thyroxine. At the age of 4 years, height was 88.0 cm (-3.3 SD). Endocrinological tests at this time showed GH deficiency. The pituitary was hypoplasia on MRI and posterior pituitary bright spot was orthotopically noted. Cortisol peak to CRH showed adrenocortical insufficiency at the age of 5 years, when hydrocortisone therapy was started. 9q34.3 deletion does not include the genes, which reportedly cause congenital hypopituitarism. We will analyze the mutations for congenital hypopituitarism. **Conclusion;** ;This is the first report of hypopituitarism with 9q34.3 subtelomeric deletion syndrome. We should check endocrinological tests when following up patients with 9q34.3 deletion who have retarded growth, genital abnormalities or hypothyroidism.

823W

Prevalence of an *Alu* insertion mutation in the *CASR* gene. *B.Y.L. Wong[1], G.N. Hendy[2], D.E.C. Cole[3].* 1) Molec Gen Lab, Sunnybrook Hlth Sci Ctr, Toronto, ON, Canada; 2) Experimental Therapeutics and Metabolism, McGill University Health Centre-Research Institute, and Departments of Medicine, Physiology and Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Dept of Clinical Pathology, Sunnybrook Health Sciences Centre; Depts of Lab Medicine & Pathobiology, Pediatrics, and Medicine, University of Toronto, Toronto, Ontario, Canada.

   Inactivating mutations in the *CASR* gene cause Type 1 familial hypocalciuric hypercalcemia (FHH) if heterozygous, and neonatal severe hyperparathyroidism if homozygous. Most *CASR* mutations involve a single nucleotide mutation or small insertion/deletion (indel). Copy number variations due to large rearrangements are reportedly quite rare. The only large insertion reported within the *CASR* gene is a founder *Alu*-insertion (~300bp) mutation described in 3 large inter-related kindreds from south west Nova Scotia in 1995. It has not been seen since. Recently, however, we identified the very same *Alu*-insertion mutation c.[2631_2643dup{ins{T(88);oS77605.1}}] in 4 geographically distant and apparently unrelated FHH individuals. Two of the cases were identified when the genomic samples were screened for *CASR* mutations by Sanger sequencing. Normally, the *Alu* insertion would not be detected by routine PCR based methods because of failure to amplify due to the long polyT tail in the *Alu* element. However, the possibility that an insertion might be present arose when allelic dropout of the nearby p.A986S SNP signal was noted in one of two overlapping primer sets. Using an *Alu* specific primer set, the two individuals were confirmed by sequencing to carry the *Alu*-insertion identical to that reported for the Nova Scotian deme. The other 2 cases were found when 49 putative FHH samples, previously screened negative for *CASR* mutation by denaturing HPLC and either heterozygous or homozygous for the p.986S minor allele, were analyzed with the *Alu* specific primer set. Among the 4 new cases we found, one came from the west coast of Canada (Greater Vancouver Area), two from central Canada (Toronto) and one from western Canada (Calgary). Given the prevalence in our case series, we have elected to add a screening step for all of our p.986S positive subjects putatively identified as FHH but with an otherwise negative *CASR* mutation screening result. The origin and full geographic dispersal of this mutation is not known. Surveys at other test centres would be of interest and of practical importance in the molecular diagnostic workup of FHH and related conditions.

824T

Identification of pathogenic variants in Osteogenesis Imperfecta by massively parallel sequencing. *T. Felix[1,2], L. Todeschini de Souza[1,2], E. Brizola[2], B. Pinheiro[2], M. Zambrano[2], A.P. Vanz[2].* 1) Servico Genetica Medica, Hosp Clinicas Porto Alegre, Porto Alegre, RS, Brazil; 2) Reference Center for Osteogenesis Imperfecta, Porto Alegre, RS, Brazil.

Osteogenesis imperfecta (OI) is a heritable bone fragility disorder associated with dominant inheritance in 85-95% of cases due to mutations in type I collagen genes.. Genes associated with recessive forms (10-15% of cases) are involved in type I collagen post-synthetic modifications or trafficking. **Objective:** The aim of this study was to identify pathogenic variants in 18 genes associated with type I collagen pathway in patients with clinical diagnosis of OI by Massively Parallel Sequencing. **Methods:** The study comprised 28 individuals with a clinical diagnosis of OI type III and IV without previous molecular analysis and a case with type V OI with known *IFTIM5* c.-14C>T mutation as a control. The panel was costumized by Ion AmpliSeq Designer (www.ampliseq.com/ Life Technologies) and cover 99.6% of the coding region of *COL1A1, COL1A2, CRTAP, LEPRE1, PPIB, WNT1, TMEM38B, SERPINH1, BMP1, SP7, SERPINF1, FKBP10, SMPD3, CREB3L1, PLOD2, P4HB, PLS3, IFITM5* and 5`UTR region of *IFITM5.*The sequencing was performed by the Ion Torrent Personal Genome Machine (Ion-PGM) platform. Data from the sequencing run were processed using Torrent Suite software (version 5.0; Life Technologies) for base calls, read alignments, and variant calling using the reference genomic sequence (hg19) of target genes. Called variants were annotated using Ion Reporter (version 5.0; Life Technologies). The variants were classified according to ClinVar-NCBI and LOVD 2.0 databases. In silico prediction tools were used for analysis of variants of uncertain significance and included SIFT , Polyphen-2, PredictSNP and Mutation taster **Results:** Of the 28 cases, 404 variants were identified. Of these, 266 missense and small indel mutations were located in intronic regions, 101 in exonic regions and 37 in splice sites, 3' or 5' UTR regions. Several mutations were previously reported as pathogenic variants according to the databases. Mutations in 4 genes were identified: *COL1A1*(10 mutations), *COL1A2* (6 mutations), *LEPRE1*(1 mutation) and *SERPINF1*(1 mutation). Ten variants predicted to be pathogenic were identified in *COL1A1*, *COL1A2* , *LEPRE1* , *SERPINF1, PLOD2*, *BMP1,, CREB3L1and P4HB*. **Conclusion:** At least, one pathogenic mutation (reported or predicted) was found in each individual. This study confirms that this panel with 18 genes using based semiconductor sequencing may identify missense and small indel mutations in disorders with genetic heterogeneity as OI. .

825F

From phenotype to genotype and back again: Whole exome sequencing in the diagnosis of patients with skeletal dysplasia. *A. Beleza-Meireles[1], M. Jansson[2], P. Louro[1,3], S. Lillis[2], J. Hoyle[2], J.W. Ahn[2], K.J.P. Ryan[2], A. Calder[4], C. Hall[4], A. Offiah[5], M. Simpson[6], S. Yau[2], S. Mohammed[1], M. Melita[1].* 1) Clinical Genetics Department, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; 2) Viapath Analytics LLP, London, United Kingdom; 3) Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal; 4) Radiology Department, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; 5) Academic Unit of Child Health, Sheffield Children's NHS Foundation Trust, Sheffield, United Kingdom; 6) Division of Genetics and Molecular Medicine, King's College London School of Medicine, London, United Kingdom.

Background: Skeletal dysplasia (SD) is a genetically heterogeneous group of over 350 distinct disorders. An experienced clinician may be able to identify the clinical diagnosis of a specific SD condition with the aid of a skeletal survey. Multidisciplinary work, involving geneticists, radiologists, orthopaedic surgeons and growth specialists is often necessary in the diagnostic work-up of these patients. Nonetheless, it is frequently challenging to reach a clinical diagnosis, and genetic analysis using a single gene testing strategy is often unhelpful owing to difficult as a result of significant genetic heterogeneity. Increasingly, however, the use of whole exome sequencing (WES) is overcoming diagnostic uncertainty and providing new opportunities for accurate, rapid characterisation of skeletal dysplasia conditions. Our purpose is to highlight the clinical utility of WES analysis as a diagnostic tool for SD, discuss new challenges it can present, and emphasise the continued importance of clinical phenotyping in the interpretation of WES data.   Methods: Twenty-three patients with presumed SD were selected from the specialist multidisciplinary clinic. WES was performed using Agilent SureSelectXT Human All Exon V5 followed by sequencing on an Illumina HiSeq 2500. Raw sequence data was aligned using Novoalign, and variants called with Samtools. Qiagen Ingenuity Variant Analysis was used to aid assessment of variant pathogenicity after applying a virtual panel of 220 SD genes in combination with multidisciplinary clinical interpretation.   Results: Pathogenic variants were detected in ten patients; while in five other patients we found variants of uncertain pathogenicity, which require additional work such as inheritance studies. Of the remaining eight cases, no pathogenic variants were detected in the SD gene panel. Post-WES clinical reassessment suggested that three of these eight cases are unlikely to have a SD.   Conclusions: WES has confirmed, or provided a diagnosis, in 43% (10/23) of our patients, and identified possible candidates in another 22% (5/23). These results demonstrate that WES analysis using targeted bioinformatic analysis is a cost-effective diagnostic tool for SD. The importance of thorough clinical evaluation and the close liaison with clinical scientists has proven essential in providing an optimal and efficient service. Detailed phenotyping of patients with presumed SD remains of paramount importance in the diagnostic pathway.

826W

RIPPLY2: **Believe it or not.** *N. Vena[1], L. Rochford[2], R.J. Hopkin[2], C. Rajakuruna[2], T.A. Burrow[2], S.A. Hosseini[1].* 1) Claritas Genomics, Cambridge, MA; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

We present on a 9 mo. old male with multiple vertebral anomalies associated with compound heterozygous *RIPPLY2* variants. The proband was born to non-consanguineous parents via vaginal delivery. Cervical spine instability due to complete absence of the C2 vertebra, hypoplasia of the other cervical vertebrae, craniocervical junction stenosis and spinal cord compression lead to spinal cord injury during labor and delivery resulting in respiratory failure, hypotonia and decreased mobility. He subsequently became hypertonic with spasticity and muscle spasms and continues to be trach-dependent. Absence of cervical vertebrae is extremely rare and no genetic causes have been identified. Exome testing identified two variants in *RIPPLY2* in the compound heterozygous state. *RIPPLY2* encodes a nuclear protein belonging to a family of proteins required for vertebrate somitogenesis and functions in the *NOTCH* signaling pathway; several other genes involved in this pathway are associated with segmentation defects of the vertebrae (SVD). Null mutant mice show axial skeleton segmentation defects due to defective somitogenesis. In humans, limited case reports of individuals with SDV and Klippel-Feil syndrome associated with *RIPPLY2* variants have been reported. The c.240-4T>G variant in *RIPPLY2* was previously reported in two siblings with multiple regional segmentation defects of the vertebrae. Functional studies have not been performed, but this variant is predicted to affect the splice acceptor consensus sequence. The c.286G>A variant in *RIPPLY2* is unreported in the literature or control datasets. It occurs at a highly conserved nucleotide and amino acid, and is predicted to be deleterious by *in silico* prediction models. The two variants were observed on opposite strands in the NGS data, suggesting a *trans* configuration. Due the limited information available regarding these variants, they were classified as uncertain. However, due to the clinical diagnosis in the proband, the *trans* configuration, and the role of *RIPPLY2* in somitogenesis, they likely represent the molecular diagnosis for the proband's phenotype. This case highlights the role of *RIPPLY2* in vertebral development and provides further evidence that *RIPPLY2* is a gene responsible for autosomal recessive malformation of the cervical vertebrae. Functional studies are necessary to determine the pathogenicity of the identified variants and to fully evaluate the phenotypic spectrum of disease associated with this gene.

827T

Next generation sequencing as a confirmatory diagnostic tool in a Brazilian Skeletal Dysplasia Tertiary Center. *W.A.R. Baratela[1], G.L. Yamamoto[1,2], L.C. Testai[2], R. Honjo[1], L.A.N. Oliveira[7], M.J.R. Doriqui[4], R. Rosa[8], J.G.C. Meira[3], J.M. Pina-Neto[6], D.P. Cavalcanti[5], C.A. Kim[1], M.R. Passos-Bueno[2], D.R. Bertola[1,2].* 1) Unidade de Genética, Instituto da Criança - Hospital das Clínicas - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Centro de Pesquisas sobre o Genoma Humano e Células-tronco, Instituto Biociências – Universidade de São Paulo, SP, Brazil; 3) Serviço de Genética - Hospital Universitário Professor Edgard Santos/ Universidade Federal da Bahia, Salvador, Bahia, Brazil; 4) Hospital Infantil Dr Juvêncio Mattos Rua Sao Pantaleao, s/n centro, Sao Luis, Maranhão, Brazil; 5) Department of Medical Genetic, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil; 6) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil; 7) Departamento de Radiologia - Hospital das Clínicas - Universidade de São Paulo, São Paulo, Brazil; 8) Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

Skeletal dysplasias (SD), a heterogeneous group of genetic conditions impairing normal bone growth and development, currently encompass 436 disorders classified in 42 groups based on molecular, biochemical and/or radiographic criteria. Our group (Hospital das Clínicas/University of São Paulo) started a SD clinic in 2012, aiming not only to provide accurate diagnosis, but also to address clinical aspects of these conditions. Approximately 400 individuals have been included in the SD cohort after clinical and radiographic analysis. An algorithm was established to test patients by Sanger sequencing with the current available gene analysis (*FGFR3*, *TRPV4*, *LIFR*, *SOST*, *CHST3*, *CANT1*, or *WNT7A*) or the new implemented next generation sequencing (NGS): a targeted panel for 120 genes related to SD or exome sequencing. The latter for cases in which the disease associated gene was not included in the panel or the etiology unknown. Approximately 113 cases were tested by Sanger sequencing (84 for *FGFR3*), 94, by the NGS panel, and 9 families by exome sequencing. A molecular diagnosis was achieved in 83/113 (73%) by Sanger, in 60/94 patients (64%) by panel, and 3/9 by exome. Applying the current classification of SD to the cohort of NGS panel, the positivity in groups was: I (3/4), II (7/11), III (2/4), IV (8/10), VII (4/5), VIII (3/4), IX (2/5), X (3/3), XI (1/1), XIII (4/4), XV (1/2), XVI (1/2), XVIII(1/1), XVIII (1/1), XX (1/1), XXI(0/1), XXII (0/1), XXIV (1/2), XXV(10/13), XXVI (2/2), XXVII (1/1), XXXV (4/15), XLII (0/1), while in exome: II (0/1), V (*HSPG2* 1/1), IX (0/1), XII (*PCYT1A* 1/1), XVI (0/1), XXI (*PEX7* 1/2), XXXVI (0/1), XXXVIII (0/1). The two most prevalent groups in SD (I and XXV) were under-represented in NGS due to Sanger testing for *FGFR3* and selection of only the severe or consanguineous cases of OI for testing. The lowest sensitivity in XXXV probably occurred due to a less strict clinical criteria for the disorder. Of interest, a case presenting brachyolmia and mutations in both copies of *PAPSS2* showed skeletal abnormalities similar to *TRPV4* spectrum, influencing genetic counseling. Furthermore, through exome, we found mutations in *HSPG2* in a case of Rolland-Desbuquois. Although analysis of this sample of patients is ongoing, initial findings show that NGS sequencing has been concordant to the clinical/radiographic diagnosis in around 64% of patients. Due to heterogeneous nature of SD, 33% have their molecular basis yet to be discovered. FAPESP.

828F

Mutation screening of the γ-secretase genes in sporadic and familial Hidradenitis Suppurativa (acne inversa) and acne conglobata. *D.D Jhala[1], U. Ratnamala[2,3], N.K. Jain[2], N.M. Saiyed[4], B.C. Gorijala[5], P.V.S. Prasad[6], U. Radhakrishna[7].* 1) Department of Zoology, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India; 2) Department of Life Sciences, School of Sciences, Gujarat University, Gujarat University, Ahmedabad, India; 3) Department of Pharmacology, Creighton University, Omaha, NE, USA; 4) Biotechnology, Nirma Institute of Science, Nirma University, Ahmedabad, India; 5) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, India; 6) Department of Dermatology, Annamalai University, Chidambaram, India; 7) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

   Hidradenitis Suppurativa (acne inversa) is a chronic, inflammatory and recurrent, debilitating skin disease that usually affect**s** people between puberty and middle age. It is a progressive disease where single boil-like, pus-filled abscesses become hard lumps, followed by painful, deep-seated, often inflamed clusters of lesions with chronic seepage. It most commonly affects areas bearing apocrine glands. The γ-secretase genes have been identified as the cause for a subset of familial HS. Twenty-five γ-secretase gene mutations have been published, including two from our research group, accounting for almost 88.5% of Asian and European familial HS. Remarkably, the genetic variants have been described mostly in families with a history of the disease, including acne conglobata; however, only a very small proportion of sporadic patients have reported mutations. Direct sequencing of DNA samples from 2 large families and 95 sporadic cases with HS and acne conglobata screened in the entire coding sequence and splice junctions of the γ-secretase complex genes *NCSTN*, *PSENEN,* and *PSEN1* showed only one pathogenic variant in a family with acne inversa and acne conglobata and excluded one family and 95 sporadic cases. A nonsense mutation, 1876C > T in exon 16, which would result in a substitution of Arg at codon 626 (p. Arg626X) and premature stop codon was observed in one family. The 95 sporadic cases and a large HS family with multiple affected members did not reveal pathogenic variants in either coding or noncoding regions including splice sites. This report will be valuable to delineate possible genotype-phenotype associations in HS. The present study provides evidence that HS is a heterogeneous disease as γ-secretase genes are involved only in the pathogenesis of some familial and/or sporadic HS cases. Epigenome-wide studies also should be considered in HS individuals excluded with γ-secretase complex gene mutations. .

829W

Acrodermatitis enteropathica-like forms of atopic dermatitis revealed by whole-exome sequencing. *S. Küry[1], T. Besnard[1], X. Latypova[1], D. Quinquis[1], S. Schmitt[1], I. Ricordeau[1], M. Sedghi[2], J. Nasiri[2], M. Salehi[2], S.P. Paul[3], D. Candy[3], E. Blouin[4], S. Bezieau[1].* 1) CHU de Nantes, Nantes, France; 2) Isfahan University of Medical Sciences, Isfahan, Iran; 3) St Richard's Hospital, Chichester, UK; 4) Laboratoire LABCATAL, Montrouge, France.

   Background: Acrodermatitis enteropathica (AE; MIM #201100) is a rare and severe autosomal recessive zinc deficiency disorder molecularly diagnosed by the identification of a biallelic anomaly of zinc transporter gene *SLC39A4*. Its pathognomonic symptom is an acral and periorificial dermatitis occurring in early life that responds in a spectacular fashion to zinc therapy. Yet, our experience in the molecular diagnosis of AE showed that more than a half of the cases of suspected AE were not due to a primary zinc dyshomeostasis that would be associated by alterations of *SLC39A4*. **Purpose:** Our goal was to determine the genetic cause of genodermatoses very suggestive of an AE, albeit with no *SLC39A4* mutation. Method: We sequenced the whole-exome of a series of simplex and multiplex families with a AE-like dermatitis. **Results:** Beside the identification of a few exceedingly rare syndromes mimicking AE, the most striking result was the presence of dominant mutations in filaggrin and in laminins, some of them were already known to be associated with an increased susceptibility to atopic dermatitis (AD). **Conclusions:** Our work highlighted atopic dermatitis as the main clinical diagnosis of AE. This suggests that a sub-group of the common disease represented by AD could be diagnosed by genetic testing. Further studies are needed to determine whether these severe cases of AD are exceptional or if the efficiency of zinc therapy can be extended to the treatment of all AD patients with similar anomalies.

830T

Delineation of genetic alterations in patients with pigmentary mosaicism.
C. Salas-Labadia[1], V. Ulloa-Aviles[1], M.P. Navarrete-Meneses[1], A. Reyes-Leon[1], D. Martinez-Anaya[1], R. Criz-Alcivar[1], O. Castro-Ayala[1], S. Gomez-Carmona[1], C. Duran-McKinster[1], E. Yokoyama-Rebollar[1], E. Lieberman-Hernandez[1], A. Gonzalez-Del Angel[1], V. Del Castillo-Ruiz[1], P. Perez-Vera[1], D. Cervantes-Barragan[2], A. Ruiz-Herrera[3]. 1) Instituto Nacional de Pediatria, Mexico City; 2) Hospital Central sur de Alta Especialidad, Pemex, Mexico City; 3) Hospital de Especialidades Pediatricode Leon, Leon, Guanajuato, Mexico.

Patients with pigmentary mosaicism (PM) are a heterogeneous phenotypic group, characterized by specific skin lesions (Blaschko lines pattern), but in association with a great variety of central nervous, musculoskeletal and ocular alterations. Miscellaneous chromosomal aberrations have been observed in 30-60% of PM cases, and also other types of genetic alterations have been reported like microdeletions and mutations. The aim of this study was: a) To establish a complete cytogenetic and molecular characterization of patients with PM and b) To determine whether our strategy analysis contributes to the establishment of a better characterization of this entity. We included 63 patients with PM and other systemic anomalies. The GTG-bands from lymphocytes, light and dark skin samples were analyzed, reviewing 50 metaphases per tissue. Molecular analysis with SNP array and/or FISH were performed to finely characterize chromosomal abnormalities (CA), and also in some patients with normal karyotype but strong clinical suspicion of chromosomal alterations below the detection level of cytogenetics. A total of 60 cases have been completely analyzed by cytogenetics and the results could be classified as follows: **1)** Normal karyotype n=39; **2)** Mosaics with 2 or more different cell lines with structural alterations n=12: **a)** mos del(18)(q21)/normal (nl) (2 patients); **b)** mos +i(12)(p10)/nl; **c)** mos r(22)(p11.2q13.3)/dup r(22)(p11.2q13.3)/nl (2 patients); **d)** *mos r(7)(p22q36.3)/dup r(7)(p22q36.3)/-7; **e)** t(1;8)(p?36;q?22)/nl; **f)***mos +mar/+14/nl; **g)** *mos +mar/45,X/nl; **h)** *mos +mar/nl (3 patients); **3)** Balanced translocations X;autosomes n=2: **a)** t(X;16)(p11.2;p13.3); **b)** t(X;22)(p11.2;q13.3); **4)** Polyploidy (mosaic) n=1; **5)** *Others n=6. In *6 patients with abnormal karyotype, the molecular analysis delineated the structural alteration, identified other structural alteration (<5Mb) and characterized the different cell lines proportions. In 1/5 patients with normal karyotype was identified a 603Kb deletion in 16p11.2. Our analysis strategy could contribute to a better characterization of patients with PM. It was possible to demonstrate that cytogenetics is a very important first filter for identification of CA. Molecular analysis allowed the delineation of the abnormalities and identified CA not detected previously by GTG-bands. The combination of both methodologies allowed us to perform better associations with the phenotype. CONACyT 1882277, Recursos Federales 2016, Genetadi.

831F

Linked-reads accurately map variants to *NEMO* and its pseudogene (*IKBKGP*). *F.A. Mafra, C. Kao, R. Pellegrino, J. Garifallou, C. Kaminski, A. Cederquist, E. Ryan, T. Wang, L. Tian, H. Hakonarson.* Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Anhidrotic Ectodermal Dysplasia with ImmunoDeficiency (EDA-ID) and Incontinentia Pigmenti (IP) are two rare diseases arising from mutations in *IKBKG/NEMO*. Located in Xq28, this gene encodes the regulatory subunit of the IKK complex that activates NF-kB, which is involved in a variety of physiological and cellular processes, such as immunity, inflammation, cell proliferation, and survival. *NEMO* hypomorphic mutations, have been identified in EDA-ID and IP patients, while null mutations seem to be exclusively associated with IP. The structural architecture around *NEMO* imparts genomic instability that can lead to pathogenic alterations. On the 5' side, *NEMO* partially overlaps with the *G6PD* gene, which is a recombination hotspot containing multiple *Alu* sequences; on the 3'-side *NEMO* is part of a low-copy-repeat (LCR1) with an inverted duplicated copy (LCR2) located ~75kb away that contains a highly conserved pseudogene (*IKBKGP*) with >99% sequence identity to *NEMO* between exons 3-10. The LCRs are prone to inversion/gene-conversion events causing the repositioning/copying of mutations from pseudogene to gene. Current whole exome and targeted NGS-based sequencing represents a challenge. Short reads arising from regions such as *NEMO* may not be mapped accurately since it would be difficult to know if they are coming from the gene or the pseudogene. Clinical testing for variants in *NEMO* traditionally requires customized methods such as long-range PCR, which can be laborious and slow. We illustrate an NGS-based method here to resolve reads and variants in *NEMO* and its pseudogene, using linked reads generated by 10x Genomics Chromium technology in combination with whole exome sequencing using Agilent SureSelect baits. We used linked read WES on a number of "truth" samples with established variants in *NEMO* and/or *IKBKGP*. One candidate immunodeficiency family in particular is highlighted, where a candidate variant was identified in the putative splice junction site of exon 9. The linked read WES results was in accordance with the long-range PCR and Sanger sequencing, demonstrating that this is a non-pathogenic SNP present in both *NEMO* and *IKBKG*. We also used linked read WES to accurately map and confirm variants in samples from known female carriers of *NEMO* mutations.

832W

Hereditary neuralgic amyotrophy and intellectual disability in a patient with 3.85 MB 17q25.1-q25.3 duplication involving *SEPT9.* C. Esmer[1], M. Villegas[2], C. Gómez-Elias[2], S. Avila[3], J.L. Castrillo[3], A. Bravo-Oro[2]. 1) Genetics, Hospital Central "Dr. Ignacio Morones Prieto", San Luis Potosi, Mexico; 2) Child neurology, Hospital Central "Dr. Ignacio Morones Prieto", San Luis Potosi, Mexico; 3) Genetadi, Biotech, Bilbao, España.

Several unbalanced rearrangements of terminal 17q have been described associated with monosomy or trisomy of other chromosomes; pure 17q25 duplications are infrequent and remain inadequately delineated. Partial duplications in septin 9 (*SEPT9*) was shown to be causal for hereditary neuralgic amyotrophy, an episodic peripheral neuropathy with autosomal dominant inheritance but duplication of the entire gene have been reported only once. We report an individual with pure non-recurrent 17q25.1-q25.3 duplication. The case was evaluated for the first time at 8 years of age referred to genetic consultation due to developmental delay, poor scholar performance, attention deficit disorder and craniofacial dysmorphism. At physical examination showed short forehead, hypotelorism, short palpebral fissures, high nasal bridge, small mouth with protruding lips, retromicrognatia. Physical examination at the age of 18 months showed relative macrocephaly, a large anterior fontanel, triangular facies, a prominent forehead, hypertelorism, down-slanting palpebral, low-set ears and abnormal posterior hairline. The mother reveals a history of episodic pain and impaired movement of the left shoulder and arm. The muscles in the affected area showed mild atrophy and weakness, mild scapular winging was also present. The array CGH revealed a gain in 17q25.1-q25.3 spanning approximately 3.85 Mb (72,695,356-76,548,866) and contains at least 86 genes. The 17q23-qter portion is the region most commonly duplicated in almost all the dup 17q cases reported in literature. Duplication 17q has been associated with a wide range of clinical findings, including psychomotor/mental retardation, growth retardation, and dysmorphic features that were also present in our case. This region includes genes involved in several diseases, such as retinitis pigmentosa, ciliary dyskenesia, microcephaly, bradyopsia, etc. This is the first time that a patient with a large chromosomal 17q duplication shows clinical evidence of hereditary neuralgic amyotrophy.

833T

A 30 gene clinical panel for dementia: The first 100 cases. S.A. Bucks[1], C.J. Hale[2,3], D.S. Hanna[2,3], M.B. Dauer[2,3], L. Mathew[2,3], D. Sarroza[2,3], S. LaDuke[2,3], M.A. Weaver[2,3], M. Rumbaugh[1], D. Tsuang[1,4], S. Jayadev[1], T.D. Bird[1,4,6], M.O. Dorschner[2,3,5]. 1) Department of Neurology, University of Washington, Seattle, WA; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Center for Precision Diagnostics, University of Washington, Seattle, WA; 4) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 5) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 6) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle, WA.

Dementia, or cognitive impairment, affected an estimated 35.6 million people worldwide in 2010. As the elderly population continues to grow, the number of dementia cases and the associated socio-economic burdens will swell in concert. Several neurological syndromes cause or associate with dementia, including Alzheimer disease (AD), Parkinson disease (PD), frontotemporal dementia (FTD), and vascular dementia. Identifying specific genetic changes associated with dementia in individual patients contributes to accurate diagnosis, inclusion into appropriate clinical trials, and potential discovery of novel gene targets that could be pursued for therapeutic development. Additionally, if a known genetic cause is discovered, unaffected family members may opt for genetic testing to help determine their risk status and assist with life planning. We have developed a clinical exome-based panel for over 30 genes known to have common causative, rare causative, and risk variants for dementia. The clinical panel was developed by a group of experts actively engaged in neurodegeneration research and clinical practice at the University of Washington, and it includes genes from a variety of diseases associated with a dementing phenotype, including AD, PD, FTD, and amyotrophic lateral sclerosis. We used an IDT Exome v1.0 capture system for library enrichment followed by Illumina sequencing. More than 100 individuals with early onset dementia have been subjected to testing with this panel to evaluate its performance. We have identified potential disease-associated variants in >16 genes known to cause familial AD or related disorders. Further analysis of these variants is ongoing and includes the use of disease-specific and exome-derived variant databases, literature searches, and *in silico* predictive algorithms for altered mRNA and/or protein structure and function. Results are returned to patients and their families using a multidisciplinary team including neurologists and genetic counselors to help convey the appropriate utility and caveats of test results. Overall, our clinical dementia panel is comprehensive and can identify disease-causing variants in patients.

834F

Evaluating the strength of evidence for gene/condition relationships: Impact on multi-gene panel testing design. *C. Tan, D. Riethmaier, M. Hogue, N. Patil, K. McDonald Gibson, T. Winder, J. Westbrook, S. Topper.* Invitae Corp., San Francisco, CA.

    Background: As multi-gene panels have become routine in evaluating patients for genetically heterogeneous conditions, growing inter-laboratory variability has occurred in the number of genes offered in a given panel for many disorders. Although some of this variability can be attributed to the relationship between the timing of assay design and the rapid discovery of new gene–condition relationships, it is becoming increasingly important to understand the clinical validity of the diverse offerings of multi-gene panels in today's molecular diagnostic setting. Establishing the clinical validity of a multi-gene panel depends on an accurate and detailed understanding of the validity of each included gene. We proposed a method for establishing the clinical validity of genes and evaluated that method with a set of pediatric and neuro-muscular gene–condition relationships. Method: A robust variant classification method based on ACMG guidelines was used to evaluate the pathogenicity of published, clinically observed variants. Gene–condition relationships were categorized as strong (convincing evidence that an observed variant meets criteria for pathogenicity) or suggested (preliminary evidence). A third category, emerging, was used to describe an additional condition purportedly caused by a gene already determined to have a strong relationship with a different condition. Our method was applied to 187 genes implicated in a range of pediatric-onset and neuromuscular conditions. Results: We categorized 175 genes as having a strong association with at least one condition, of which 18 also had an additional suggested or emerging association with another condition. Twelve genes were categorized as having only a suggested association with a condition. Conclusion: We created a framework for categorizing gene–condition relationships, thereby establishing a method to distinguish between genes proven to cause a condition and genes for which only preliminary evidence suggests an association. Although there can be legitimate benefits to testing a gene before its clinical validity is conclusively established, understanding the rationale for its inclusion on a panel is critical for clinicians. The clinical utility of the findings in any gene ultimately depends on the strength of the evidence linking that gene to disease.

835W

The many faces of *SCN1A***.** *E. Weltmer, A. Bergner, A. Yussuf, Z. Powis, N. Ghahramani, J. Wang.* Ambry Genetics, Aliso Viejo, CA.

    Clinical presentations of *SCN1A*-related disorders can be highly variable, which can prove challenging for diagnostic test selection. We describe several different testing pathways to an *SCN1A* diagnosis and highlight the benefit of having multiple phenotypic-based approaches available. A total of 105 cases from next generation (panel-based) sequencing and 180 cases from exome sequencing with a referral indication of epilepsy, intellectual disability, and/or autism spectrum disorder were reported by our laboratory from October 2015-May 2016. Five cases were positive for an *SCN1A* mutation. We reviewed the clinical phenotypes and testing strategies employed for these five cases. Among the five *SCN1A*-positive cases, four different testing pathways were represented. Case 1 is a toddler with febrile seizures and intractable epilepsy. A nonsense mutation was identified on a targeted 13-gene febrile seizures panel. Case 2 is a toddler with febrile seizures and suspected Dravet syndrome, and case 3 is an adult with intractable epilepsy and intellectual disability. A comprehensive 100-gene epilepsy panel was ordered for both, revealing a nonsense mutation and a missense mutation, respectively. Case 4 is an adult with unspecified seizures and an autism spectrum disorder. A missense mutation was revealed on a broad 196-gene neurodevelopment panel. Whole exome sequencing (~20,000 genes targeted) identified a *de novo* nonsense mutation in case 5, an adult with unspecified seizure disorder, intellectual disability, dysmorphic features, and positive family history. Four different testing pathways were employed, all resulting in a diagnosis of an *SCN1A*-related disorder. Cases 1 and 2 presented as toddlers with febrile seizures, but clinicians opted for different test strategies. Case 2 may have benefitted from a targeted fever panel, resulting in more timely and cost-effective diagnosis. Cases 3 and 4 both presented as adults with somewhat non-specific phenotypes; however, individual features (intractable epilepsy vs. unspecified seizures plus autism) warranted testing via different broad panels. Case 5 presented with the most complex phenotype involving multiple organ systems, and utilized the broadest testing option. Although variability in clinical presentation may be challenging for diagnosis, the availability of a variety of phenotypic-based testing options allows clinicians the opportunity to tailor testing to each patient while not missing an important diagnosis.

836T

A robust, PCR-based assay for detecting *SMN1* copy number variation and performance using three commercial instruments. *M. Cai, L. Jiang, Z. Kabir, S. Paquerault, H. Huuskonen, I. Knight.* Canon U.S. Life Sciences, Inc., Rockville, MD.

Introduction: Spinal muscular atrophy (SMA) is an autosomal recessive neurodegenerative disorder caused by a mutation in the *SMN1* gene on chromosome 5. About 95% of SMA is due to a *SMN1* deletion, resulting in *SMN1* copy number variations (CNV) of 2 copies, 1 copy, or no copies per genome. Rapid detection of *SMN1* CNV is critical for the diagnosis of SMA. We developed a fast and easy analysis approach for *SMN1* CNV detection using a duplex PCR assay with high-resolution melting (HRM) analysis. **Methods**: Three different duplex PCR assays were designed and then evaluated using genomic DNA samples with 2, 1, or no *SMN1* copies per genome. Each duplex assay contained a pair of primers for the *SMN1* sequence and another for the ultra-conserved element (UCE) sequence that contains 2 identical copies per genome. The UCE serves as an internal control to normalize the PCR reactions, which is essential for melting curve analysis. Using deoxynucleotides (dNTP) as the limiting reagent of the PCR reaction, the duplex PCR and HRM analysis were performed on three commercial HRM-enabled thermocyclers. Analysis was performed using each instrument's corresponding software package. During data analysis the melting temperature range was selected to include both the *SMN1* and UCE melting domains. The melt curves were normalized, and the software automatically clustered the results. A difference plot was used to determine the *SMN1* copy number for each sample, and the one copy *SMN1* curve was manually selected as the baseline for the difference curve. Accordingly, the two-copy *SMN1* curves were those above the baseline, and the zero-copy *SMN1* curves were those below the baseline. **Results**: The *SMN1* copy number was accurately determined, with respect to the documented copy number, for the genomic samples isolated from Coriell cell lines. The CNV results were replicated using all three instruments. Analysis of the melting curves using difference plots clustered the results in three distinct groups according to their 2, 1, or no *SMN1* copy number. **Conclusion**: We accurately detected *SMN1* CNV by developing a robust, single-tube, duplex PCR assay combined with HRM analysis. The assay performed consistently using three different commercial instruments. Our study demonstrates that using the difference plot available on each commercial software package is a simple solution to detect *SMN1* CNV.

837F

Association of *ELAVL2* and *PTPRD* genes with epilepsy in purebred Chinese Crested dogs. *B.C. Ballif[1], A.L. Siniard[2], I.S. Piras[2], M. Krug[1], V. Zismann[3], J. Trent[3], M.J. Huentelman[2], L.G. Shaffer[1,4].* 1) Paw Print Genetics, Genetic Veterinary Sciences Inc, Spokane, WA; 2) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 3) Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix, AZ; 4) School of Molecular Biosciences, College of Veterinary Medicine, Washington State University, Pullman, WA.

We performed a genome-wide case-control study in a sample of epileptic Chinese Crested dogs and matched controls to investigate the contribution of genetic variants to the high epilepsy risk in the breed. The genotyping was performed using the Canine HD Whole-Genome Genotyping BeadChip (Illumina). After quality controls, we obtained a dataset of 27 cases and 38 controls (Call Rate > 80%) for a total of 106,392 informative autosomal SNPs. Association analysis at the SNP level was conducted using PLINK running a logistic regression including age, gender and correcting for population stratification. None of the SNPs analyzed reached genome-wide significance (P = 4.7E-07). However, at nominal significance levels we observed 7, 15 and 2 SNPs with P < 0.001 in the additive, dominant and recessive models, respectively. Among the top genes observed were *KHRDBS3*, *ELAVL2*, *TCF25*, *DEF8*, *PTPRD*, *RNF25*, *GALNT10*, *DPYD*, *ATAD2* and *WDYHV1*. The association test was rerun for the SNPs located in the most potentially biologically relevant genes for epilepsy (*KHRDBS3*, *ELAVL2* and *PTPRD*) with additional samples not included in the GWAS due to call rate, for a total of 4 or up to 9 additional samples depending on the SNP. The results confirmed the significance for the SNPs located in *PTPRD* and *ELAVL2* (P = 9.9E-05 and P = 9.9E-05, respectively). The results for the variant located in *KHRDBS3* were not considered due to the low quality of the additional samples for that particular SNP. Our preliminary results suggest the involvement of variants located in *ELAVL2* and *PTPRD* with epilepsy in the Chinese Crested dog. *ELAVL2* is highly expressed in hippocampal CA3 pyramidal neurons and hilar interneurons. It has also been reported in animal models that ELAVL2 protein level in the hippocampus is acutely down regulated after a kainic acid induced seizure. *PTPRD* has been identified by GWAS of 889 newly treated patients to be significantly associated with remission of seizures after starting treatment. We also observed 2 SNPs located in *ADAM23* (P = 0.0012), a gene previously associated with idiopathic epilepsy in 4 different dog breeds. Due to our limited sample size, the results are not significant at the genome-wide level after correction for multiple testing. However, the biological significance of some of these genes justifies a potential role in epilepsy in the Chinese Crested dog. A larger sample size is being used to help to confirm the results.

838W

Motor neuron disease associated genes: Next generation target resequencing in sporadic and familial ALS cases. *C. Gellera[1], V. Pensato[1,2], B. Castellotti[1], S. Magri[1], F. Taroni[1], C. Mariotti[1], E. Dalla Bella[2,3], G. Lauria[2,3].* 1) Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS - Istituto Neurologico Carlo Besta, Milan, Italy; 2) Motor Neuron Diseases Centre, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy; 3) Unit of Neurology III, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy.

ALS is an adult-onset, progressive, neurodegenerative disease caused by the selective loss of motor neurons with a large prevalence of sporadic cases (SALS). The pathogenesis is not completely known, but genetic factors play a major role in the disease. Although mutations in several genes have been described, with a major genetic cause in the repeat expansions in the C9orf72 gene, the underlying cause is still unknown in about 30% of familial ALS (FALS) and in the large majority of SALS (80%). In order to better define the genetic contribution in SALS and to fill the genetic gap in FALS, we set up a NGS approach, using the Illumina technology and the MiSeq apparatus, for a customized analysis for the screening of 78 ALS and ALS-like syndromes causative genes. We selected 213 consecutive ALS cases (37 FALS and 176 SALS), recently referred to our Institute for clinical and genetic definition of the disease. Diagnosis of ALS was made according to the El Escorial revised criteria (Brooks et al. 2000). They were negative for *SOD1*, *FUS*, *TARDBP* and *C9orf72* gene mutations at the preliminary molecular screening. Data analysis showed that 139 patients were negative for variants (filtered for MAF > 1%). The remaining cases (n=74) presented at least one variants, in heterozygous form. In particular, we have identified 7 patients carrying pathogenetic mutations in *VCP* (p.Arg155Cys; p.Arg93Cys), *OPTN* (p.Leu304Phe; p.Gln314Leu; p.Ala481Val) and *TRPV4* (p.Arg269His; p.Asn833Ser). We also report the identification of 56 novel variants of possible pathogenic significance and 16 extremely rare variant (MAF< 0.05%), whose pathogenetic significance should be assessed, and currently classified as VUS. Moreover, we have identified a number of variants in heterozygous form in genes associated with a recessive trasmission of the trait and thus considered as non causative. This study allowed the identification of causative mutations in three genes in 7 patients, and the identification of extremely rare variants or novel possibly causative variants in a significative number of patients (n=67; 31%).

839T

Clinical exome sequencing for identification of undiagnosed genetic disorders. *Y. Ko[1], H. Lee[1], H.D. Kim[2], J.S. Lee[2], C. Lee[1], J.S. Lee[1].* 1) Department of Clinical Genetics, Department of Pediatrics, Yonsei University College of Medicine, Ph.D; 2) Department of Pediatrics, Division of Pediatric Neurology, Yonsei University College of Medicine.

Recent advances in next-generation sequencing (NGS) technologies make it possible to comprehensively explore and identify disease-associated variants and genetic variants that could be associated with disease susceptibility. Here, we investigated the utility of NGS for the diagnosis of 196 patients with unknown genetic origin and molecular characterization. Intractable epilepsy or seizure like phenotype (60% of patients) is the most prevalent type in our cohort. Our pipeline for the analysis of disease-associated variants, could successfully identified the definite genetic causes as well as likely pathogenic variants. In this study, we utilized the TruSight One sequencing panel providing comprehensive coverage of < 4,800 clinically known disease-associated genes. Sequencing was performed using the Illumina MiSeq. The sequenced reads were mapped to the human reference (UCSC hg19) with Burrows-Wheeler Aligner (BWA), and variants were identified with the Genome Analysis toolkit (GATK). Molecular diagnoses could be made for 110 patients (56.1%), which reveals higher diagnostic yield than other study. All mutations have been confirmed by conventional Sanger sequencing and validated by parent testing. Especially, significantly higher diagnostic yield (59.3%) for seizure patients was observed. 50% of confirmed variants are predicted as functionally damaging or deleterious. Many pathogenic variants including known SCN8A (p.Glu710Lys) and SCN2A (p.Thr435Pro) genes are identified as genetic origin for the epilepsy. However, novel pathogenic variants associated with GRIN2A, STXBP1, and CDKL5 were also identified. In this study, we showed that targeted NGS has been successfully applied as a diagnostic tool for patients with undiagnosed genetic disorders. We conclude that NGS are likely to be a very useful in a diagnostic setting.

840F

SMN **copy number analysis: A comparison of three different methodologies.** *P.S. Lai[1], O.S. Yim[1,2], S. Marrs[3], J. Regan[3], P.S. Low[1].* 1) Dept Pediatrics, National Univ Singapore, Singapore, S128044; 2) Dept Psychology, National Univ Singapore, Singapore, S117570; 3) Digital Biology Center, Bio-Rad Laboratories 5731 West Las Positas Blvd, Pleasanton, CA 94588, USA.

Spinal muscular atrophy (SMA) is a neuromuscular disorder characterized by loss of motor neurons in the anterior horn of the spinal cord. The survival motor neuron (*SMN*) genes that influence SMA status are present on 5q13 as two homologous copies, a telomeric *SMN1* gene and a centromeric *SMN2* gene. Homozygous deletion of *SMN1* is responsible for more than 95% of all SMA cases. *SMN2* gene copy number can influence disease severity although it cannot compensate for *SMN1* loss. Thus, determining the *SMN* copy number is important for clinical diagnosis and prognosis, and the most commonly used molecular methods for this screening are multiplex ligation probe amplification (MLPA) and real-time PCR (qPCR). In this study, we set out to compare a new technology, droplet digital PCR (ddPCR), with MLPA and qPCR for *SMN* copy number determination. To do so, we tested 44 subjects for *SMN1* and *SMN2* copy number using a commercial assay for MLPA, previously validated home-brew qPCR assays, and newly commercialized assays for ddPCR. ddPCR and MLPA were found to be concordant in 43 out of 44 samples, and qPCR analyses were discordant with both MLPA and ddPCR in 8 and 10 samples for *SMN1* and *SMN2*, respectively. Two samples initially gave conflicting ddPCR results for *SMN2* copy number and upon reanalysis, only one of them remained discordant which was with the qPCR results. There was also one unique sample for each gene in which ddPCR showed discordant results with both MLPA and qPCR. ddPCR displayed tight coefficient of variation (CV), as only 9% of *SMN1* samples and 16% of *SMN2* samples had CVs higher than 5%, with the maximum CVs being 15.2% and 10.1%, respectively. In conclusion, we found ddPCR to accurately measure 0 – 4 copies of both genes. The methodology is easy to perform and the analysis is straightforward. Furthermore, its high-throughput nature makes it amendable for not only diagnosis and disease characterization, but also for carrier screening.

841W

A NGS strategy for molecular diagnosis of Duchenne/Becker muscular dystrophy in patients without gross DMD gene rearrangements. *O.P. Ryzhkova, N.V. Komarova, A. Loginova, A.V. Polyakov.* Research Centre for Medical Genetics, Moscow, Russian Federation.

Duchenne and Becker muscular dystrophies (DMD/BMD) are the most common inherited neuromuscular disease. The genetic diagnosis is not easily made because of the large size of the dystrophin gene, complex mutational spectrum and high number of tests patients undergo for diagnosis. Multiplex ligation-dependent probe amplification (MLPA) has been used as the initial diagnostic test of choice. Although MLPA can diagnose 70% of DMD/BMD patients having deletions/duplications, the remaining 30% of patients with small mutations require further analysis, such as Sanger sequencing. We worked out a new system to analyze of "point" DMD mutations. This method based on AmpliSeq technology with the addition of fragmentary analysis step between multiplex PCR and sequencing. For this purpose we used specially designed primers including the fragment complementary to fluorescently labelled universal primer. The fragmentary analysis made possible to achieve the necessary level of quality and quantity of each amplicon. This added stage helped us to equalize coverage between some regions. We use the system described above by analysis of 97 patients with DMD/BMD without gross DMD gene rearrangements. Forty-two patients (43%) had LOF-mutations in DMD gene. Previously described missense-mutation was found in one case. A total of 43 variants were Sanger sequenced and all were confirmed to be real (100%). Of these mutations 41% were nonsense-mutation, 23% small insertion, 20% small deletion, 14% splice-site mutation and only one (2%) was confirmed disease-causing missense-mutation. The most of them were not previously reported.

842T

Two novel panels for detection of genetic determinants of epileptic diseases. *C. Clear, R. Hrdlickova, J. Nehyba, C. Firmani, D. Fox, A. Lewis, M. Toloue.* Bioo Scientific, Austin, TX.

   Epilepsy is the most common neurological disease, affecting over 50 million people worldwide. It is characteristized by recurrent seizures due to abnormal neuronal firing in the brain. It is estimated that over 70% of these conditions have genetic background and are associated with inherited mutations in one of several genes encoding proteins involved in neural signalization. Next generation sequencing is a powerful tool for the precise identification of mutations which are associated with specific diseases. Targeted sequencing further economizes the diagnosis by focusing on genes known to be associated with an observed phenotype. Bioo Scientific developed two sequential panels for the detection of epilepsy. The first panel detects the sequences of two neuronal voltage-gated sodium channel genes, *SCN1A and SCN1B*, and the ligand gated gamma aminobutyric acid (GABA) receptor gene, *GABRG2*. The mutations found in these genes are the most frequent cause of early epileptic encephalopathies and have been associated with a spectrum of phenotypes including Dravet Syndrome, Generalized Epilepsy with Febrile seizure (GEFS+), Borderline Severe Myoclonic Epilepsy and other epileptic disorders. The second panel identifies the sequences of another two genes which encode the subunits of voltage-gated sodium channels *SCN2A* and *SCN9A*. The mutations in *SCN2A* are associated with benign neonatal and infantile seizures (BFNIS), but also with other syndromes such as autism, intellectual disability, and schizophrenia. The mutations in *SCN9A* were identified in *SCN1A* negative Dravet Syndrome patients. These amplicon panels cover 100% of the coding regions of *SCN1A, SCN1B, GABRG2, SCN2A,* and *SCN9A* genes and are compatible with Illumina (including compatibility with the MiniSeq) and Ion Torrent sequencing. These panels have been validated with a unique library preparation protocol which only takes 3 hours. Both panels are 100% uniform, which allows for the economical multiplexing of hundreds of libraries in one run by using the accompanying set of 384 barcodes. The development of these two economical, flexible, and well-designed panels for sequential detection of mutations associated with common neurological diseases will serve as a powerful tool to further research the genetic determination of these disorders.

843F

Diagnostic targeted next generation sequencing in patients with epilepsy. *Y.J. Vos[1], P. Rump[1], D.R.M. Vlaskamp[1,2], P.M.C. Callenbach[2], C.M.A. van Ravenswaaij[1], R.J. Sinke[1], O.F. Brouwer[2].* 1) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 2) University of Groningen, University Medical Center Groningen, Department of Neurology, Groningen, the Netherlands.

   Introduction: Epilepsy has a heterogeneous aetiology with a substantial genetic component. We used a targeted next generation sequencing (NGS) gene panel for clinical genetic testing in a heterogeneous cohort of 125 patients with epilepsy. **Materials and Methods:** A gene panel, based on Agilent Sure Select Target Enrichment®, including 144 genes and validated for diagnostic purposes, was used. Within this panel nine partly overlapping subpanels were designed based on different epilepsy phenotypes, namely: benign familial neonatal/infantile epilepsy (5 genes), epileptic encephalopathy (39 genes), focal epilepsy (12 genes), fever related epilepsy (12 genes), progressive myoclonic epilepsy (19 genes), metabolic disorders with epilepsy (43 genes), generalized epilepsies (14 genes), epilepsy in combination with paroxysmal disorders (11 genes), and syndromes with epilepsy and intellectual disability (63 genes). Based on the phenotype of the patient one or more of these subpanels were analysed. **Results:** A (likely) pathogenic variant was identified in 12% of the patients. All confirmed by Sanger sequencing. The yield was highest in patients with benign familial neonatal/infantile epilepsy (23%) followed by fever related epilepsy (17%). The genes involved were: *CACNA1A, CDKL5, GRIN2A, KCNQ2, KCNQ3, KCNT1, MECP2, NHLRC1, PCDH19* (n=3)*, PRRT2, RANBP2* (n=2)*,* and *SYNGAP1.* Using this approach we were also able to detect a mosaic pathogenic variant in the *PCDH19* gene in a male patient. A variant of unknown significance was reported in 12.8% of the patients. In most instances clinical relevance could be excluded. **Conclusion:** Our results demonstrate that targeted next generation sequencing offers a quick, comprehensive and efficient molecular screening in patients with epilepsy.

844W

Early infantile epileptic encephalopathy (EIEE) in females due to an X-linked *ALG13* variant: 4 new cases and a review of the literature. *S.A. Wong, S.L. Klemm, J.L. Schuette, D.J. Cardeiro, C.M. Stanley, N.M. Le. Courtagen Life Sciences, Woburn, MA.*

The *ALG13* gene, along with *ALG14*, encodes a subunit of UDP-GlcNAc transferase, which catalyzes a key step in endoplasmic reticulum N-linked glycosylation. This gene has been previously associated with a severe seizure phenotype in a male patient associated with an N-linked congenital disorder of glycosylation, CDG type1s (Timal et al., 2012) and intellectual disability in four brothers from a single family (Bissar-Tadmouri et al., 2014). The *ALG13* c.320A>G (p.N107S) variant has been reported in only female patients in the literature, with five cases to date discovered through exome sequencing and all of which were *de novo* mutations (Smith-Packard et al., 2015, Dimassi et al., 2016). One of these patients had biochemical testing, which showed a normal glycosylation pattern. We ascertained four additional female patients with this variant via a 471-gene NextGeneration sequencing panel, epiSEEK*ℝ*. The patients ranged in age from 16 months to 13 years. We observed the frequency of this variant to be approximately 0.34% of our tested population (4/1166). The variant met ACMG-based criteria for being pathogenic. Targeted parental testing was performed for two of the four cases, and both variants were *de novo,* consistent with the previous reports. Some patients had additional variants of uncertain significance (*RAI1, PRICKLE1, PRICKLE2, KMT2D, SETBP1, PTCH1* and *NOTCH3*). Of the three cases with detailed clinical information, all had early onset infantile spasms, epileptic encephalopathy, and developmental delay. The 13-year-old patient's spasms had progressed to generalized seizures; she also had dysmorphic features and microcephaly. As this variant was not identified in any male patients, it is possible that it is sex-limited, or as previously suggested in other papers, exhibits a dominant-negative effect. Isoelectric focusing results were not available to us, so it is unclear if these patients exhibited an abnormal profile. *ALG13* is present in several commercially available CDG gene panels and in select epilepsy gene panels. In females with severe early onset infantile spasms and global delay, testing of *ALG13* should be considered via gene panels, or if identified by whole exome sequencing, considered causative/pathogenic, as this variant may be more common in these patients than previously expected.

845T

Clinical detection of pathogenic copy number variation in epilepsy-related disorders. *H. Yang, M.B. Stosser, A. Lindy, O. Cano, D. McKnight. GeneDx, Inc., 207 Perry Parkway Gaithersburg, MD 20877.*

Copy number variants (CNVs) have been reported to play an important role in epilepsy; studies have demonstrated that CNVs account for 5-9% of pathogenic variants in epilepsy-related disorders. We conducted a retrospective study to assess the frequency and type of pathogenic CNVs detected in patients with epilepsy-related disorders who had undergone targeted multi-gene testing. Exon-level array CGH and next generation sequencing of up to 70 genes associated with epilepsy-related disorders was performed on 9314 samples. In 1442 positive cases, pathogenic CNVs were identified in 9% (n=136/1442). Of the cases found to have a single pathogenic CNV in a recessive gene, 4% (5/136) were also found to have a sequencing variant in the same gene. Additionally we reported single pathogenic CNVs in recessive genes that may or may not contribute to the phenotype (n=31). The amount of exomic sequence included in each CNV ranged from one or two exons (30%, 51/167) or 3 or more exons (27%, 45/167) to whole gene (18%, 31/167) or multiple genes (24%, 41/167). CNVs in positive cases were most frequently reported in the following genes: *CDKL5* (10%, 13/136), *SCN1A* (7%, 9/136), *KCNQ2* (6%, 8/136) *TSC2* (8%, 11/136), *PRRT2* (7%, 9/136), *STXBP1* (4%, 6/136), *GRIN2A* (4%, 5/136), *CLN3* (3%, 4/136), *MEF2C* (3%, 4/136), *PCDH19* (2%, 3/136), and *TSC1* (2%, 3/136). Multiple gene CNVs were also reported in positive cases, and often involved genes known to be part of microdeletion or duplication syndromes such as *UBE3A* (15%, 20/136), *KCNQ2/CHRNA4* (5% each, 7/136), *CHRNA7* (5%, 7/136), *SCN1A/2A* (4% each, 4/136). These microdeletion/duplication syndromes were in previously reported, recurrent CNV hotspots such as the 15q11.2 hotspot for Prader-Willi/Angelman syndromes (*UBE3A*), 16p11.2 region that contains *PRRT2* and *CLN3*, and the 15q13.3 region that contains *CHRNA7*. Additionally, the majority of positive findings reported for the *UBE3A* (60%, 20/33), *CHRNA7* (100%, 7/7), and *MEF2C* (57%, 4/7) genes were CNVs and not sequence based changes. Some CNVs were identified in genes not previously reported to have exon-level deletions/duplications (*GABRG2, MFSD8, PNKP,* and *TPP1,* n=5/136). Partial gene CNVs, which made up 57% (96/167) of the CNVs detected in this cohort, may not be detected by whole genome microarrays and could be missed by that technology. These data demonstrate the utility of exon level array technology as part of a comprehensive genetic analysis for patients with epilepsy-related disorders.

846F

NF1 gene deletion in neurofibromatosis type I patients from Egypt. *O. Eid[1], M. Eid[1], A. Kamel[1], G. El-Kamah[2], G. Abdel-Salam[2].* 1) Human Cytogenetic Department, National research centre - Egypt, Giza, Egypt; 2) Clinical Genetics Department, National research centre - Egypt, Giza, Egypt.

   Neurofibromatosis type I (NF1) is the most common hereditary neurocutaneous disorder. Its incidence is 1 in 3,000 to 4,000 births. It is an autosomal dominant disorder characterized by café au lait spots, neurofibromas, Lisch nodules, freckling of the axillary and inguinal regions, optic nerve gliomas and bone dysplasia. It is associated with an elevated risk for malignant tumors of tissues derived from neural crest cells. NF1 is diagnosed on the basis of clinical criteria, while identifying the genetic background of the disease is important mainly for genetic counseling and malignant predisposition. NF1 is caused by defects in the *NF1* gene, a tumor suppressor gene, located at 17q11.2. Point mutations in the *NF1* gene found in 90-95% of NF1 patients, while deletion/duplication found in 5-10% of the patients. Whole *NF1* deletions tend to cause a more severe phenotype. Here, we report the clinical and molecular cytogenetic findings of 28 patients, 16 unrelated patients and 12 patients from five families with NF1 from Egypt. Using FISH technique, *NF1* gene deletion were found in 17.9% of the patients. We also evaluated the frequency of *NF1* gene deletion, emphasizing the importance of deletion analysis in NF1 patients. A more sensitive genetic test such as MLPA, is necessary in order to detect partial *NF1* gene deletion that couldn't be detected by FISH.

847W

Diagnosis of spinal muscular atrophy: A simple method for quantifying the relative amount of *SMN* using Sanger DNA sequencing. *Y. Cao, W. Zhang, Y. Qu, J. Bai, Y. Jin, H. Wang, F. Song.* Medical Genetics, Capital Institute of Pediatrics, Beijing, China.

   Background: Spinal muscular atrophy (SMA), a common genetic disease in childhood, is caused by deletion or mutations in the survival motor neuron 1 gene (*SMN1*). Therefore, *SMN1* genotyping is very important for SMA diagnosis. **Aims:** We developed a novel method to simultaneously screen for *SMN1* homozygous deletion and *SMN1* point mutations. **Methods:** This relative quantifying method is based on PCR and Sanger DNA sequencing. *SMN1* and *SMN2* were simultaneously amplified by the same pair of primers, serving as an internal reference for each other to indicate the relative amount, as observed from the base peaks on the Sanger sequencing chromatogram at differential sites between genes. The accuracy, effectiveness and reproducibility of our method were evaluated using recombinant plasmids with various ratios of *SMN1* to *SMN2*, SMA patients' samples with known *SMN* genotype, respectively. Moreover, 223 SMA patients and 216 healthy controls were screened in parallel, using our method in comparison with multiplex ligation-dependent probe amplification (MLPA). **Results:** Our method showed experimental sensitivity and specificity of 98.6% compared with MLPA. And ROC curve analysis indicated high consistency between the results of two methods ($P < 0.001$). In addition, among 10 patients with diagnostic *SMN1* heterozygous deletion, nine *SMN1* point mutations were identified. **Conclusions:** Under certain conditions such as *SMN1* and *SMN2*, Sanger DNA sequencing can be used in relative quantification analysis. Based on this peculiarity, we developed a relative quantifying method not only for *SMN1* homozygous deletion but also for *SMN1* point mutations.

848T

Detection of copy number and structural variation in the clinical lab using massively parallel sequencing. *J.M. Devaney, B.W. Meltzer, M.B. Seprish, S.L. Everhart, J.M. Campos, K.P. Cusmano-Ozog, S.E. Hofherr.* Laboratory Medicine, Children's National Health System, Washington, DC.

   Statement of Purpose: Massively parallel sequencing (MPS) is offered as a clinical tool to discover genetic variation including single nucleotide variants and small insertions or deletions. Recently, the significance and frequency of pathogenic intragenic deletions or duplications (copy number variation) and large structural variants have become increasingly evident. We sought to validate the use of the Clinical Exome (4,813 genes) combined with bioinformatics tools to discover CNVs and SVs. **Methods:** The genomic DNA was prepared for targeted sequencing using the TruSight One (TSO) panel that contains the amplification, amplicon enrichment, and indexing for samples. The samples were analyzed on a NextSeq 500 system (Illumina, USA) using a High Output Kit v2 (300 cycles). Using BaseSpace Sequence Hub, we performed alignment to the hg19 reference genome with Burrows-Wheeler Aligner (version 0.7.7) with the Genome Analysis Toolkit (GATK v1.6) to generate BAM files. The BAM files were analyzed with the NextGENe software v2.4.0.1 (SoftGenetics, USA) for exon level copy number variation (CNV) detection. CNVs are detected by comparing the coverage (Reads Per Kilobase per Million mapped reads) of specified regions in a "sample" project and a "control" project to provide a coverage ratio. A beta-binomial model is fit to the coverage ratio in order to model the amount of dispersion (noise). Likelihood values are calculated based on the dispersion measurements and coverage ratios. Using a Hidden Markov Model, the probabilities are used to CNV classifications for each region as Normal, Duplication, or Deletion. We examined samples that were previously analyzed on the Affymetrix CyctoScan DX system (35 samples) and samples from Coriell Institute for Medical Research that contain CNVs (15 samples). In addition, we confirmed two samples with CNVs called from MPS data on the Affymetrix system. **Results and Conclusions:** We were able to use the NexGENe software to call single exonic deletions, exonic duplications, partial gene deletions and duplications, gene deletions, gene duplications, and two different trisomy conditions. The sensitivity was 100% with detection of all known deletions and duplications. The analysis requires multiple samples of the same sex as controls for one sample and all samples must be part of the same run. Overall, this tool was successful for the calling of CNVs and we are in the process of making it part of our clinical pipeline for MPS.

849F

Expanding the genomic search for etiology of the congenital microcephaly in Brazil. *E.A. Zanardo[1], J.G. Damasceno[1], A.M. Nascimento[2], M.M. Montenegro[1], G.M. Novo-Filho[1], R.L. Dutra[1], F.A.R. Madia[2], Y.G. Oliveira[1], T.M.M. Costa[1], A.T. Dias[1], C. Milani[2], F.B. Piazzon[1], M.I. Melaragno[3], C.A. Kim[4], L.D. Kulikowski[1,2,5].* 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; 2) Laboratório de Citogenômica, Departamento de Pediatria, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; 3) Departamento de Morfologia e Genética, Universidade Federal de São Paulo, São Paulo, Brazil; 4) Unidade de Genética, Departamento de Pediatria, Instituto da Criança do HC/FMUSP, São Paulo, Brazil; 5) Centro de Reprodução Humana e Genética, Departamento de Saúde Coletiva, Faculdade de Medicina do ABC, Santo André, Brazil.

   Congenital Microcephaly (CM) is a birth defect associated with small head circumferences that result in severe neurological consequences. CM seems to arise from a combination of various genetic and environmental factors that can alter the developing brain, resulting in structural and functional abnormalities, however most cases of clinically diagnosed microcephaly remain unexplained. Recent Brazil is facing a disturbing increase in cases of CM which coincided with Zika virus (ZIKV) infections in mothers during pregnancy. Zika is a neurotropic virus that can lead to microcephaly still in the placental environment and is suggestive of a possible relationship of birth defect rise in Brazil. Although Infections are a known cause of microcephaly, most of the reports are based on circumstantial evidences. Systematic investigations in congenital brain malformations using array analysis have implicated copy number variations (CNVs) in their etiology. In the past years, about several different genomic syndromes were described linking congenital microcephaly with pathogenic CNVs suggesting that other factors than ZIKV could be responsible for CM. Thus in this study, we evaluated 109 patients with multiple congenital abnormalities and developmental delay without genetic conclusive diagnostic using different genomic array platforms. One-third (29,4%) of patients studied (32/109) presented congenital microcephaly, among them about 46,9% (15/32) showed pathogenic CNVs that were not found in patients without microcephaly phenotype. The genotype-phenotype relationship between CM and specific pathogenic CNVs is still unclear in the literature and insufficient notified in Brazil. Also the array analysis in these patients may possibly identify other risk factors to the clusters of microcephaly and neurological disorders and help us better understand the relationship between ZIKV and the rise of detected cases of congenital malformations in this moment. Grants: FAPESP: 14/02565-8 and FINEP-CT INFRA 0160/12 SP8.

850W

Molecular combing reveals structural variations in the Spinal Muscular Atrophy locus in African-American population. *M. Pierret, S. Berthoumieux, A. Thomas, C. Chypre, S. Barradeau, A. Bensimon.* Genomic Vision, Bagneux, France.

The Spinal Muscular Atrophy (SMA) is an autosomal recessive motor neuron disease, which is the most common genetic cause of infant death, due to deletions/mutations in the *SMN1* gene. Improvement of the detection of SMA carriers is important in genetic counseling, especially in African-American population in which undetectable carriers are particularly frequent. The *SMN1* gene, and its homologous *SMN2* gene are localized on chromosome 5q13.2 in a complex region characterized by an inverted duplication of around 500 kb sequence. However, the precise mapping of this locus is extremely difficult with the current technologies, such as sequencing or DNA microarray, due to high density of segmental duplications and other structural variations. Molecular Combing allows direct visualization of single DNA molecule, combining constant and uniform stretching of DNA molecules with a unique detection strategy, the Genomic Morse Code (GMC). The GMC, fluorescently provides a specific color-coded pattern for the direct and high resolution visualization of loci of interest. In order to precisely characterize the *SMA locus*, we developed a specific GMC that cover the entire *SMA locus* over 2 Mb. This GMC was hybridized on combed genomic DNA extracted from amnyocyte-derived cell cultures from African-American individuals. The image acquisition of fluorescent array signals is performed using an automated epifluorescence microscope, FiberVision®. After acquisition, SMA fluorescent array signals are pinpointed by the dedicated FiberStudio® software. The alignment of the different fluorescent array signals to the theoretical GMC deduced from the human genome reference sequence (GRCh38/hg38) reveals major discrepancies. First, it appears that the two *SMN* genes are not in a head to-tail orientation as annotated but are in a head-to-head orientation. Moreover, a color pattern from the theoretical GMC was not observed in African-American samples indicating the absence of the corresponding sequence. Finally, we also identified a repeat sequence with a variable number of repeated units located at the telomeric and/or centromeric regions indicating the presence of an unknown copy number variation sequence. In conclusion, Molecular Combing is a powerful technology that allows us to precisely and accurately map the *SMA locus* in the African-American population. This corrected map gives information that would be helpful to develop relevant SMA screening test for this population.

851T

Novel *UBE3A* variant detected in a family with familial Angelman syndrome and osteogenesis imperfecta type I. *E.A. Leeth[1,2], E.M. McGinnis[1], D.A. Kirschmann[1], C. McCabe[1], E. Caparelli[1], V.R. Sanders[1], J. Charrow[1,3], K.M. Leuer[1,2], L.J. Jennings[1,2].* 1) Ann and Robert H Lurie Children's Hospital, Chicago, IL; 2) Pathology and Laboratory Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Pediatrics-Genetics, Feinberg School of Medicine, Northwestern University, Chicago, IL.

Approximately 10% of individuals with Angelman syndrome (AS) possess causative loss of function variants in the maternally derived E6-AP ubiquitin-protein ligase (UBE3A) gene. UBE3A consists of 16 exons where the last two nucleotides of exon 7 and exons 8-16 make up the coding regions of the gene. To date, over 200 pathogenic variants have been reported, spanning across the entire region. We describe a patient who initially presented to our institution for an evaluation due to a maternal family history of osteogenesis imperfecta (OI) type I. Analysis of the COL1A1 gene identified the familial mutation in the proband. Subsequently, developmental delay and seizures prompted a neurological evaluation. Analysis of the SNRPN region showed normal methylation. A comprehensive epilepsy panel revealed a novel UBE3A missense variant in the proband, his mother and affected maternal half-sibling. The c.1763G>T (NM_000462.3) variant is located in exon 9 and codes for an amino acid in the C-terminal HECT domain, which plays an essential role in transferring the ubiquitin to proteins targeted for degradation. Although two nearby splicing pathogenic variants have been published, to our knowledge, no missense variants at this location have been reported. Additionally, this variant is absent from population databases and several prediction algorithms suggest a deleterious or possibly damaging effect. Based on family segregation analysis, we currently interpret the c.1763G>T variant as likely pathogenic. This case example describes a novel likely pathogenic UBE3A variant as well as adds to the literature indicating the importance of screening UBE3A for causative variants in patients with suspected AS. Sadikovic et al. (2014) suggests that approximately 29% of causative variants identified in UBE3A are inherited, which poses a significant risk of recurrence for families and highlights the need for genetic counseling. Genetic counselors can help educate providers on the various genetic etiologies of AS as well as the importance of screening UBE3A. They can also encourage patients and their families to participate in additional family studies when a novel UBE3A variant is identified. Additionally, the finding of two familial genetic disorders, OI type I and AS, also emphasizes the importance of continued genetic evaluations and testing even in light of an established genetic diagnosis.

852F

Expert recommendations for the laboratory diagnosis of neuronal ceroid lipofuscinosis type 2 (CLN2 disease): Diagnostic algorithm and best practice guidelines for a timely diagnosis. *N. Miller[1], M. AlSayed[2], D. Burke[3], J. Cohen-Pfeffer[1], J.D. Cooper[4], L. Dvorakova[5], R. Giugliani[6], E. Izzo[1], H. Jahnova[5], Z. Lukacs[7], S. Mole[8], I. Noher de Halac[9], D. Pearce[10], A. Schulz[11], N. Specchio[12], W. Xin[13], M. Fietz[14].* 1) Medical Affairs, BioMarin Pharmaceutical Inc., Novato, CA; 2) Department of Medical Genetics, Alfaisal University and King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 3) Chemical Pathology, Camelia Botnar Laboratories, Great Ormond Street Hospital, London, UK; 4) Institute of Psychiatry, Psychology, & Neuroscience, King's College London, London, UK; 5) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic; 6) Serviço de Genética Médica HCPA, Departamento de Genética UFRGS, Porto Alegre, RS Brazil; 7) Newborn screening and Metabolic Diagnostics unit, Hamburg University Medical Center, Hamburg, Germany; 8) Molecular Cell Biology, MRC Laboratory for Molecular Cell Biology, University College London, London, UK; 9) Universidad Nacional de Córdoba, Facultad de Ciencias Médicas, Córdoba, Argentina; 10) Sanford Children's Health Research Center, Sioux Falls, USA; 11) Department of Pediatrics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 12) Department of Neuroscience, Bambino Gesù Children's HospitalRome, Italy; 13) Neurogenetics DNA Diagnostic Lab, Massachusetts General Hospital, Harvard Medical School, Boston, USA; 14) Department of Diagnostic Genomics, PathWest Laboratory Medicine WA, Nedlands, Australia.

Neuronal ceroid lipofuscinoses (NCLs), a heterogeneous group of lysosomal storage disorders, include the rare autosomal recessive neurodegenerative disorder CLN2 disease (CLN2). CLN2 is due to mutations in *TPP1/CLN2* gene causing tripeptidyl-peptidase-1 (TPP1) enzyme deficiency. Classic late-infantile CLN2 has pediatric onset with initial symptoms of seizures and language delay followed by progressive dementia, motor and visual deterioration and early death. Variant phenotypes occur more rarely. CLN2 diagnosis is based on laboratory testing following clinical suspicion. Early diagnosis is key to optimizing clinical care and future therapies outcomes, yet delays in diagnosis are common due to low disease awareness, non-specific initial symptoms and limited diagnostic testing access in some regions. In May 2015, international experts met to recommend best laboratory practices for early CLN2 diagnosis. When clinical signs suggest NCLs, TPP1 activity should be the first test performed (along with palmitoyl-protein-thioesterase-1 to exclude CLN1). However, since reaching initial suspicion of CLN2 and NCLs is challenging, where available, use of epilepsy gene panels to investigate unexplained seizures in childhood is endorsed. These panels should include *TPP1/CLN2* besides genes for other NCLs lacking biochemical tests. Diagnostic TPP1 enzyme test in leukocytes is well established and robust and in DBS is considered diagnostic if followed by molecular testing. Future methods to measure TPP1 activity via MS/MS may improve DBS-based TPP1 testing sensitivity allowing also future newborn screening.   To confirm clinical suspicion of CLN2, the recommended gold standard for laboratory diagnosis is demonstrating deficient TPP1 activity and detecting causative mutations in each allele of *TPP1/CLN2* gene..

853W

Genetic testing strategies for children and adults with peripheral neuropathy. *T. Brandt, H. Zghal Elloumi, F. Millan Zamora, A. Bamford, K. Retterer.* GeneDx, Inc., Gaithersburg, MD.

Inherited peripheral neuropathies are a genetically heterogeneous group of disorders with considerable phenotypic overlap, making it challenging to choose the most appropriate genetic test. This retrospective study examines the positive cases resulting from two different testing options for patients with peripheral neuropathy in our clinical laboratory: a 50-gene custom capture-based NGS panel (including genes most commonly associated with neuropathy) and whole exome sequencing (WES). Data include results from 311 neuropathy panel tests and 262 WES tests for patients with an indication of peripheral neuropathy. Results are defined as positive if pathogenic or likely pathogenic variants related to the neuropathy phenotype were identified. There was no significant difference in the positive rates for neuropathy cases between the targeted panel testing (70/311, 23%) and WES (46/262, 18%). However, the positive rate of WES was significantly higher for pediatric neuropathy patients (100 patients <18 y.o., 28% positive) compared to adult neuropathy patients (162 patients ≥18 y.o., 11% positive) defined by age at testing (p<0.01). This finding was independent of the WES testing approach, and applied to concurrent parental testing (trio) with 30% (pediatric) versus 13% (adult), as well as to cases where both parental samples were not available, with 19% (pediatric) versus 11% (adult). In contrast, age at testing did not influence the diagnostic rate for panel testing. Half of the WES positive findings in children with neuropathy involved genes not on the neuropathy panel, representing 14% of all pediatric WES neuropathy cases (14/100). Those genes are associated with syndromic disorders with a broad phenotypic spectrum that includes neuropathy. For example, multiple pediatric cases were found to have diagnostic findings for Brown-Vialetto-Van Laere syndrome (*SLC52A2* gene) by WES. In contrast, only 4% (7/162) of adult WES patients with neuropathy had positive findings in genes not on the panel, and each of these seven cases had findings in a different gene. Therefore, the 50-gene panel is well-suited for adult patients with neuropathy, with WES testing offering limited additional gains. On the contrary, as half of the diagnostic findings in children with neuropathy were identified in genes associated with syndromic disorders not typically included on neuropathy panels, WES provides a significant benefit in the molecular diagnosis of childhood neuropathy.

854T

A comprehensive NGS gene panel for the challenging genetic diagnosis of spinocerebellar ataxias and spastic paraplegias. *S. Magri, D. Di Bella, E. Sarto, L. Nanetti, C. Gellera, C. Mariotti, F. Taroni.* Unit of Genetics of Neurodegenerative and Metabolic Disease, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy.

Spinocerebellar ataxias and hereditary spastic paraplegias are genetically highly heterogeneous groups of neurological disorders. Although >50 genes have been identified for each group, >50% of the patients remain undiagnosed. Molecular diagnosis is made difficult also by the clinical overlap with continuity of the phenotypic spectrum (spastic ataxia). Aim of the study was to analyze all the patients referred to our laboratory with two subsequent NGS approaches: 1) Two different TruSeqCustomAmplicon (TSCA, Illumina) panels including 76 genes for both dominant (SCA) and recessive (SCAR) spinocerebellar ataxias, and 50 spastic paraplegia genes; 2) One single probe-based panel (Nextera Rapid Capture, Illumina) containing >200 genes associated with spastic paraplegia (112) or ataxia (139) phenotypes. Coverage analysis revealed a higher sensitivity and uniformity of the probe-based *vs* the amplicon-based strategy (99% *vs* 78% of target region with ≥20X coverage). We analyzed 243 index cases (70-80% sporadic), 120 with spastic paraplegia and 123 with ataxia previously screened for the most frequent forms as appropriate. Overall, pathogenic mutations were identified in 53/243 patients (21.8%), 25/123 ataxic and 28/120 HSP patients. In particular, we identified mutations in challenging genes such as *SYNE1* (5 index cases), *SACS* (2), *SETX* (1), *SPG11* (10) and *PNPLA6* (2) which are difficult to be studied by conventional sequencing because of their length. Moreover, we identified mutations in rare genes such as *SPG8*, *SPG21*, *SPG47*, *SPG48*, *PNKP* (AOA4), *GBA2* (SPG46), *KIF1A* (dominant SPG30), *SPTBN2* (SCA5), and *PRKCG* (SCA14). CNV analysis based on depth-of-coverage comparison allowed us to identify a single exon deletion in *ANO10* (SCAR10) gene and a large deletion (exon 1-43) of *ITPR1* (SCA15). This comprehensive approach allowed also to identify mutations in genes unexpected based on the clinical diagnosis: *ATP13A2* (PARK9), *CYP27A1* (cerebrotendinous xanthomatosis), *EXOSC3* (PCH1B) and *PLP1* (a mutation in a female patient with an SPG-TCC phenotype). These results prompted reevaluation of clinical phenotypes and diagnostic reclassification. The usefulness of this combined ataxia-HSP multiplexed gene panel approach is further demonstrated by the identification of *FA2H* (SPG35) and *AP4B1* (SPG47) mutations in 2 predominantly ataxic patients, and *SYNE1* mutations in a patient with a complex spastic paraplegia phenotype. (E-Rare EUROSCAR and Italian MoH RF-2011-02351165 grants to FT).

855F

HMSN type Lom (*CMT4D*) in a non-Roma Czech boy caused by an isodisomy of the whole chromosome 8. *D. Safka Brozkova[1], M. Krutova[1], M. Trkova[2], R. Valkovicova[1], J. Haberlova[1], P. Seeman[1].* 1) DNA laboratory, Charles University 2nd Medical School and University Hospital Motol, Prague, Czech Republic; 2) Centre for Medical Genetics and Reproductive Medicine GENNET, Prague, Czech Republic.

Inherited peripheral neuropathies (IPN) are a heterogeneous group of disorders characterized by distal weakness and atrophies on lower limbs and also the sensory loss. IPN are genetically extremely diverse with more than 90 genes implicated with the IPN. Therefore the classical methods as Sanger sequencing is useful for the four most frequent causes and massively parallel sequencing (MPS) with panel of all the known IPN genes is suitable for the very many rare causes. Five years old patient with walking problems from two years of age was referred to our clinic from the local pediatrician. The nerve conduction study showed very low nerve conduction velocities (20 m/s), in a range of demyelinating neuropathy. CMT1A/HNPP , *MPZ* and *GJB1* gene were initially tested and not confirmed. MPS revealed two candidate mutations as disease causing – homozygous p.Arg148X in *NDRG1* gene and heterozygous p.Arg113X in *REEP1* gene. The patient is not Roma. Interestingly the *NDRG1* mutation p.Arg148X, in homozygous state responsible for HMSN type Lom in patients of Roma origin, was detected only in father, not in mother. Additional haplotype analysis with use of dinucleotide markers for chromosome 8 and 17 revealed homozygous haplotype inherited only from father, with missing information from mother for chromosome 8 and one haplotype from each parent for chromosome 17. Whole genome genotyping on Illumina Human Cyto SNP chip revealed complete isodisomy of chromosome 8, probably resulting from monosomy rescue. The segregation analysis in family revealed the *REEP1* mutation also in clinically healthy father and grandfather; therefore it was excluded as disease causing. HMSNL in our patient is caused by p.Arg148X mutation occuring in heterozygous state in healthy father. Due to an isodisomy of whole chromosome 8 with the mutated *NDRG1*, the p.Arg148X occurs in a homozygous state in patient. The family is not aware of Roma ancestors, but due to a presence of this mutation the very distant ancestor of Roma origin is probable. Support- AZV 15-31899A.

856W

Exon analysis of 74 genes potentially involved in hereditary spastic paraplegias through targeted next generation sequencing in a large cohort. *G. Stevanin[1,2,3], L. Raymond[1,2], M. Mairey[1,2], S. Morais[1,2,4], L. Parodi[1,2,5], R. Valter[1,2], C-S. Davoine[1], G. Banneau[3], E. Leguern[3], F. Barreca[1,2,5], L. Elsayed[1,2,6], G. Vazza[5], I. Alonso[4], A. Brice[1], A. Durr[1,3].* 1) Institut du Cerveau et de la Moelle, NEB, INSERM 1127, Paris, France; 2) Ecole Pratique des Hautes Etudes, PSL research University, Paris, France; 3) APHP, Department of Genetics, Pitie-Salpetriere Hospital, Paris, France; 4) UnIGENe, IBMC, Porto, Portugal; 5) University of Padova, Department of Biology, Padova, Italy; 6) University of Khartoum, Khartoum, Sudan.

Hereditary spastic paraplegias (HSP) are neurological diseases. They are genetically heterogeneous (59 spastic paraplegia genes, SPG) and are inherited through all known inheritance patterns accounting for a prevalence of 1-5:100 000 in most populations. HSPs are characterized by spastic gait, hyperreflexia, extensor plantar responses and proximal weakness, frequently associated to other neurological and extra-neurological signs. This heterogeneity makes classical molecular tests time-consuming and difficult to interprete, and this prompted us to develop a kit that couples a targeted capture and next generation sequencing to analyse 74 genes potentially involved in HSPs and related diseases. The kit has been used to sequence and analyse 324 index patients (153 French, 98 Portuguese, 32 Italian and 41 patients from Sudan). DNAs libraries were prepared through KAPA library preparation kit (KAPA Biosystems). The panel included the target region probes (1042 regions) corresponding to all coding exons belonging to 74 HSP genes and was designed by Roche-NimbleGen (SeqCap EZ). The DNA sequencing was realized using the MiSeq or NextSeq sequencers (Illumina). Alignment and variants detection were done through the Genomics Workbench software (CLC bio, Qiagen). Variants were filtered and prioritized, then verified through Sanger sequencing on patients and other family members, in order to establish the segregation patterns. A mutation responsible for the disease phenotype was identified in 1/3 of the patients overall analysed. For 1/3 of the patients, no pathogenic variant could be detected. One or more variants with unknown significance was identified among the remaining 1/3 of the cohort; in these cases the actual knowledge regarding the gene involved didn't allow to draw a conclusion regarding a possible pathogenic effect, and therefore its role in HSP onset. In addition, we identified a series of 8 patients carrying heterozygous variants in the motor domain of KIF1A inherited in a dominant pattern or de novo which contrast with the classical recessive transmission mode of this gene in SPG30. Through the coupling of targeted capture and NGS it was possible to increase the molecular diagnostic rate if compared to more classical strategies based on decision trees. Our results therefore show the importance of testing the same set of genes in patients with different transmission patterns, allowing the detection of multiple and unexpected transmission pathways.

857T

Genetic epidemiology of inherited peripheral neuropathy in 481 Japanese pedigrees. *A. Yoshimura, M. Ando, Y. Hiramatsu, J. Yuan, Y. Higuchi, A. Hashiguchi, Y. Okamoto, H. Takashima.* Department of Neurology and Geriatrics, Kagoshima University, Kagoshima, Japan.

Background and Objective Inherited peripheral neuropathies (IPNs) are a group of related diseases primarily affecting the peripheral motor and sensory neurons and have been linked to over hundred genes or loci. These diseases are clinically heterogeneous and mainly comprise of Charcot-Marie-Tooth disease (CMT), hereditary sensory and autonomic neuropathy (HSAN), hereditary neuropathy with liability to pressure palsy (HNPP), and hereditary motor neuropathy (HMN). In this study, we aim to describe the clinical and genetic spectrum of patients with IPNs in Japanese population. **Methods** In the period from October 2014 to March 2016, we collected 481 index cases with IPNs who were referred by neurological and neuropediatric departments throughout Japan. All of these cases had been proved negative of *PMP22* duplication/deletion mutation; cases with HSAN were excluded from this study. We designed an Ion AmpliSeq gene panel encompassing 72 IPN-related genes. All target genes were amplified and subsequently sequenced using Ion Proton system (Life Technologies). Candidate variants were validated by Sanger sequencing, and if available, segregation analysis was performed. **Results** Among 481 cases with IPNs, mutations were detected in 159 cases (33.1%), consisting of 118 pathogenic mutations and 41 likely pathogenic mutations. Of all patients identified with pathogenic mutations, *GJB1* gene mutation was the most common cause in our cohort (29 cases, 24.6%), followed by *MPZ* (22 cases, 18.6%), *MFN2* (21 cases, 17.8%), *BSCL2* (6 cases, 5.1%), *MME* (5 cases, 4.2%), *PMP22* (5 cases, 4.2%). Besides, less than five cases were found harboring mutations in the other eighteen disease-causing or candidate genes, respectively. The mutation detection rate was 17.0% in cases suspected with autosomal recessive (AR) CMT (8/47), 37.9% in autosomal dominant (AD) CMT (44/116), and 21.3% in sporadic CMT (61/287). Mutations in recently described *MME* gene was found to be the primary cause for genetically confirmed axonal AR-CMT (5/9). **Conclusion** Targetted resequencing confirmed the diagnostic yield of the most frequent genetic defects besides *PMP22* duplication/deletion mutation, *GJB1*, *MPZ* and *MFN2*, total of which account for 61.02% of all mutation-positive cases. *MME* is the most frequent cause of AR-CMT2 in our cohort. The mutation detection rate of suspected AR-CMT was relatively lower than AD or sporadic CMT, suggesting more genetic diversity and require further study.

858F

Clinical diversity caused by novel *IGHMBP2* mutations. *J. Yuan[1], A. Hashiguchi[1], Y. Higuchi[1], A. Yoshimura[1], H. Yaguchi[2], K. Tsuzaki[3], A. Ikeda[4], K. Wada[5], M. Ando[1], Y. Hiramatsu[1], Y. Okamoto[1], H. Takashima[1].* 1) Department of Neurology and Geriatrics, Kagoshima University, Kagoshima, Japan; 2) Kashiwa hospital, Jikei University School of Medicine, Chiba, Japan; 3) Department of Neurology, Kansai Electric Power Hospital, Osaka, Japan; 4) Kanagawa Children's Medical Center, Yokohama, Japan; 5) Department of Neurology, Tottori University Faculty of Medicine, Yonago, Japan.

Background and Objective Charcot-Marie-Tooth disease (CMT), as the representative disease of inherited peripheral neuropathies (IPNs), is typically characterized by progressive motor and sensory polyneuropathy. However, a certain CMT disease-causing gene might also produce distinguish phenotypes and therefore clinically diagnosed as other inherited neuropathies. Like *IGHMBP2*, which was linked to both autosomal recessive axonal CMT and spinal muscular atrophy with respiratory distress type 1 (SMARD1). The aim of this study is to identify the genetic spectrum of patients with IPNs in Japan and demonstrate their phenotypic features. **Methods** From June 2014 to December 2015, we collected 408 cases diagnosed with IPNs on the basis of clinical findings and electrophysiological study. We conducted mutation screening by applying an Ion AmpliSeq gene panel, which comprises of 72 disease-causing or candidate genes of IPNs. **Results** We identified novel compound heterozygous or homozygous mutations of *IGHMBP2* in four cases, consisting of p.Ala345Glu and p.Arg595Trp in patient 1, p.Thr115Met and p.Gly399Ser or c.1060+5G>C in patient 2, p.Gln276* and p.Gln568* in patient 3, and homozygous p.Tyr920Cys in patient 4. Segregation study was carried out in pedigrees of patient 1 and 3, confirming that a single mutated allele was inherited from each parent. Parental consanguinity was only recorded in patient 4. Three patients (1, 2, and 4) were diagnosed with CMT, manifesting childhood-onset axonal polyneuropathy. Patient 3 was characterized by low birth weight, weak cry, reduced spontaneous movement, respiratory distress from four months after birth, and thus comparable to SMARD1. **Conclusion** We present the original report of patients with axonal CMT caused by recessive *IGHMBP2* mutations in Japan. In our cohort, the *IGHMBP2* mutation was detected in less than 1% of IPNs. We report four novel mutation genotypes of *IGHMBP2* leading to significant clinical diversity, whereas the interaction between genotype and phenotype remains controversial and requires further research.

859W

Mutational analysis of sequence variants in *GCDH*: Frequency of variants in NBS-positive versus symptomatic patients. *B.D. Apple, H.-C. Yu, N. Meeks, K. Kronquist, M. Woontner, S. Goodman, E. Spector.* Department of Pediatrics, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO.

Glutaric acidemia type 1 (GA1, OMIM 231670) is an autosomal recessive disorder caused by sequence variants in *GCDH*, encoding the protein glutaryl-CoA dehydrogenase (GCDH). GA1 is one of the 31 metabolic disorders included in all newborn screening (NBS) panels. We present 12 years of data from clinical Sanger sequencing in our laboratory. Overall, 1331 probands were analyzed. The ACMG Guidelines for Interpretation of Sequence Variants was used to interpret the pathogenicity of all sequence variants not reported in the literature. In a subset of 341 cases where a reason for testing was provided, the indications included: abnormal NBS (245 or 72%), clinical suspicion of GA1 (64 or 19%) and abnormal biochemical testing in an older patient (32 or 9%). Of the 245 cases in the NBS positive group, 26 possessed *GCDH* variants; 16 were carriers and 5 were homozygotes or compound heterozygotes for a pathogenic variant. This data indicates a yield of 5/245 or 2.0% for a molecular diagnosis of GA1 among NBS positive patients. This detection rate is in line with the high false positive rate of the NBS. Of the patients with a clinical phenotype suspicious of GA1, 4/64 cases were either homozygous or compound heterozygotes giving a 6.2% diagnostic yield. Of the group of older patents with abnormal biochemical testing, 9/32 samples were either homozygous or compound heterozygous, giving a 28.1% diagnostic yield. Among all of the 1331 probands tested we identified three variants known to be associated with a low-excretor phenotype. The variants were: Val400Met (6.6%), Arg227Pro (5.2%) and Met405Val (1.9%). Interestingly, only two individuals possessing one of these low-excretor alleles were detected in the NBS positive group. One hundred and thirty-four variants were identified: 64 were pathogenic, 28 likely pathogenic, 38 VUS, 1 likely benign and 1 benign. Of these 134 variants 48 have not been previously reported. The majority of variants were missense variants (107). In addition to the missense variants, 6 nonsense, 5 splice-site, 10 indels causing a frame shift and 3 in-frame indels were found. No deletions or duplications at the exonic level were identified in 54 patient samples tested by aCGH. Of all the variants (known and novel), almost 70% of them were either pathogenic or likely pathogenic. Our data confirm that *GCDH* is a highly conserved gene and that the majority of new sequence variants identified are either pathogenic or likely pathogenic.

860T

Diagnostic yield for chromosome microarray analysis in patients with autistic features. *A. Bandholz, V. Mehta, Z. Powis, E. Spiteri, M. Rossi.* Ambry Genetics, Aliso Viejo, CA.

Autism Spectrum Disorders (ASDs) are clinically heterogeneous, highly heritable neurodevelopmental disorders seen in 1-2% of children. Chromosomal microarray (CMA) is a first-tier test in the evaluation of individuals with ASDs with the diagnostic yield varying based on population examined, ranging from 5-25% in recent literature. Subjects in the majority of published studies are formally evaluated for ASD prior to inclusion; however, in the clinical setting formal autism evaluations may not occur prior to genetic testing. Herein we calculate the diagnostic rates for copy number variations (CNV) by CMA in a cohort of patients with ASD, pervasive developmental disorder not otherwise specified (PDD-NOS), or autistic features (AF) with or without formal evaluation submitted for clinical testing. The most recent 614 unselected cases referred for postnatal microarray to Ambry Genetics (Aliso Viejo, CA) with indication for testing provided were reviewed. Patients with clinical indications including ASD, PDD-NOS and AF were included in the cohort. Of the 614 cases, 128 (21%) were reported to have ASD/PDD-NOS/AF with or without additional features excluding two patients with prior abnormal diagnostic lab results. For cases with more than one CNV identified, the CNV of highest pathogenicity determined the overall case classification. Of the 128 ASD/PDD-NOS/AF cases, 7 (5.5%) were pathogenic, 2 (1.6%) variant likely pathogenic (VLP), 37 (28.9%) variant of uncertain significance (VUS), 10 (7.8%) variant likely benign (VLB), and 72 (56.3%) normal. Overall, 9/128 (7.0%) patients had a pathogenic or VLP CNV; 6/9 (66.7%) with microdeletions, 1/9 (11.1%) with a microduplication, and 2/9 (22.2%) with sex chromosome aneuploidy. Recurrent CNVs demonstrating reduced penetrance and variable expressivity were frequent amongst the pathogenic and VLP cases including single patients with 15q11.2 BP1-BP2 microdeletion, 16p11.2 microdeletion, 15q13.3 microdeletion, distal 1q21.1q21.2 microduplication, and intragenic *NRXN1* deletion. Among VUS CNVs, two autism candidate genes were noted, *CHRNA7* and *CNTN6*. The majority of the pathogenic/VLP results (67%) were identified in patients with additional physical findings suggesting a higher detection rate for the more complex cases. The rate of pathogenic and likely pathogenic findings is comparable to the conservative published estimates of diagnostic yield for CMA in patients with ASD.

861F

Array comparative genomic hybridization findings in autism spectrum disorders. *A. Battaglia, G. Scarselli, R. Tancredi, A. Cosenza, R. Igliozzi, F. Apicella.* Dev Neurosciences, Stella Maris nst/Univ Pisa, Pisa, Italy.

Early presumptions opined that autism spectrum disorder (ASD) was related to the rearing of children by emotionally-distant mothers. Advances in the 1960s and 1970s clearly demonstrated the biologic basis of autism with a high heritability. To date, specific etiologic factors in ASDs can be identified in 30%–40% of patients. The development and refinement of array comparative genomic hybridization (a-CGH) has led to expanded applications as a diagnostic tool. Recent reports suggest a high diagnostic yield for a-CGH in ASDs. The objective of this study was to determine the diagnostic yield in a-CGH for ASDs at the Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry. We report the diagnostic yield of a-CGH in 200 samples with a primary indication of autism. All propositi were personally assessed with a detailed dysmorphologic, neurological, and neuropsychological evaluation. ASD diagnosis was based on criteria from the DSM-IV-R, and DSM5 since 2013, which was confirmed by ADOS and ADI-R assessments. 20% of the patients had abnormal findings on a-CGH. This study further supports a-CGH in the etiologic evaluation of ASDs, and elevation of array to a first tier diagnostic test. In the current era of emerging efforts in "personalized medicine", identifying an etiology will be critical in identifying endo-phenotypic groups and individual variations that will allow for tailored treatment for individuals with ASD.

862W

Identification of structural variants in neurodevelopmental disorders from sequence data in a New Zealand cohort. *W. Whitford[1], I. Hawkins[1], E. Glamuzina[2], D.R. Love[3], J. Taylor[4], R. Hill[5], K. Lehnert[1], R.G. Snell[1,6], J.C. Jacobsen[1,6].* 1) School of Biological Sciences, The University of Auckland, Auckland 1142, New Zealand; 2) Adult and Paediatric National Metabolic Service, Auckland City Hospital, New Zealand; 3) Diagnostic Genetics LabPLUS, Auckland City Hospital, Auckland, New Zealand; 4) Genetic Health Service New Zealand, Auckland City Hospital, New Zealand; 5) Department of Neurology, Auckland City Hospital, New Zealand; 6) Centre for Brain Research, The University of Auckland, New Zealand.

There is a significant overrepresentation of genetic structural variants (SV) in individuals with Neurodevelopmental Disorders. New Zealand has a unique and diverse population for studying genetic diseases, however current standard diagnostic practices do not include next generation sequencing (NGS). SVs including copy number variants (CNV) and balanced chromosomal rearrangements are currently identified using low resolution techniques such as karyotyping and array comparative genomic hybridisation (aCGH). To overcome the inherent limitations of these techniques we have employed NGS by means of whole genome sequencing (WGS) data in a New Zealand cohort of ASD and other related neurodevelopmental disorders, and we present the results from one of the cases here. The family consists of two affected siblings, who presented with recurrent episodes of neurological deterioration, bilateral basal ganglia strokes, and in one sibling, autism spectrum disorder. aCGH revealed two shared CNVs that could not be linked to the disease phenotype. We subsequently employed WGS which revealed an inherited heterozygous missense mutation in exon 2 (c.68C>V) of the solute carrier family 19 (thiamine transporter), member 3 (*SLC19A3* [MIM 606152]) gene. This p.Gly23Val missense mutation affects a conserved amino acid in the transmembrane domain of the carrier protein. No other putative SNP was identified in this gene, however subsequent analysis of the WGS revealed the presence of a 4808 base pair deletion in the 5' untranslated region of *SLC19A3* (Chr2:228,582,251–228,587,059). Mutations in this gene have previously been found to cause Thiamine metabolism dysfunction syndrome 2 (THMD2 [MIM 607483]) which aligns with the clinical manifestation in these children. The syndrome is known to be biotin and thiamine responsive, with an alternate name of thiamine-biotin responsive basal ganglia disease. Subsequent treatment with biotin and thiamine has resulted in significant clinical improvements. Thus, WGS was able to identify a compound heterozygous mutation that was not detected by current diagnostic assays, and has resulted in a successful targeted treatment.

863T

Novel mutations and further phenotypic expansion for early onset epileptic encephalopathy genes. *E. Bettella[1], R. Polli[1], E. Leonardi[1], F. Cesca[1], M. Vecchi[1], I. Toldo[1], C. Boniver[1], S. Sartori[1], A. Murgia[1,2].* 1) Department of Women's and Children's Health, University of Padua, Padua Italy; 2) Department of Neuroscience, University of Padua, Padua Italy.

Early Onset Epileptic Encephalopathies (EOEEs) are a common and a heterogeneous group of neurological disorders of early infancy, characterized by premature and often drug-resistant seizures mainly resulting in motor and cognitive deficits. Indeed EOEEs represent a final common pathway for a broad spectrum of genetically heterogeneous conditions. As single gene testing is no longer a practical approach, the use of Next Generation Sequencing (NGS) should be applied as a routine molecular diagnostic strategy in patients with EOEE. With the use of a custom targeted multigene panel and data analysis pipeline, we have identified and characterized 7 disease-causing variants that seem to further expand the phenotypic range of five early onset epilepsy genes. We have detected four novel *de novo* mutations: *SCN1A* c.5732A>G(p.K1911R) (MIM#182389; NM_001165963), *SCN2A* c.4993C>T(p.L1665F)(MIM#182390;NM_021007), *KCNQ2* c.845A>T(p.D282V) (MIM#602235;NM_172107) and *FOXG1* c.1135C>T(p.L379F) (MIM#164874;NM_005249); one novel maternally inherited variant in *SCN1A* c.3572G>A(p.C1191Y) and one *de novo* known variant in *SCN2A* c.3631G>A(p.E121K). Furthermore, we have identified a novel homozygous mutation of the *ALDH7A1* gene c.1256C>T(p.S419L) (MIM#107323;NM_001202404) in the newborn child of consanguineous Pakistani parents. The phenotypes of all the patients carrying these mutations were peculiar both from the clinical and electroclinical standpoints and can further expand the genotypic and clinical spectrum of these disorders. Our findings contribute to highlighting the extreme variability of phenotypes associated with mutations in known early onset epilepsy genes and show that the more insights we get into the molecular bases of genetic disorders which definition is based uniquely on clinical grounds, the more we have to accept that the relationship between the genotype and phenotype may be anything but straightforward. Finally, a careful, refined and comprehensive description of the patients' clinical features is the key factor to expand the phenotypic spectrum associated with single nucleotide variants, depending on their nature and location .

864F

Splice site mutation of *TCF4* gene in a patient with suspected Pitt-Hopkins syndrome. *S. Chahrokh-Zadeh, A. Munzig, M. Ziegler, S. Doelken, I. Rost.* Center for Human Genetics and Laboratory Medicine, Martinsried, Germany.

Pitt-Hopkins syndrome (PHS) is a rare neurodevelopmental genetic disorder, remaining underdiagnosed due to similarities with other known genetic syndromes like Angelman or Rett Syndrome. It is mainly characterized by extensive developmental delay, severe intellectual disability (ID), and a typical facial gestalt. Other features include episodic hyperventilation, seizures, eye abnormalities and constipation. Mutations of the *TCF4* gene were identified as disease-causing and haploinsufficiency of TCF-4 has been reported to cause classical PHS.   We report on a 20-year old girl, born as a second child to healthy, non-consanguineous parents, presenting with severe developmental delay, severe speech delay, facial dysmorphism and movement disorder. Initial genetic investigations at age 4 including molecular genetic diagnostics of Angelman-Syndrome (Methylation assay, Deletion-/UPD-testing, sequencing of *UBE3A*-gene) and Fragile X syndrome (Southern Blot analysis), classical karyotyping and array-based comparative genomic hybridization (array-CGH) revealed negative results.   We performed molecular genetic diagnostics of *TCF4* gene consisting of Sanger sequencing followed by MLPA (multiplex ligation-dependent probe amplification) analysis. By sequence analysis we were able to identify the splice donor variant c.1486+3A>C of the *TCF4* gene (NM_001083962.1; chr18:52901776(hg19)). This substitution is localized at position +3 of intron 16 (IVS16+3 A>C) and was not reported in relevant databases (ExAC Browser (Exome Aggregation Consortium); dbSNP; EVS) or the literature to date. By the result of parental analysis the variant had occurred de novo in the patient. Bioinformatic algorithms (MaxEntScan, Human Splicing Finder a.o.) predicted not to significantly alter the splice site activity.   In order to determine the in vivo effect of this substitution on the splice donor site of intron 16 in RNA extracted from the patients' platelets and leukocytes we performed functional RNA studies to elucidate the true effect of the variation detected in this individual. We were able to show, that this variant results in the production of a critical amount of aberrantly spliced transcripts, ultimately leading to haploinsufficiency of tcf4 protein. Aberrant transcripts contain translational stop codons leading to productions of variant proteins, that most likely will be completely degraded.

865W

Use of chromosomal microarray in Hong Kong patients with autism spectrum disorder: Implications of a copy number variation involving *DPP10*. *H. Chung[1,2,4], S.L. Mak[3], T.G.A. Chiu[1,4], K.C. Leung[1], C.Y. Mak[1], W.Y. Chu[1], T.K. Mok[1], Y.K. Chan[2], S.Y.A. Kan[2], H.Y.M. Tang[2], T.K. Lau[2], C.W. Fung[1,4], S.L. Lee[1,4].* 1) Pediatrics and Adolescent Medicine, Queen Mary Hospital, HKU, Hong Kong; 2) Obstetrics and Gynaecology, Queen Mary Hospital, HKU, Hong Kong; 3) Obstetrics and Gynaecology, Queen Elizabeth Hospital. , Hong Kong; 4) Pediatrics and Adolescent Medicine, Duchess of Kent Children's Hospital, HKU, Hong Kong.

Background Array Comparative Genomic Hybridization (aCGH) is now the recommended first-tier genetic test for children with autism spectrum disorder (ASD). However, its interpretation can be challenging. Genetic counselling can be further complicated by the fact that contribution of CNVs in non-European ASD patients has not been well studied. We report the CNV findings in a clinical cohort of Chinese children with ASD in Hong Kong.   **Methods** DNA samples were obtained from 288 ASD patients recruited from a child assessment centre in Hong Kong between January 2011 and August 2014. aCGH was performed using the NimbleGen CGX-135k oligonucleotide array or Agilent CGX 60k oligonucleotide array in a clinical laboratory. Results were classified based on existing guidelines with reference to medical literature and CNV databases.   **Results** Of the 288 patients, 5 had pathogenic and 4 had likely pathogenic CNV. The diagnostic yield was ~3%. A small duplication overlapping the *DPP10* gene (hg[19]chr2:116,534,689-116,672,358), which was previously implicated in the pathogenesis of ASD, was identified once in the ASD cohort (0.35%). The same variant was observed at a frequency of 0.96% in our internal controls, and 1.2% amongst the parental samples (all of typical development) sent for prenatal diagnosis respectively. Multiple individuals of typical development carrying a similar DPP10 duplication were also identified in regional CNV databases with predominantly Chinese population.   **Conclusions** Our findings suggest that the *DPP10* duplication is likely a benign copy number polymorphism frequently seen in southern Chinese. This highlights the importance of (a) using ancestry-matched controls in clinical interpretation of aCGH findings and (b) recruiting cases and controls of well-mixed ethnicity backgrounds in genomic studies.

866T

Maternal derived 15q11.2-q13.1 duplication in a patient with autism and seizure. Case report. *M. Perez Sanchez[1,2], M. Martinez[1,2], S. Garcia[1,2], A. Mora[1,2], A.R. Gonzalez[2,3].* 1) Genetics, Complejo Hospitalario de Granada., Granada, Granada, Spain; 2) Instituto Biosanitario de Granada, Spain; 3) FIBAO, Complejo Hospitalario de Granada, Spain.

   Autism is a neurodevelopmental disorder charazterized by impaired social interaction and comunication and a restricted range of interests and activities, with onset during thethe first 3 years of life. Duplications of the 15q11-q13 region are the most frequently reported chromosomal aberration in individuals with autism Thie region includes the Prader-Willi/Angelman syndrome (PWS/AS) critical region, wich is subjet to genomic imprinting. Most duplications of this interval are caused by supernumerary chromosomes formed by the inverted duplication of proximal 15q, know as isodicentric chromosome 15. Intersticial duplications of this region are less frequent, but many cases have been reported in association with autism. The majority of cases are associated with maternally derived duplications, whereas paternal inheritance usually leads to normal phenotypes.We describe here a female patient in wich autism segregates with maternal inherance of a proximal 15q duplication. The patient was born at 37 gestation weeks with a weigh of 3150 gr. and no incidences in neonatal period. In posterior pediatric revission was detected language and cognitive delayed, supected due to autistic disorder (AD), at 3 years age preseted epilepsy episodes. The familiar history include fater and maternal grandmother with epilepsy, maternal uncle with autism and a sister healty.Comparative Genomic Hybridization (CGH-Array) with the Nimblegen CGX Cytogenetic Microarrays platform, suplied by PerkinElmer, and karyotype was performed for the patient and parentsThe CGH- Array was normal in the father, but both mother and patient presented a duplication of 5,56 Mb in 15q11.2-q13.1 PWS/AS region (arr(hg19)15q11.2-q13.1(22,822,019-28,379,369)x3). No others chromosomal anomalies were detected.Karyotype was normal in the patients and parents.The Array-CGH and karyotype from paternal grandmother, maternal uncle and sister were no done.This results are according with previous studies that show the correlation between this duplication and autistic disorder when have been maternal in origin. The family history with the maternal uncle diagnosed of autistic disorder also agree this posibility. The epilepsy, presented in some members of father`s family, is possible that can be originated for another unknow genetics alteration, without any correlation with the microduplication in PSW/AS region described in the patient.

867F

Co-existence of a complex, three-way translocation with a 4.6 Mb deletion in 8q22.3-8q23.1. *K. Swisshelm[1], S. Toomey[1], J. LeRoux[1], R. Rowe[1], K. Ha[1], B. Carstens[1], E. Elias[2].* 1) Dept Pathology, Colorado Genetics Laboratory, Univ Colorado Denver Anschutz Medical Campus, Aurora, CO; 2) Department of Pediatrics and Colorado Childrens Hospital, Aurora, CO.

   We present a case of a non-verbal 18 year old male with skeletal abnormalities, including mild scoliosis, severe intellectual disability, autism and generalized convulsive and intractable epilepsy.   Initial high resolution chromosome studies revealed a complex karyotype involving a three-way translocation of chromosomes 7, 11 and 12, with an insertion of 12p material into the short arm of chromosome 7. The ISCN for this finding is: 46,XY,t(7;11;12)(12pter->12p12.2::7p15.1->7qter;11pter->11q22.3::7p15.2->7pter;11qter->11q22.3::7p15.1->7p15.1::12p12.2->12qter).   Single nucleotide polymorphism chromosomal microarray testing revealed no gains or losses of genomic material involving chromosomes 7, 11 and 12. However, an additional finding of a 4.6 Mb loss within the long arm of chromosome 8, at 8q22.3-8q23.1 (101,930,909-106,510,518; hg19) was observed. Upon inspection of the high resolution chromosomes, this finding could be visualized at the 450-550 band level. Parental studies were negative for a rearrangement of chromosomes 7, 11 and 12 and were negative for the 8q22.3-8q23.1 loss by fluorescence in situ hybridization. Therefore, the findings of a complex chromosomal rearrangement and an 8q deletion in this patient are *de novo*.   An overlapping region of deletion within 8q has been reported in the literature in a case of an 8 year old female with intellectual disability and epilepsy (Kuroda Y et al., Am J Med Genet 164A:2104-2108, 2014), and in five patients with similar deletions, all with intellectual disability and dysmorphic features, and 4 of 5 with epilepsy (Kuechler A et al., Am J Med Genet 155A:1857-1864, 2011). Our patient's deletion involved seven disease genes, including *GRHL2, RRM2B, FZD6, CTHRC1, SLC25A32, DPYS,* and *ZFPM2.* In this case, the complex three-way translocation appears to be balanced, with no gain or loss of genomic material. While disruption of genes involved in the rearrangement of chromosomes 7, 11, and 12 cannot be excluded, our patient's phenotype can be attributed to the 4.6 Mb loss within 8q. Our findings underscore the importance of a complete genomic laboratory testing to include both karyotype and chromosomal microarray in the absence of other molecular finding.

868W

What can go wrong in the molecular assays when the results conflict the clinical diagnoses. *C. Tsai.* pediatrics/genetics, childrens Hospital colorado, Aurora, CO.

BACKGROUND: Molecular genetics testing have been applied to our daily medical practice. Current methods for single gene disorders including Sanger sequencing and high through-put next generation sequencing (NGS): PCR or captured based panels as well as exome sequencing. The predictive value and clinical applications are variable from one condition to the other. Numerous questions remain including technology deficits, data interpretation and unpredicted phenotypes in complex disorders.OBJECTIVE: In this study, the author reviewed all patients with conflicting results between molecular results and clinical diagnosis. Addition testing on or include the suspected genes were studied which subsequently revealed or confirmed the diagnosis.DESIGN/ METHODS: From 2011-2015, 22 patients with classic or highly suggestive clinical diagnosis of congenital malformation syndrome yet initial molecular assay did not identify an anomaly were reviewed and offered additional study. 15 patients were able to establish a final diagnosis.RESULTS: A patient with Pompe disease had 2 mutations missed by Sanger but picked up by NGS and microarray due to allele dropout and a partial exon deletion. 2 siblings with MELAS mutation each has a ratio of 19 % and 13% mosaicism originally missed by sanger but picked up by NGS. 1 patient with classic Noonan syndrome due to PTPN missense mutation was missed by Sanger and picked up by NGS due to allele dropout. 2 patients with FOXG1 mutation missed by NGS panel but picked up by exome sequencing due to script and filtering errors. 4 Angelman syndrome pt with normal Sanger sequencing of UBE3A were found to have CRIN2B, SLC6A9, CTNNB1 and PQBP1 mutations by panel assay. 2 patient thought to have Waardenburg syndrome and NF1 respectly were proved to have KIT mutations, 1 patients with retinopathy were missed by captured based NGS panel but picked up by PCR based NGS ones. CONCLUSIONS: Most capture based panel assays continue to miss gene changes from the capturing process, exon 1 mutations, GC rich genes and pseudogenes. Bioinformatics analysis remain critical for calling predicting the consequences of the variant. The author demonstrated a series of cases of laboratory pitfalls. Clinicians need to recognize that a mutation can be missed by one reputable lab and detected by another using a different platform/methodology; understand the limitations of each testing is critical in making the correct diagnosis. .

869T

Microarray comparative genomic hybridization analysis (aCGH) in patients with multiple congenital abnormalities and global developmental delay. *R.I. Ursu[1,2], V.E. Radoi[1,2], G.P. Chelu[2], C. Arsene[2], R.E Bohiltea[1], L.C. Bohiltea[1].* 1) "Carol Davila" University of Medicine and Pharmacy, Faculty of General Medicine, Bucharest, Bucharest, Romania; 2) SYNEVO Romania, Central Laboratory, Medical Genetics Department, Bucharest, Romania.

Background: The array comparative genomice hybridization technique (aCGH, molecular karyotype) has become one of the most important diagnostic tools in identifying genomic abnormalities associated with developmental disabilities (congenital malformations, cognitive impairment and behavioral disorders, microdeletion and microduplication syndromes). **Purpose:** The purpose of this presentation is to describe the optimal medical genetics evaluation and diagnosis of the child with intellectual disability and global developmental delay. **Material. Method:** Study group consists of 8 children with epileptic enephalopathy, hypotonia, psychomotor impairment, different forms of epilepsy, microcephaly, intellectual disability. Constitutional karyotyping has been performed in all the cases with normal findings. The next diagnostic step in the management of the patients has been the molecular karyotype (aCGH) on an Agilent platform using SurePrint G3 CGH ISCA v2 Microarray, 8X60K. Data analysis and clinical interpretation have been completed using the Agilent CytoGenomics 3.0.2.11 Software, ADM2 algoritm. **Results:** The results revealed pathogenic variants in 4 cases: Microduplications with pathogenic significance (correlated with the clinical indications): 2 cases, 1p36.32; 8p23.1 Microdeletions with pathogenic significance (correlated with the clinical indications): 2 cases, 14q31.1; 1p36.33p36.23 Microdeletions and microduplications with benign or uncertain significance *(according to current genetics knowledge):* 4 cases. **Conclusions:** Although arrayCGH is currently being used as a complementary test to standard cytogenetic techniques, it is likely to become the genetic test of choice, especially in cases of idiopathic developmental delays, epilepsy, multiple congenital anomalies microdeletion and microduplication syndromes.

870F

22q11.2 duplication: Medical co-morbidities, intellectual functioning and neurodevelopmental concerns in probands and non-proband carriers.

M.T. Carter[1], B. McConnell[2], B. Mackinnon[2], S. Jilderda[2], J. Falcon[2], S. Saleh[2], G. Kamboj[2], A. Noor[3], A.S. Bassett[4], M. Speevak[5], D.J. Stavropoulos[6], I. Drmic[2]. 1) Regional Genetics Program, The Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) Autism Research Unit, SickKids Hospital, Toronto, ON, Canada; 3) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada; 4) Dalglish Family 22q Clinic, University Health Network & Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 5) Department of Genetics, Trillium Health Partners Credit Valley Site, Toronto, ON, Canada; 6) Department of Pathology and Laboratory Medicine, SickKids Hospital, Toronto, ON, Canada.

   A wide range of phenotypic manifestations have been described in probands with duplication of the 22q11.2 (LCR-a to LCR-d) region in case reports. Non-proband carriers of the duplication (i.e. parents and siblings of probands) are very rarely described, despite the fact that this duplication is usually inherited. We collected medical history information on 46 probands and 12 non-proband carriers of the 22q11.2 duplication; we also collected psychometric data on 10 of the probands and all 12 of the non-proband carriers. While 28% had one or more major congenital anomalies, no specific pattern of anomalies emerged. Dysmorphic facial features were absent in non-probands and rare in probands. One-quarter of the individuals tested had no clinically significant findings based on best-estimate clinical diagnosis that included file review and psychological testing (which included standardized measures of intelligence, academic achievement, adaptive functioning, social communication, attention, anxiety and depression). Thirty-three percent had a significant split between verbal IQ (VIQ) and performance IQ (PIQ); all but one individual had significantly higher VIQ. Of the remaining individuals for whom a Full Scale IQ was calculated, the majority were within the Average range. Similarly, reading and math skills were in the average range for the majority (58 and 70%, respectively). Learning disability and attention deficit-hyperactivity disorder symptoms were more common in probands versus non-probands, while anxiety and depression showed the reverse trend. In conclusion, our results show that most individuals with 22q11.2 duplication have average intellectual functioning, and those with an uneven cognitive profile tend to have stronger verbal reasoning skills. They may be at slightly elevated risk for congenital anomalies, but no specific pattern of anomalies is characteristic. Future studies will examine whether congenital anomalies and/or intellectual deficits are correlated with the presence of additional copy number variants.

871W

***De novo* 17q21.31 microdeletion in two girls with intellectual disability detected by Chromosomal Microarray Analysis.**

I.P. Pinto[1,2], L.B. Minasi[2,3], G.R. Nascimento[2,3], A.V. de Melo[1,2], D.M.C. Cunha[2], C.L. Ribeiro[1,2], C.C. da Silva[1,2,3,4,5], A.D. da Cruz[1,2,3,4]. 1) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil.

   The Chromosome 17q21.31 Deletion Syndrome (MIM 610443) is a genomic disorder caused by the microdeletions between 500-650 Kb that are mediated by non-allelic homologs recombination between flaking low-copy repeats regions. The deletion at 17q21.31 shows clinical recognizable manifestation characterized by intellectual disability, hypotonia, a friendly or amiable behavior, distinctive facial features, epilepsy, heart defect and urogenital malformation. Here, we report two cases of girls who presented moderate intellectual disability, friendly behavior, mild facial dysmorphisms with long face and tubular or "pear shaped" nose with bulbous nasal tip, and epileptic seizure with de novo 17q21.31 microdeletion in Central Brazil. Conventional cytogenetics analysis by GTG banding using the software IKAROS® (Metasystems Corporation, Germany) showed the normal female karyotype 46,XX for both girls. Chromosomal Microarray Analysis (CMA) using GeneChip® CytoScan-HD™ array revealed in all two girls a de novo microdeletion spanning ~ 500 kb in 17q21.31, encompassing 9 genes (*CRHR1, MGC57346, CRHR1-IT1, MAPT-AS1, SPPL2C, MAPT, MAPT-IT1, STH, KANSL1*). The evaluation of progenitor's CMA confirmed de novo genomic imbalances in their each child. *CRHR1* and *MAPT* genes are known to be highly expressed in brain and have been implicated in behavioral phenotypes and several neurodegenerative disorders, and *KANSL1* gene is expressed in human tissues, including in the central nervous system. Haploinsufficiency of one or more of these genes within the deleted region in both girls is the most probable cause of the probands phenotype and is responsible for the phenotype seen in the Chromosome 17q21.31 Deletion Syndrome. The CMA analysis has proving to be a powerful tool to carry out genetic diagnostic and it was useful method to identify at the first time in Central Brazil the 17q21.31 microdeletion associated with intellectual disability phenotypes. The authors believe that the rare and complex phenotypes need to be investigated to allow specify phenotypes classification. Therefore, the CMA is an efficient strategy to help in delineating phenotypic variation and allow adequate clinical management and better follow up of the probands and the family.

872T

Identification of a *KAT6B* mutation in a family with developmental delay and intellectual disability leading to the diagnosis of SBBYSS syndrome via diagnostic exome sequencing. *H. Kim[1], Y. Kim[2], J. Park[2], J. Jang[3], E. Cho[3].* 1) Dept.of Medical Genetics, Konyang University Hospital, Deajon, South Korea; 2) Dept. of Rehabilitation Medicine, Konyang University College of Medicine, Daejon, South Korea; 3) Green Cross Genome, South Korea.

Developmental delay (DD) and intellectual disability (ID) are major neuropsychological disorders in children, which affect 1-3% of the general population and substantial portion of cases of DD/ID are proven to be genetic.(17.4-47.1%) Recently, heterozygous *de novo* mutation of the histone acetyltransferase encoding gene KAT6B has been recognized to cause a syndrome with congenital anomalies and intellectual disability, namely Say-Barber-Biesecker/Young Simpson syndrome (SBBYSS). Here, we present a SBBYSS syndrome of a Korean Family affected in three generations with a missense mutation in *KAT6B*, c.2292C>T (p. His767Tyr) identified by diagnostic exome sequencing(DES), which has not been reported in Korea yet. The proband was 3years old female with facial dysmorphism of bilateral blepharophimosis and strabismus and microcephaly in addition to DD. And she was diagnosed as SBBYSS syndrome with the identification of mutation in KAT6B and subsequentcy, the same mutation was found in her father, uncle, and grandmother. Thus SBBYSS syndrome was diagnosed in four affected members in three generations in the family with DD/ID. DES was useful diagnostic tool for the identification of molecular diagnosis of DD/ID and leading to accurate and appropriate genetic counseling for the family with DD/ID.

873F

Severe intellectual disability in a patient with Turner syndrome features. *U. Kappes[1,2], D.G. Basel[3], D.P. Bick[4], B. Chirempes[5], L. Weik[3], P. van Tuinen[6], J.F. Peterson[6].* 1) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI; 2) Department of Pediatrics, Division of Genomic Pediatrics, Medical College of Wisconsin, Milwaukee, WI; 3) Department of Pediatrics, Division of Genetics, Medical College of Wisconsin, Milwaukee, WI; 4) Hudson-Alpha Institute for Biotechnology, Huntsville, AL; 5) The Advanced Genomics Laboratory, Children's Hospital of Wisconsin, Milwaukee, WI; 6) Department of Pathology, Medical College of Wisconsin, Milwaukee, WI.

Turner syndrome and its variants typically result in short stature, premature ovarian failure, and variable dysmorphic features. We report a 20-year-old female patient with a history of short stature, premature ovarian failure, sacral teratoma, neurogenic bladder, and severe intellectual disability. At age 18 classical cytogenetic analysis (50 PHA-stimulated cells) revealed a terminal Xq24 deletion in 37 cells, a ring X chromosome in 6 cells, and monosomy X in 7 cells. At that time a diagnosis of Turner syndrome was rendered. Two years later further cytogenetic studies were initiated to better understand the etiology of a phenotype that surpasses the characteristic features of Turner syndrome. Array comparative genomic hybridization (aCGH) analysis was performed on purified DNA from the proband using the CytoScan HD microarray (Affymetrix, Santa Clara, CA), and scanned with a Genechip Scanner 3000 (Affymetrix). Microarray analysis revealed a concomitant 38.5-Mb duplication of 3q25.33-q29, and an 82.9-Mb heterozygous Xq13.2-q28 deletion. The 3q25.33-q29 duplication contained 420 genes, and the Xq13.2-q28 deletion contained 962 genes including *XIST*. To confirm the location of the 3q25.33-q29 duplication, FISH analysis was performed on metaphase cells using a *BCL6* (3q27) break-apart DNA probe (Abbott Molecular, Des Plaines, IL). Positive hybridization for *BCL6* was observed on both copies of chromosome 3, in addition to the long arm of the abnormal X chromosome. The duplicated 3q segment located on the abnormal X chromosome accounts for the widely discrepant Xq breakpoints seen by the classical cytogenetic (Xq24) and microarray (Xq13.2) analyses. The mosaic cell lines observed in the patient's karyotype likely indicate a *de novo* event. Parental chromosome studies have yet to be performed to investigate balanced translocation carrier status. The intellectual disability together with a Turner-like phenotype is due to an unbalanced partial X;autosome translocation resulting in partial trisomy 3q. Genes on the abnormal X chromosome remain activated due to loss of the X-inactivation center (*XIST*). This case demonstrates that a multifaceted molecular approach may be necessary to resolve complex constitutional disorders.

874W

Functional disomy of Xp22 and *de novo* t(X;3). *A. Mohamed[1], A. Kamel[1], M. Zaki[2], M. Mekkawy[1], P. Safwat[1], I. Fadel[1].* 1) Human cytogenetics, National Research Centre, Cairo, Egypt; 2) Clinical Genetics, National Research Centre.

We report on a female patient aged 2 years complaint of delayed milestones and abnormal facies. She is the third child of healthy non consanguineous parents and had normal sibs. Since birth there were dysmorphic facial features. On examination the weight was 10.5 Kg (-1SD), Ht. 88 (mean) and head circumference 46cm(-1SD). She had hypertrichosis, depressed nasal bridge,sagging lower eye lids and low set ears. She had hypotonia and the reflexes were present. MRI, EEG and Echo were normal, abdominal sonar was normal. Her karyotype showed 46,XX,t(X;3)(3q12;Xp22.1). FISH analysis using whole chromosome paint for chromosomes 3 and X, subtelomeres 3p and q, Xp subtelomere, approved that translocation is a reciprocal translocation which is apparently balanced one. Chromosomes of both parents were normal.X chromosome inactivation performed using Late replicating chromatin technique in order to detect the pattern of X inactivation. The abnormal X chromosome revealed that the normal X chromosome is the active one in all analyzed cells and the translocated X is the inactive one, the translocated 3q part to the X chromosome escape X inactivations**.** The Xp22.1 segment translocated to 3q is out of XIST gene control and escaped X inactivation. In this patient the delayed milstones and dysmorphic features bay be due to functional disomy of Xp22,1 or other cryptic chromosome abnormalities involving Xp and 3q.

875T

Impact of long contiguous stretches of homozygosity in three girls with intellectual disability. *D.C. Silva[1,6], I.P. Pinto[1,3], L.B. Minasi[1,2], D.M.C. Cunha[1], C.L. Ribeiro[1,3], C.C. da Silva[1,2,3,4,5], A.D. da Cruz[1,2,3,4].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil; 6) Federal University of Goias, Animal Science PhD Program, Campus Samambaia, Goiânia, GO, Brazil.

The genetic causes of intellectual disability (ID) and autism include almost all types of genomic variations (chromosomal rearrangements, copy number variations and single gene mutations). Epigenetic alterations due to genomic variations affecting genes involved in epigenomic regulation and uniparental disomy resulting from chromosomal or segmental homozygosity are shown to contribute to the etiology of neurodevelopmental disease. Probably the commonest type of epigenomic variations in humans is long contiguous stretches of homozygosity (LCSH), and the presence of LCSH can be indicative for parental consanguinity, uniparental disomy, or segmental homozygosity for single gene recessive mutations. Herein, we report three girls with moderate intellectual disability who presented LCSH at Xq11.1q13.1 detected by Chromosomal Microarray Analysis (CMA). Karyotyping at > 550 band resolution for both three girls showed a female karyotype (46,XX). Affymetrix's GeneChip® CytoScanHD™ array revealed the same alteration in all three girls, LCSH spanning ~ 6.20 Mb in Xq11.1q13.1, with 21 genes (*SPIN4, LOC92249, ARHGEF9, MIR1468, AMER1, ASB12, MTMR8, ZC4H2, ZC3H12B, LAS1L, FRMD8P1, MSN, MIR223, VSIG4, HEPH, EDA2R, AR, OPHN1, YIPF6, STARD8, EFNB1*). Epilepsy and ID are the common phenotypic effects of all *ARHGEF9* mutations. Point mutations, rearrangements, and small deletions of *ZC4H2* gene cause a clinically variable broad-spectrum neurodevelopmental disorder of the central and peripheral nervous systems. Mutations in the *OPHN1* gene are responsible for OPHN1-related X-linked mental retardation with cerebellar hypoplasia and distinctive facial dysmorphisms. Clinically, LCSH of any size can aid in diagnosis of autosomal recessive disease through homozygosity mapping. Furthermore, it is now recognized that the effect of LCSH is not only significant for monogenic recessive disorders, but there is a role for recessive variants in complex genetic disorders. In several ways, LCSH loci appear comparable to large genomic CNVs that are well known to be associated with autism and ID, and some LCSH loci can have features of a contiguous gene disorder. Therefore, the CMA with polymorphic and non-polymorphic markers is an important diagnostic tool to identify both CNV and LCSH in patients with a wide variety of clinical indications. However, the mechanisms and consequences of LCSH are poorly understood, and the interpretation of these epigenetic mutations can represent a challenge.

876F

Use and application of chromosome microdissection technique to isolate the 4p16.3 region. *A.P. Amancio[1,2], I.P. Pinto[1,2], C.C. da Silva[1,2,3,4], A.D. da Cruz[1,2,3].* 1) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 4) State University of Goias, UnU Eseffego, Brazil.

Around 5% of pregnancies result in the birth of a child with some congenital disorder, disability or genetic disease that compromises the development and the quality of life. Monogenic diseases or chromosomal alterations cause more than 40% of severe intellectual disability cases. The Wolf-Hirschhorn Syndrome (WHS) is a genetic disorder characterized by severe growth retardation, severe intellectual disability, severe facial dimorphism and seizures. This syndrome is caused by a terminal microdeletion at 4p16.3 region from the short arm of chromosome 4, in which the size of deletion varies among affected individuals. Conventional cytogenetics analysis by GTG banding does not have enough resolution to diagnosis of WHS. On the order hand, molecular cytogenetic techniques such as FISH and Chromosomal Microarray Analysis (CMA) are necessary for full characterization of chromosomal rearrangements associated with WHS. The aim of this study was to perform chromosome microdissection technique to isolate the chromosomal segment from 4p16.3 region. A total of 5 mL of peripheral blood was drawn using a standard vacuum extraction blood-collecting system containing heparin from a non-syndromic individual with intellectual disability to obtain the chromosomes using the cell culture. The sample from cell culture was dripped on glass coverslips for scraped chromosomal, and while manipulating the coverslip and microscope table, the needle was maintained at the high position to avoid breakage. Detachment of the chromosomes from the coverslip was achieved by gently scraping the chromosome along its vertical axis, causing the chromosome to fold around itself and finally bind to the needle. The microdissection was performed using an inverted microscope (Zeiss, Germany) with the manipulator (Eppendorf, Germany). The 4p16.3 region from chromosome 4 was isolated by chromosome microdissection method for further development and validation of the DNA probe using the FISH technique. The chromosome microdissection technique is an important approach to isolate the chromosome and construction of DNA probe that can detect many genetic syndromes and hematological disorders. Furthermore, this technique is useful for decreasing the cost of production of probes, allowing access to detection of genetic syndromes caused by microrearrangements in public health service.

877W

Microarray analysis of an infant with a rare *de novo* 1p31.3p31.1 deletion from Central Brazil. *C.C. da Silva[1,2,3,4,5], I.P. Pinto[1,3], D.M.C. Cunha[1], C.L. Ribeiro[1,3], L.B. Minasi[1,2], A.D. da Cruz[1,2,3,4].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil.

The Chromosome 1p32-p31 Deletion Syndrome (MIM 613735) is rare disorder and at least 19 cases have been reported. Patients with 1p interstitial microdeletions display a number of clinical features due to a phenotype expressivity and deletion size be variable including intellectual disability, craniofacial anomalies, seizures, tooth abnormalities, urinary tract anomalies, and skeletal, cardiac as well as limb defects. Here, we report an 8 years old girl who was born to non-consanguineous parents at 37 weeks gestation, and her birth weight was 2095 g and length of 44 cm. Physical examination revealed small nose and bulbous nasal tip, low-set ears, small chin, tooth abnormalities, hypotonia, moderate intellectual disability, behavioral disturbs, and seizures requiring the use of phenobarbital. The family history revealed that her paternal relatives had intellectual disability and epileptic seizures. Karyotyping at > 550 band resolution showed a female karyotype (46,XX). Chromosomal Microarray Analysis using Affymetrix GeneChip® CytoScanHD™ array revealed a *de novo* pathogenic 10.89 Mb microdeletion at 1p31.3p31.1 (68,693,129-79,580,916) x1. This region encompassing 49 genes (*WLS, RPE65, DEPDC1, LRRC7, PIN1P1, LRRC40, SRSF11, ANKRD13C, HHLA3, CTH, PTGER3, ZRANB2-AS1, ZRANB2, MIR186, ZRANB2-AS2, NEGR1, NEGR1-IT1, LRRIQ3, FPGT, FPGT-TNNI3K, TNNI3K, C1orf173, CRYZ, TYW3, LHX8, SLC44A5, ACADM, RABGGTB, SNORD45C, SNORD45A, SNORD45B, MSH4, ASB17, ST6GALNAC3, ST6GALNAC5, PIGK, AK5, ZZZ3, USP33, FAM73A, NEXN, FUBP1, DNAJB4, GIPC2, MGC27382, PTGFR, IFI44L, IFI44,* and *ELTD1*). The progenitor's CMA confirmed *de novo* genomic imbalances in their child. Definition of chromosomal deletions can help to identify genes contributing to pathological phenotypes. The *AK5* gene is likely candidate gene that could contribute to the intellectual disability phenotype. This gene encodes a member of the adenylate kinase and is expresses exclusively in brain, thus, is associated to a specific needs of different cellular functions. The other genes in the deleted region did not appear to play a causative role for the phenotype, but we cannot exclude their involvement. Genetic evaluation of patients with intellectual disability using microarray analysis allowed obtain precise information about CNVs, identifying candidate genes contained within these regions, and provided knowledge about the clinical implications of these CNVs.

878T

Identification of microdeletion at 15q11.13 using fluorescence *in situ* **hybridization in Brazilian probands from public health service.** *V.C.P. Santos[1], S.S.S. Pereira[1,2], L.S. Teodoro[1,2], D.M.C. Cunha[1], C.L. Ribeiro[1,3], I.P. Pinto[1,3], C.C. da Silva[1,2,3,4,5], L.B. Minasi[1,2], A.D. da Cruz[1,2,3,4].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil.

   Prader-Willi (PWS) and Angelman (AS) syndromes are two distinct neurogenetic disorders associated with multiple anomalies and intellectual disability. Three genetics mechanisms are known to be involved in the pathogenesis: microdeletions, uniparental disomy (UPD) or loss of the imprinted gene expression in paternally or maternally 15q11-q13 region. Both disorders have clinically overlapped features with other diseases during infancy, which makes them difficult to diagnose solely relying on the clinical manifestations, therefore, definite diagnosis of PWS/AS is dependent of genetic tests. The aim of this study was report 32 cases from public health system in Central Brazil to confirm the suspect of PWS and AS. Fluorescence *in situ* Hibridization (FISH) analysis were performed using Prader-Willi/Angelman Probe that include 15q11.13 (spectrum red) and 15qter (spectrum green) according to the standard FISH protocols (Cytocell, USA), and chromosomes were counterstained with 4,6-diamino-2-phenyl-indole (DAPI). Capture and analysis of metaphases and interphase nuclei were released by Epifluorescence Microscope (Carl Zeiss, Germany) and the software ISIS® (Metasystems Corporation, Germany). Since 1997, assistant physicians from Goiás state public health system referred 32 patients with clinical indication of PWS/AS to our genetic service at both the Laboratory of Human Cytogenetic and Molecular Genetics and the Biology Department at Pontifical Catholic University in Central Brazil to perform FISH test. FISH analysis of 15q11.13 microdeletion revealed 6 (18.75%) patients with positive FISH results and 26 (81.25%) patients with negative FISH results. The method that is commonly used for PWS/AS detection is fluorescence *in situ* hybridization because 75% of the cases occurring by microdeletion and the remaining 25% could be due to UPD and imprinting gene mutation. In our cases, only 18.75% were diagnosed with 15q11.13 microdeletion, on the other hand, 81.25% of cases without 15q11.13 microdeletion could present different genetics mechanisms for PWS/AS. Therefore, it is recommended to carry on others techniques to identify these disorders such as methylation-specific PCR, multiplex ligation-dependent probe amplification, short tandem repeat linkage analysis, DNA sequencing, array comparative genomic hybridization, and chromosomal microarray analysis. In addition, it is important the correct diagnose for accurate genetic counseling for the patients and their family.

879F

The diagnosis experience for Cockayne syndrome-Complex *ERCC8* **variants in two Chinese sibings.** *H. Xie.* Capital Institute of Pediatrics ,Beijing Municipal Key Laboratory of Child Development and Nutriomic, BeiJing, China.

   Cockayne syndrome is an autosomal recessive disorder principally characterized by postnatal growth failure and progressive neurologic dysfunction. Mutations of the *ERCC6* and *ERCC8* genes are the predominant cause of Cockayne syndrome. We reported the diagnosis experience for two clinical suspected Cockayne syndrome patients in a Chinese family. Using multiple molecular techniques including whole exome sequencing, array comparative genomic hybridization and Quantitative polymerase chain reaction, we identified a maternal splicing mutation (chr5:60195556, NM_000082:c.618-2A>G) and a paternal complex exonic deletion (chr5:60211534-60213756, chr5:60212086-60217114) of the *ERCC8* and confirmed that is the pathogenesis of the two suspected patients. The microhomologies (TAA and AGCT) at the breakpoints proved microhomology-mediated FoSTeS events were involved in this complex *ERCC8* deletion. Our diagnosis experience also demonstrated the advantage of high-throughput genomic technologies in the clinical genetic diagnosis.

880W

Deep phenotypic and genetic investigation of array-CGH data reveals novel genomic disorders, position effects and refines genotype/phenotype correlations. *A. Brusco[1,2], E. Di Gregorio[1,2], E. Riberi[3], E.F. Belligni[3], E. Biamino[3], U. Ala[4,5], M. Spielman[13], I. Bagnasco[6], D. Carli[3], G. Gai[2], R. Keller[7], G. Mandrile[2], V.G. Naretto[2], F. Sirchia[2], L. Sorasio[8], A. Zonta[2], F. Talarico[2], P. Pappi[2], S. Cavalieri[1], E. Giorgio[1], C. Mancini[1], M. Gandione[9], A. Pelle[10], P. Provero[4,5], M. Cirillo Silengo[3], E. Grosso[2], J. Buxbaum[11,12], B. Pasini[1,2], S. De Rubeis[11,12], G.B. Ferrero[3], Neurowes.* 1) Medical Sciences, University of Turin, Torino, TO, Italy; 2) Città della Salute e della Scienza University Hospital, Medical Genetics Unit, 10126, Turin, Italy; 3) University of Torino, Department of Public Health and Pediatrics, 10126, Turin, Italy; 4) Molecular Biotechnology Center (MBC), Computational Biology Unit, Turin, Italy; 5) University of Torino , Department of Molecular Biotechnology and Health Sciences, CAP, Turin, Italy; 6) Martini Hospital, Maternal and Infantile Department, ASL TO1, Torino, Italy; 7) Adult Autism Center, ASL TO2, CAP, Torino, Italy; 8) Hospital "Santa Croce e Carle", Cuneo, Italy; 9) University of Torino, Department of Neuro-psychiatry, 10126, Turin, Italy; 10) University of Torino, Department of Clinical and Biological Sciences, and Medical Genetics, San Luigi Gonzaga University Hospital, Orbassano (TO), Italy; 11) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, CAP, New York, USA; 12) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, CAP, New York, USA; 13) Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, Berlin, 14195, Germany.

Using an array-CGH 60K platform, we analyzed 1,023 children with unexplained developmental delay/intellectual disability (DD/ID) or multiple congenital anomalies (MCA). We identified 11.1% of pathogenic, 4.4% of possibly pathogenic and 13.8% of unknown significance deletions/duplications. We recognized rare genomic disorders, such as 4q35 duplication (Coat's like disease), 2q24.2 deletion, 18q12.2 deletion, centromeric 3q29 deletion, and 17p13.1 duplication. We detected CNVs spanning genes associated with single gene disorders in ten cases, among which deletions of *CAMTA1, NRXN1* and *WAC* genes. Lack of phenotypic segregation for a deletion in the 5'UTR region in *MBD5*, together with similar cases described in DGV, suggested rearrangements involving this part of the gene are irrelevant to *MBD5* associated phenotype. We deeply investigated our CNVs to dissect possibly novel candidates for ID/DD. We searched for FMRP mRNA targets, synaptic proteins, histone-modifying enzymes, GO ontology enrichment. We identified 62 remarkable genes, 16 of which showed strong constraints both for missense and loss-of-function mutations (*ACTR3B, BCL11B, BRSK1, BRWD3, CACNA1D, EP400, GRIA2, KIF2A, LRRC41, NAA15, NR2E1, NRXN3, OTUD7A, U2AF2, ZNF536* and *ZNF628*). We also considered strong candidates such as *DLG2* and *ASTN2* based on literature data. We identified two potential Topological Associated Domain (TAD) boundary deletions: i) a *de novo* CNV overlapped the mesomelia-synostoses syndrome critical region (MIM 600383) in a patient without skeletal anomalies. We suggest that this disease is due to a position effect rather than to *SULF1* and/or *SLCO5A1* haploinsufficiency; ii) in a family with three cases associated with psychiatric disturbances, we showed the segregation of small deletion upstream *CTNND2* possibly altering its regulatory landscape. In conclusion, deep phenotypic and genetic analyses of our cohort allowed identifying complex cases, possibly rare disease genes involved in ID/DD, and novel position effects.

881T

Identification of *PSMD12* and *PITPNC1* as candidate genes for 17q24.2 microdeletion syndrome. *T. Gambin[1,2,3], J.A. Rosenfeld[1], B. Yuan[1], C.D. Luzzi[4], A. Gomez Gutierrez[5], E. Roeder[6], K. Stringer[7], S. Sachorow[8], C.A. Bacino[1,9,10], F. Xia[1], C.A. Shaw[1], Y. Yang[1], A. Patel[1], P. Stankiewicz[1].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 3) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 4) Memorial Behavioral and Developmental Pediatrics, South Bend, IN; 5) Instituto de Referencia Andino, Bogota, Columbia; 6) Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, San Antonio, TX; 7) University of Mississippi Medical Center, Jackson, MS; 8) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 9) Department of Medicine, Baylor College of Medicine, Houston, TX; 10) Department of Pediatrics, Texas Children's Hospital, Houston, TX.

Large genomic deletions involving 17q24.2 have been reported in patients with intellectual disability (ID) truncal obesity, mood swings, and hallucinations (PMID: 22166941, 24297458), and *PRKCA, CACNG5, CACNG4, CACNG1,* and *HELZ* have been proposed as candidate genes. In our CMA database of 59,092 patients who underwent exon targeted aCGH testing, we have found four individuals with smaller-sized deletions on 17q24.2 (including three *de novo* events). These four patients presented with variable features of developmental delay/ID, dysmorphic features, hypotonia, and failure to thrive. The smallest deletion of ~ 125 kb only involved exons 5-8 of *PITPNC1*; its inheritance is unknown. The smallest region of overlap for the two *de novo* deletions (220 kb and 580 kb) and one *de novo* del-norm-del rearrangement (620 kb) contained two genes: *PSMD12* and the first exon of *PITPNC1*. By exploring our database of >7000 clinical exomes analyzed, we found a *de novo* splicing variant g.17:65340898A>G in *PSMD12* in a patient with seizures, ADHD, learning/developmental delays, autistic features, dysmorphic features, velopharyngeal insufficiency, cryptorchidism, chordee, lymphangioma, and pineal cyst. In summary, we have identified one critical region responsible for abnormal neurodevelopment in the 17q24.2 microdeletion syndrome, which contains two candidate genes, although functional information about these genes is limited. *PSMD12* encodes a non-ATPase subunit of the 19S regulator which is a part of 26S proteasome, was found to be highly expressed in frontal cortex and has very high haploinsufficiency score (HI index=0.0557). *PITPNC1* encodes a phosphatidylinositol transfer protein that plays important roles in cell signaling and lipid metabolism, and the probability of this gene being haploinsufficient is lower (HI index=0.147).

882F

Further clinical delineation of PACS1-related syndrome due to a recurrent *de novo* **pathogenic variant.** *T. Funari[1], D. Stern[2], M. Cho[1], R. Chikarmane[1], J. Hoffman[1], R. Willaert[1], K. Retterer[1], H. Crawford[3], F. Kendall[4], M. Deardorff[5], S. Hopkins[6], E. Bedoukian[7], A. Slavotinek[8], S. Vergano[9], B. Spangler[9], M. McDonald[10], A. McConkie-Rosell[10], B. Burton[12], K. Kim[12], B. Baskin[1], H. McLaughlin[1], D. McKnight[1], J. Juusola[1], W. Chung[2,11].* 1) GeneDx, Gaithersburg, MD; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY; 3) Clinical and Metabolic Genetics, Cook Children's Health Care System, Fort Worth, TX; 4) VMP Genetics, Roswell, GA; 5) Department of Pediatrics, Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 6) Division of Neurology, The Children's Hospital of Philadelphia, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 7) Individualized Medical Genetics Center, Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 8) Department of Pediatrics, UCSF Benioff Children's Hospital, University of California, San Francisco, San Francisco, CA; 9) Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, VA; 10) Division of Medical Genetics, Duke University Medical Center, Durham, NC; 11) Department of Medicine, Columbia University Medical Center, New York, NY; 12) Division of Genetics, Birth Defects & Metabolism, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL.

Identifying the etiology of intellectual disability (ID) is challenging due to the genetic heterogeneity and clinical variability associated with these disorders. Whole exome sequencing (WES) has proven to be an effective tool for identifying the genetic etiology of ID in many patients. Using WES, *PACS1* was recently recognized as a novel gene for ID associated with autosomal dominant mental retardation syndrome 17 (MRD17) (MIM#615009). A recurrent, *de novo PACS1* pathogenic variant, p.Arg203Trp, has been reported in 19 individuals with a similar phenotype including ID, dysmorphic facial features, and urogenital abnormalities. Using WES, we have identified seven additional unrelated individuals who were heterozygous for this *de novo* missense variant in *PACS1.* All seven probands in our series also have ID, impaired language development, and overlapping dysmorphic facial features , such as hypertelorism, downslanting palpebral fissures, full/arched eyebrows, down-turned corners of the mouth, long eyelashes, simple ears, bulbous nasal tip, and widely-spaced teeth. Other common features shared among this cohort and the previously reported patients include seizures (in 6/7 and 12/19, respectively), structural brain abnormalities (3/5 and 12/16 who had imaging studies), oral aversion (3/7 and 11/19), hypotonia (5/7 and 8/19), aggressive outbursts (3/7 and 10/19), cryptorchidism (3/5 and 6/12 males), eye abnormalities (3/7 and 8/19), gastroesophageal reflux (5/7 and 6/19), and congenital heart disease (3/7 and 8/19). Two of the seven individuals also have hypertonia of the extremities in addition to central hypotonia, which has not been previously reported in affected individuals. PACS1 protein is a trans-golgi-membrane traffic regulator responsible for localization of protein cargo and several viral envelope proteins. PACS1 is up-regulated during fetal brain development and down-regulated after birth. The p.Arg203Trp variant is located in the furin cargo binding domain, adjacent to a CK2-binding motif necessary for PACS1 autoregulation. Previous studies in zebrafish embryos demonstrate that PACS1 plays a role in cranial neural crest migration. Currently, no other missense variants in the *PACS1* gene have been reported as pathogenic, suggesting a specific disease mechanism for p.Arg203Trp, possibly due to a gain of function. In summary, our results add to the clinical delineation of this new syndromic form of ID (MRD17).

883W

Understanding Xp22.31 duplication. *Y. Qiao[1], F. Tang[2], C. Badduke[1], S. Martell[1], H. Bagheri[1], S. Lewis[3], W. Robinson[3], L. Arbour[4], E. Rajcan-Sepaorivc[1].* 1) Dept. Pathology (Cytogenetics), BC Child and Family Research Institute, University of British Columbia (UBC), Vancouver, BC, Canada; 2) University of Calgary, Alberta, Canada; 3) Dept. Medical Genetics, UBC, Vancouver, BC, Canada; 4) Dept. Medical Genetics, University of Victoria, BC, Canada.

Xp22.31 duplications are reported in subjects with neurodevelopmental delay (NDD) and variable phenotypic abnormalities, including seizures. However, they can also be present in controls. In order to understand the role of the Xp22.31 duplication in NDD we preformed multifaceted studies in a family with an affected boy with NDD and an Xp22.31 duplication (7515787-8123310, hg19). This duplication encompasses only the *PNLPN4* and *VCX* genes and is inherited from his normal grandfather. The proband had significant NDD, verbal apraxia and facial dysmorphism including a short forehead, bitemporal narrowing, hypertelorism, wide palpebral fissures, flattened supraorbital ridges, a broad nasal root, thin upper lip, long philtrum, a high arched palate, malar hypoplasia, and posteriorly placed and low set ears. Digits and toes were widened distally with short distal phalangeal creases and prominent finger pads, mildly splayed distal phalanges, and pes planus. He also had central hypotonia, hyperextensibility of joints, thin torso, prominent scapulae, a pectus excavatum and prominence of the lower ribs. His mother had seizures, which were controlled in pregnancy with valproic acing (1000mg/day). The proband's phenotype had overlaps with the prenatal effect of valproic acid, but was overall considered to be more severe. The expression of the *PNLPN4* gene from Xp22.31 was increased in the proband and his mother. Exome sequencing revealed that the proband also carried a *de novo* missense mutation in *PURA,* and an X-linked missense mutation in *TMLHE,* inherited from the maternal grandmother. Both genes are associated with NDD, and *TMLHE* is also linked with seizures. Skewed X-chromosome inactivation (XCI) (94:6) was observed in the mother with preferential inactivation of maternal X (*TMLHE* mutation) and activation of the paternal X (Xp22.31 duplication). The *TMLHE* mutation carrier grandmother also showed skewed XCI (91:9). We propose that a combination of dosage sensitive genes in the Xp22.31 duplication (e.g. *PNLP4* is a lipase with a role in retinoid homeostasis, critical for development and growth), as well as additional genetic "hits" (e.g. mutations in NDD genes) contribute to the severity of the phenotype in some of the subjects with Xp22.3 duplication. More widespread use of next generation sequencing and functional analysis of Xp22.31 genes in cells from carriers could contribute to better understanding of phenotypic variability in subjects with Xp22.31 duplication.

884T

Fragile-X Syndrome: Advances in diagnosis in the public health system in central Brazil. *T.C. Vieira[1,3,4], M.A.D. Gigonzac[1,2,3], L.S. Teodoro[4], L.B. Minasi[4], C.C. Da Silva[1,2,3,4], A.D. Da Cruz[1,2,4].* 1) Human Cytogenetics and Molecular Genetics Laboratory (LaGene) Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 2) Biotechnology and Biodiversity Graduate Program, Federal University of Goiás, Campus Samambaia, Goiânia, Goiás, Brazil; 3) State University of Goiás (UEG), UnU Goiânia, Avenida Anhanguera, Goiânia, Goiás, Brazil; 4) Graduate Program in Genetics (MGene)/Replicon Research Center, Catholic University of Goiás (PUC-GO), Setor Universitário, Goiânia, GO, Brazil.

Fragile X Syndrome (FXS) is the most common cause of Intellectual Disability inherited, with an estimated prevalence of 1/4000 male births. The most common etiology of the syndrome is associated with a CGG trinucleotide expansion and methylation Xq27.3 region involving the *FMR1* gene. This disorder is commonly underdiagnosed in children and adolescents, given the high clinical variability. This study aimed to describe the implementation of a method of molecular diagnosis in patients with intellectual disability suggestive of FXS from the Public Health System in Central Brazil. Genomic dsDNA of patients was isolated and amplified by polymerase chain reaction using the *FMR1* Amplidex® Kit. The fragments were separated by capillary electrophoresis in the ABI 3500 Genetic Analyzer (Applied Biosystems®), and analyzed using GeneMarker® program. It was possible to detect changes in the *FMR1* gene, like full mutation and pre-mutations, consistent with the phenotypes observed in patients. Our study reports the results of TP-PCR of 26 unrelated individuals, 9 females and 17 males, with aging ranging from 3 to 18 years old. The normal CGG range was 8–35 repeats, and 29, 28 and 30 repeats represented the most common alleles corresponding to 20%, 11.5% and 11.5%, respectively. Two (7.7%) full mutation was identified in a male individual. The size of the fragments was 977pb with 249 CGG expansions, and 927pb with 233 CGG expansions. Moreover, pre-mutation was identified in another male patient who showed a fragment with 632pb involving an expansion of 134 CGG copies. Thus, the proposed method proved to be rapid and effective for the diagnosis of FXS, permitting genetic counseling suitable for families. On the other hand, the inclusion of this molecular testing in a Brazil public health service has allowed bridge the gap between the available technologies and effective diagnosis, universalizing access to genetic testing in central Brazil.

885F

The French HUGODIMS consortium experience on intellectual disabilities. *S. Bezieau[1], A. Denomme[2,3], T. Besnard[1], X. Latypova[1], B. Cogne[1], E. Seche[1], D. Bonneau[2], P. Parent[4], B. Gilbert-Dussardier[5], S. Odent[6], A. Toutain[7], B. Redon[8,9], S. Schmitt[1], P. Boisseau[1], HUGODIMS Consortium[10], M. Vincent[1], S. Mercier[1], B. Isidor[1], S. Küry[1].* 1) CHU Nantes, Service de Génétique Médicale, Nantes, France; 2) CHU Angers, Département de Biochimie et Génétique, Angers, France; 3) UMR INSERM 1083 - CNRS 6214; 4) CHRU Brest, Génétique médicale, Brest, France; 5) CHU Poitiers, Service de Génétique, EA 3808 Université Poitiers, Poitiers, France; 6) CHU Rennes, Service de Génétique Clinique, CNRS UMR6290, Université Rennes1, Rennes, France; 7) CHU Tours, Service de Génétique, Tours, France; 8) Inserm, UMR 1087, l'institut du thorax, CHU Nantes, Nantes, France; 9) CNRS, UMR 6291, Université de Nantes, Nantes, France; 10) Western France consortium, HUGODIMS is the French acronym standing for "Projet inter-régional Français des Hôpitaux Universitaires du Grand Ouest pour l'exploration par approche exome des causes moléculaires de Déficience Intellectuelle isolée ou syndromi.

Background: Intellectual disabilities (ID) constitute a heterogeneous group of syndromic and non-syndromic disorders of variable prevalence. The number of genes accounting for the vast majority of ID is so important that a targeted analysis is questionable. Numerous studies pointed to the relevancy of exome sequencing of patients/parents trios to increase diagnostic yield. **Purpose:** Our goal was to determine the efficiency of exome sequencing to unravel the molecular cause of ID in patients with severe phenotype seen by clinical geneticists from Western France hospitals. Method: Following a trio-based exome sequencing strategy, we investigated 75 patients with severe ID which could not been explained by fragile X syndrome, copy number variations (CGH array) or even by known candidate genes. **Results:** For almost 50% of the cases tested so far, we have been able to explain the molecular cause of the disease or highlight new ID candidate genes. **Conclusions:** Our work confirmed that the trio-based whole-exome sequencing is a powerful approach for diagnosis and research. New identified genes highlighted the importance of several signaling pathways such as the NMDAR one, in the occurrence of neurodevelopmental disorders.

886W

The evaluation of CNV in an infant with intellectual disability prenatally exposed to crack-cocaine detected by Chromosomal Microarray Analysis. *A.S. da Cruz[1,2], L.G. Oliveira[1,2], D.M.C. Cunha[1], C.L. Ribeiro[1,3], I.P. Pinto[1,3], L.B. Minasi[1,2], E.O.A. Costa[1,2], C.C. da Silva[1,2,3,4,5], A.D. da Cruz[1,2,3,4].* 1) Biology, Pontifícia Universidade Católica de Goiás, Department of Biology.Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Genetics Master's Program, Pontifical Catholic University of Goias , Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Av. Anhaguera, N° 3228, Setor Leste Vila Nova.

Prenatal crack-cocaine exposure affected brains are characterized with widespread structural and functional alterations, and are associated with deficits in intelligence, language skills, executive functioning, impulse control and attention, and evidence of internalizing and externalizing behavioral traits. Herein, we reported the genomic imbalances in a boy who his parents were prenatally crack-cocaine dependent users referred by assistants from public health system of Central Brazil. A 2-years old male proband born to non-consanguineous parents, at 39 weeks gestation and his birth weight was 2800 g. Physical examination of the proband reveled difficulty of gait in his left leg, claw left hand deformity, mild intellectual disability, and a height of 94 cm, a weight 1400 g. Both parents are crack depends users and even during the proband's pregnancy. Conventional cytogenetics analysis at about 550 band resolution using the software IKARO® (Metasystems Corporation, Germany) showed the proband with a male karyotype (46,XY), without any suggestion of chromosome alteration. Chromosomal Microarray Analysis (CMA) using Affymetrix GeneChip® CytoScanHD™ array with the reduced standard segments filters recommended by manufacturer, comprising for losses and gains a minimum of 25 and 50 markers, respectively, in a > 1kb length. The CMA revealed small CNVs between 1 to 17 kb. Three *de novo* microduplications at Xq13.3, Xq24 and Xq28 regions, and three *de novo* microdeletions involving 2q33.2, Xq26.1 and Xq27.1 regions were identify. None genes were identified in Xq13.3, Xq26.1 and Xq27.1 regions. The 17 kb microduplication in the *SLC25A43* gene at Xq24 region is widely expressed in the central nervous system, and the 10 kb microduplication in the *MECP2* gene at Xq28 region is required for maturation of neurons and mutations in this gene can cause Rett Syndrome, mental retardation, encephalopathy, and have been implicated in autism susceptibility. *De novo* small CNVs are also observed in different types of exposure to ionizing radiation and chemical agents. The authors suggest that the exposure before and during the pregnancy of crack-cocaine could increase the genomic instability and if not adequately repaired, these DNA damage would cause high frequency of small CNVs seeing in the proband.

887T

***De novo* single nucleotide and structural variants in the *CACNA1C* gene detected with array and WES in a diagnostic cohort of 10,731 patients with intellectual disability: Disease-causing or incidental findings?** *N. de Leeuw[1], L. Vissers[1], H.G. Yntema[1], C. Marcelis[1], T. Rinne[1], M. Ruiterkamp-Versteeg[1], W.M. Nillesen[1], N. Leijsten[1], S.J. Stevens[2], A.P.A. Stegmann[2], A.D.C. Paulussen[2], C. Gilissen[1], J.Y. Hehir-Kwa[1], R. Pfundt[1].* 1) Department of Human Genetics; Radboud Institute for Molecular Life Sciences and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; 2) Department of Clinical Genetics, Maastricht University Medical Center+, Maastricht, the Netherlands.

The *CACNA1C* gene is a known dominant disease gene associated with Timothy syndrome and Brugada syndrome, both characterized by heart arrhythmias with a high incidence of sudden death. So far, only missense variants have been reported in this gene that encodes the alpha-1C subunit of voltage-sensitive calcium channels present in cardiac muscle, skeletal muscle and brain. There is evidence that *CACNA1C* plays a role in brain morphology and function and genetic variations in *CACNA1C* have been found to be associated with neuropsychiatric disease, in particular schizophrenia and autism. So far, no disease-causing *CACNA1C* variants have been described in the aetiology of intellectual disability (ID). We report a total of 10 *de novo* variants involving *CACNA1C*, including one nonsense, one frameshift and three missense variants as well as one terminal and four interstitial copy number losses encompassing *CACNA1C* in five males and five females from a total of 10,731 patients with ID who were tested in our diagnostic laboratory by genome wide SNP-based array analysis (n=8,480) or Whole Exome Sequencing (WES) trio analysis (n=2,251), including Copy Number Variant (CNV) analysis in WES data. The three missense variants were detected with WES in three unrelated patients with ID, without a cardiac phenotype being reported. Because none of these missense variants had been reported in literature and were not listed in the ExAC database, these variants were considered as incidental findings (IFs) and hence anonymously discussed in our multidisciplinary committee of IFs. All three were classified as a Class 3 variant and therefore none of these were reported back to the requesting clinician by the committee, but were labelled as potential candidate gene variants for ID. Subsequent analysis of >2,000 control trio studies using WES or Whole Genome Sequencing did not reveal any *de novo* variants in *CACNA1C*. Moreover, this gene has a Residual Variation Intolerance Score of -2.09, a probability of Loss-of-Function (LoF) intolerance of 1.00, a high Z score of 6.41 (for synonymous and missense variants), and a high rank Haploinsufficiency Index score of 8.42, all indicating that the *CACNA1C* gene is (extremely) intolerant to variation. Based on our diagnostic data from genome wide array and WES analysis in a total diagnostic cohort of more than 20,000 individuals, we conclude that both structural and nucleotide LoF variants leading to *CACNA1C* haploinsufficiency are likely to cause ID.

888F

Genetic testing of idiopathic intellectual disability patients with the 505 gene custom NGS panel DI GeneProfile.
D. Diego Alvarez[1], S. Lois Olmo[2], D. Cantalapiedra de la Fuente[1], V. Felipe[1], M.L. Bellon[1], L. Perez[1], R. Perez[1], Y. Moreno[1], A. Romera[1], A. Andujar[1], S. Valverde[1], G. Hernandez[1], D. Valero[1], E. Mesa[1], M.J. Roca[1], G. Marco[2], O. Rodriguez[2], J.C. Triviño[2], C. Collado[3], V. Fernandez[3], C. Buades[1], C. Casan[1], J.A. Quesada[1], A. Perpiñan[1], D. Rocas[1], R. Martinez[1], L. Cano[1], R. Rodriguez[3], S. Santillan[1]. 1) Medical Genetics Unit, Sistemas Genomicos, Valencia, Valencia, Spain; 2) Bioinformatics Department. Sistemas Genomicos, Valencia, Valencia, Spain; 3) New Technologies Lab. Sistemas Genomicos, Valencia, Valencia, Spain.

Intellectual disability (ID) is the most common developmental disorder with a prevalence of 1-3% of the population worlwide and a lifetime estimated cost of $1 million/individual. Benefits to establishing a diagnosis for both primary care providers and families include prognosis and treatment options, as well as genetic counselling. To date, 60% of ID patients remain without a diagnosis. Most severe forms have a genetic basis, with underlying mutations occurring at scales of resolution ranging from large cytogenetic anomalies to point mutations. At this level, there is a high heterogeneity, with >500 genes associated to ID. **Statement of purpose:** With the aim to achieve a higher diagnostic yield in patients with ID, we have designed a NGS gene panel targeting 505 genes asocciated with ID. Our bioinformatics pipeline allows the detection of SNVs and indels, as well as CNVs at a single exon resolution. **Methods & samples:** The design interrogates 2.1 Mb of 505 genes, including the entire coding region plus the intronic adjacent regions. Probes were captured using SureSelect (Agilent); sequencing was performed in MiSeq/HiSeq Illumina platforms. Nucleotide variants were confirmed by Sanger. CNVs were confirmed by MLPA, QMPSF and/or variant allele ratio observation when possible. Samples included both syndromic and non-syndromic patients, as well as one couple with offspring affected with lissencephaly, and a prenatal sample with clinical suspicion of Meckel syndrome. **Summary of results:** Before testing the clinical samples, the panel was technically validated with 8 control samples (HapMap) obtaining a sensibility >99% and a specificity of 97% for SNVs/indels. Since March 2015 to date, 51 patient samples have been analyzed, including a prenatal case that in spite of its previous assumption of Meckel syndrome, was finally diagnosed as McKusick-Kaufman. Out of the total, 15 positive diagnosis have been made so far (29%) including both nucleotide variants and CNVs, represented among some of them Mental retardation-hypotonic facies syndrome , Rett syndrome, Kabuki syndrome, Bohring-Opitz syndrome, Coffin-Siris syndrome, Lissencephaly, XL Hydrocephalus, Treacher Collins, or Mental Retardation XL 90. **Conclusions:** Our NGS gene panel proves to be a reliable and efficient diagnostic platform to detect CNVs and nucleotide variants, to well oriented cases and to not so clear ones, and hence helps the diagnosis of ID patients, which has a positive impact in their lifes.

889W

Genomic alteration research in patient with idiopathic intellectual disability.
M.A.D. Gigonzac[1,2,3,5], F.G. Dos Reis[2,6], I.P. Pinto[2], L.F Teodoro[2,4], K.S.A. Silveira[7], L.B. Minasi[2,4], D.M. Silva[2,4,6,7], C.C. Da Silva[1,2,3,4,5], T.C. Vieira[1,2,4,5], A.D. Da Cruz[1,2,3,4]. 1) Lagene, Secretary of Goias State for Public Health, (LACEN/SESGO), Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 4) Graduate Program in Genetics (MGene)/Replicon Research Center, Catholic University of Goiás (PUC-GO), Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Setor Universitário, Goiânia, GO, Brazil; 6) Genetics and Molecular Biology Graduate Program, Federal University of Goiás, Goiânia, GO, Brazil; 7) Institute of Biological Sciences, Federal University of Goiás, Goiânia, GO, Brazil.

The Intellectual Disability (ID) is one of the most prevalent development disorders, which appears before 18 years of age, characterized by adaptive and cognitive limitations with significant impairment in intellectual functioning. Due to its heterogeneity, the etiology of ID is variable and difficult to identify, often referred as idiopathic intellectual disability. Several studies highlight the genetic changes as factors that significantly contribute to this condition. The aim of this study was to investigate the genomic alteration of idiopathic intellectual disability in an 11 years-old female proband. Initially we conducted the conventional cytogenetic analysis by GTG banding using the software Ikaros® (Metasystems Corporation, Germany) to identify potential chromosomal abnormalities. Subsequently, Chromosomal Analysis of Microarrays (CMA) using GeneChip® CytoScanHD (Affymetrix, USA) was carry on to identify microrearrangements. In additional, the molecular analysis of the promoter region of the *FMR1* gene at Xq27 chromosome was performed by capillary electrophoresis genetic analyzer ABI3500 (Applied Biosystems®) to check possible CGG expansions. The karyotype showed the normal female karyotype 46,XX. Chromosomal microarray analysis revealed a *de novo* microdeletion spanning ~ 4.18 Mb in Xq27.3-q28, encompassing 33 genes (*SLITRK2, TMEM257, MIR890, MIR888, MIR892A, MIR892B, MIR891B, MIR891A, CXorf51A, CXorf51B, MIR506, MIR507, MIR508, MIR514B, MIR509-2, MIR509-1, MIR509-3, MIR510, MIR514A1, MIR514A3, MIR514A2, FMR1-AS1, FMR1, FMR1NB, AFF2, IDS, LINC00893, CXorf40A, MAGEA9, MAGEA9B, HSFX2, HSFX1, TMEM185A*). The progenitor's CMA confirmed *de novo* genomic imbalances in their child. Five genes (*TMEM257, FMR1, FMR1NB, IDS,* and *TMEM185A)* presented direct relation with the phenotype of ID and neurological disorders. Furthermore, molecular analysis of the region of the *FMR1* gene identified an allele with 28 repeats and the absence of the other allele on chromosome X. This study reinforces the importance of holding the chromosomal microarray analysis and the fragile X test for detecting CGG repeats as a diagnostic method in patients with idiopathic intellectual disabilities, contributing to the description of rare chromosomal changes or not previously described in the literature.

890T

De novo KCNH1 **mutations in four patients with syndromic developmental delay, hypotonia and seizures.** *N. Matsumoto, R. Fukui, N. Miyake.* Yokohama City University Graduate School of Medicine, Yokohama, Japan.

   The voltage-gated Kv10.1 potassium channel, also known as ether-a-go-go-related gene 1, encoded by *KCNH1* (potassium voltage-gated channel, subfamily H [eag-related], member 1) is predominantly expressed in the central nervous system. Recently, de novo missense *KCNH1* mutations have been identified in six patients with Zimmermann-Laband syndrome and in four patients with Temple-Baraitser syndrome. These syndromes were historically considered distinct. Here, we report three de novo missense *KCNH1* mutations in four patients with syndromic developmental delay and epilepsy. Two novel *KCNH1* mutations (p.R357Q and p.R357P), found in three patients, were located at the evolutionally highly conserved arginine in the channel voltage sensor domain (S4). Another mutation (p.G496E) was found in the channel pore domain (S6) helix, which acts as a hinge in activation gating and mainly conducts non-inactivating outward potassium current. A previously reported p.G496R mutation was shown to produce no voltage-dependent outward current in CHO cells, suggesting that p.G496E may also disrupt the proper function of the Kv channel pore. Our report confirms that *KCNH1* mutations are associated with syndromic neurodevelopmental disorder, and also support the functional importance of the S4 domain. Acknowledgements: Saitsu H, Tsurusaki H, Sakai Y, Haginoya K, Takahashi K, Hubshman MW, Okamoto N, Nakashima M, Tanaka F are highly appreciated for their contribution to this study.

891F

Application of a targeted next generation sequencing (TNGS) strategy for the molecular diagnosis of intellectual disability. *G. Royer[1], S. Hanein[2], A. Elmorjani[1], M. Rio[1,2], C. Fourrage[1,2], J. Amiel[1,2], V. Cormier-Daire[1,2], G. Baujat[1,2], C. Bole-Feysot[2], P. Nitschke[2], S. Lyonnet[1,2], J.P. Bonnefont[1,2], A. Munnich[1,2], L. Colleaux[2], J. Steffann[1,2], G. Barcia[1,2]. 1) Genetics Dept, Hosp Necker-Enfants Malades, Paris, France; 2) Imagine Institute for Genetic Diseases, Paris, France.*

   Purpose: Intellectual disability (ID) represents the leading cause for referral in the Genetics Department of the Necker Hospital with about 7000 patients every year. Trio whole exome sequencing (WES) is a powerful approach to gain insight into the genetic bases of ID, but it remains costly and time consuming. Moreover, it can be a source of uncertain or incidental findings, and is therefore less relevant for the diagnostic setting. We developed a custom targeted next generation sequencing (TNGS) strategy for the molecular diagnosis of ID, including all genes whose mutations have already been reported in at least two unrelated patients. Methods: Genomic DNA libraries were performed using SureSelect XT Target Enrichment Reagent Kit according to the suppliers' recommendations, and were subjected to custom targeted DNA panel enrichment. 4615 regions of interest of 253 ID genes were captured with the 120-pb cRNA baits designed using the SureDesign software (Agilent). The targeted regions, encompassing 1,556 Mb of DNA, were sequenced by synthesis on an Illumina HiSeq2500, and data were analysed through the Imagine Bioinformatics core facilities. Results: We analysed 167 patients including 32 patient-parents trios, and 135 index cases. We identified pathogenic or probably pathogenic variants in 39/167 patients (23%) and variants of unknown significance in 25/167 (15%). Pathogenic or probably pathogenic variants affected genes involved in autosomal dominant ID in 32/39 cases and in X-linked ID in 7/39. Considering the variants of unknown signification, the majority affected genes involved in X-linked ID (13/25) and was associated with atypical phenotypes. Conclusions: With a diagnostic yield of 23%, TNGS is an efficient strategy for the first-step genetic screening of ID, avoiding incidental findings, and at a lower cost compared to WES. Even if targeted approaches are restricted to a few hundred genes, interpreting variants generated by TNGS is challenging. A dynamic interaction between physicians and molecular biologists is necessary to interpret TNGS results and to draw genotype-phenotype correlations. TNGS, enabling a comprehensive screening of all known genes involved in ID, offers a rapid, cost effective and relevant selection of patients available for the identification of novel genes by WES and WGS in the research setting. .

892W

Further clinical delineation of SATB-2 associated syndrome and its possible therapeutic interventions. *A.L. Williamson[1], M. Kano-Lueckerath[2], K. Oishi[1].* 1) Departments of Genetics & Genomic Sciences, Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Departments of Internal Medicine, Pediatrics, Mount Sinai Beth Israel, New York, NY.

Introduction: SATB2–associated syndrome (MIM 612313), a rare and recently recognized genetic disorder, is associated with expressive speech delay, intellectual disability and facial dysmorphism. This condition is caused by either heterozygous mutations in the *SATB2* gene or contiguous chromosomal deletion of 2q33.1 that includes *SATB2*. SATB2 is a DNA binding protein with a role in chromatin remodeling and its haploinsufficiency causes SATB2–associated syndrome. **Case:** The patient is a Japanese female who developed severe progressive hypotonia, postnatal growth restriction, and developmental delay particularly with language difficulties in early infancy. At the age of 9, she was enrolled in extensive physical, occupational, and speech therapies. On our exam at 11 years old, she had short stature (4 %ile) with low weight (0.01 %ile) with intellectual disability, mild hypotonia, and facial dysmorphism including large frontal incisors. She was able to walk and run without difficulties and communicated with well-preserved receptive language skills. **Objective:** To identify and characterize molecular cause of the patient's condition and review the clinical outcome from the interventions she received. **Methods:** Genomic DNA from the patient and her mother were used for next generation and Sanger sequencing. Retrospective chart review was done. **Results:** Whole exome sequencing identified a heterozygous variant c.715C>T, p.R239X in the *SATB2* gene, which was predicted to be a pathogenic variant with an introduction of a stop codon. This variant was presumed to be *de novo.* The review of her clinical history and physical exam revealed that the extensive therapies remarkably improved her muscle strength, emotional expression and communication skills including recognition of written words, while she is still not able to speak meaningful words. **Discussion:** This variant has been described in two out of the eight patients with mutations in *SATB2*, suggesting that the p.R239X is a recurrent hotspot mutation in SATB2-associated syndrome. Abnormal tooth development is a part of the clinical features of this syndrome. As observed in our patient, macrodontia was seen in three previously reported cases, indicating that macrodontia should a part of the clinical phenotype of the disorder. While no large-scale Cohort studies have been employed, it is speculated that intensive interventions may be beneficial to restore social and physical functions in patients with SATB2-associated syndrome.

893T

Haploinsufficiency of E3 ubiquitin-protein ligase gene *TRIP12* **causes syndromic intellectual disability.** *J. Zhang[1,2], T. Gambin[1,3,4], J. Rosenfeld[1], M. Albalwi[5], L. Al-Gazali[6], A. Shamsi[6], E. Roeder[7], L. McAuley[8], D. Roy[9], D. Manchester[10], L. King[11], W. He[1,2], F. Vetrini[1,2], P. Ward[1,2], S.W. Cheung[1,2], W. Bi[1,2], J. Lupski[1,12,13,14], C. Eng[1,2], S. Lalani[1,2], Y. Yang[1,2], A. Patel[1,2], F. Xia[1,2], P. Stankiewicz[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 4) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 5) King Abdulaziz Medical City, Riyadh, Saudi Arabia; 6) Department of Paediatrics, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates; 7) Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, San Antonio, Texas; 8) UT Southwestern Medical Center, Children's Health Children's Medical Center Dallas, Texas; 9) Tripler Army Medical Center, Honolulu, Hawaii; 10) Genetics & Metabolism, Children's Hospital, Aurora, Colorado; 11) Lauren E. King, Vanderbilt Children's Hospital, Nashville, Tennessee; 12) Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 13) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 14) Texas Children's Hospital, Houston, Texas.

Whole exome sequencing (WES) and chromosomal microarray analysis (CMA) in individuals with intellectual disabilities (ID) and autistic spectrum disorders (ASDs) have facilitated discoveries of many novel neurodevelopmental disease genes. Here, we report eight unrelated patients referred for clinical WES or CMA in whom we identified pathogenic heterozygous variants involving the *TRIP12 (*thyroid hormone receptor interactor 12) gene, encoding a member of the HECT domain E3 ubiquitin ligases family. Most recently, variants in *TRIP12* have been reported in few patients with developmental delay (DD), ASDs, language delay, macrocephaly, hypotonia, and facial dysmorphism (PMID: 24768552, 25418537, 26777411). *TRIP12* has been classified as category 3 (suggestive evidence) in SFARI autism gene scoring. A 4-year-old boy (pt 1) with global psychomotor delay, ASDs, hyperactive and destructive behavior, and dysmorphic features carried a *de novo* c.3446_3447delCA (p.S1149X). In a 2.8-year-old male (pt 2) with speech delay, dysmorphic features, and aggressive behaviors, WES revealed c.1145-2A>C; his mother was negative for this pathogenic variant, while the father was not available for study. A 9-year-old male (pt 3) with a *de novo* c.2979dupA (p.G994fs) manifested delayed motor milestones, delayed speech, autism, intellectual disability, dysmorphic features, and obesity. Clinical evaluation of a 15-year-old male (pt 4) with a *de novo* 18 kb deletion encompassing 9 exons of *TRIP12* detected by exon targeted aCGH revealed moderate DD, minor dysmorphic features, and obesity. 7-year-old boy (pt 5) presenting with cognitive delay, autism, speech delay, and learning problems had a 29 kb deletion overlapping 17 exons of *TRIP12*. Four exon deletion of *TRIP12* was identified in a 15-year-old male (pt 6) with DD. In a 4-year-old girl (pt 7) with DD, dysmorphic features, unilateral sensorineural hearing loss, and mild obesity CMA revealed a *de novo* 940 kb deletion encompassing *TRIP12* and 7 adjacent genes. Finally, 22-month-old girl (pt 8) with ID, epilepsy, autism, mixed receptive and expressive speech delay had a *de novo* deletion of 1.7 Mb, involving *TRIP12* and 3 other genes. Our data implicate *TRIP12* as playing an important role in neurodevelopment, and its haploinsufficiency resulting in ID*,* autism, and dysmorphic features.

894F

Clinical findings in a girl with 8p23.1-p22 duplication. *I.O Focsa[1], I. Streata[2], S. Serban Sosoi[2], L.C Bohiltea[1], M. Ioana[2], M. Budisteanu[3,4,5].* 1) Carol Davila University of Medicine, Bucharest, Romania; 2) University of Medicine and Pharmacy, Craiova, Romania; 3) Prof. Dr. Alex. Obregia Clinical Hospital of Psychiatry, Bucharest, Romania; 4) Victor Babes National Institute of Pathology, Bucharest, Romania; 5) Titu Maiorescu University, Bucharest, Romania.

Introduction: Unique to chromosome 8 is a region in the p arm of about 15 Mbp, which has an abnormally high mutation rate and is genetically very divergent between humans and chimpanzees. This strongly implies that the region may be implicated in the evolution of the human brain. Many of its genes and their protein products are involved in critical process of proper brain development and function: neurogenesis, neuronal migration, and synaptic connectivity. Aim: We present a case report on 9 years-old female patient with 8p23.1-p22 duplication and complex phenotype. Material and method: The child was referred in the pediatric neurology department at 2 ½ years for global developmental delay. After a full evaluation including: clinical, neuropsychiatric and psychological examinations, biological tests, EEG, heart ultrasound, cerebral CT scan, we noted: dysmorphic features (skin pigmentary anomalies, hypertelorism, low inserted columella, micrognathia, broad malformed fingers), intellectual disability, speech delay and language development. Heart ultrasound showed atrio-ventricular septal defect and left ventricular hypertrophy. CT scan revealed a mild supra- and infratentorial cortical atrophy. Cytogenetic investigation karyotype and array-based comparative genomic hybridization (aCGH) on a 60K Agilent platform have been performed. Results and discussion: A 3.97 Mb duplication at 8p23.1-p22 was identified by aCGH investigation, as a sole aberration. The duplicated region, with genomic boundaries 12039930 - 16010296 (hg19) contains about 51 genes several listed as pathogenic in OMIM: *TUSC3* (OMIM 601385) encodes a protein involved in the vertebrate plasma membrane magnesium ion transport system and has been associated with mental retardation; *DLC1* (OMIM 604258) gene playing a critical role in biological processes such as cell migration and proliferation. Only few cases with similar duplications that partially overlap the region detected in our patient have been reported so far. Some characteristics: delayed speech and language development, intellectual disability, are reported in most of the patients. Conclusion: Our patient brings new insight in the phenotype of this rare chromosomal imbalance. The phenotype of our patient will be compared with those of other patient with the same duplication previously described to further delineate the phenotype-genotype relationship.

895W

Mosaic supernumerary marker chromosome containing euchromatin 1q21.1q24.2 in a patient with global developmental delay, facial dysmorphism and diffused nevoid hypermelanosis. *H. Li[1], A. Bibb[1], J. Sloan[2], D.G. Castillo[2], W. Kim[2], B. Bunke[2], M. Hegde[2], Z.Y. Dai[2].* 1) Division of Medical Genetics, Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Emory Genetics Laboratory, Department of Human Genetics, Emory University, Decatur, GA.

Objective: To investigate the genetic etiology in a 20 month-old boy with global developmental delay, facial dysmorphism, strabismus, laryngomalacia, diffused linear and whorled nevoid hypermelanosis. Brain MRI revealed mild brainstem dysplasia. **Methods:** Given his characteristic skin finding suggesting mosaicism, we performed skin biopsy with mixed hyperpigmented and normal skin region for cytogenetic study, including single nucleotide polymorphism (SNP) microarray, chromosome analysis and Fluorescence *In Situ* Hybridization (FISH). **Results:** A low-level mosaic (~18%) 26.6 megabase (Mb) duplication of chromosome 1q21.1q24.2 was detected by Affymetrix CytoScan HD microarray. This mosaic duplication may be even larger since there are no microarray probes for the heterochromatin region of chromosome 1q12 and the first microarray probe distal to the centromere of chromosome 1 had low-level duplication. FISH using a probe RP11-137A12 located at 1q23.3 confirmed an extra signal present in approximately 34% (17/50) interphase cells analyzed. Chromosome analysis revealed a mosaic small marker chromosome, interpreted as a supernumerary chromosome containing the centromere of chromosome 1 to chromosome 1q24.2. Therefore, the low-level duplicated region is from chromosome 1q10 to chromosome 1q24.2. **Conclusion:** This is the first reported patient carrying a mosaic small supernumerary maker chromosome containing euchromatin 1q21.1q24.2. Our study emphasizes the need of combining different techniques for low-mosaicism marker identification. It also contributes to the delineation of the partial proximal trisomy 1q phenotype.

896T

Chromosomal translocation, cnv deletion and missense mutations associated with intellectual disability in consanguineous families from Jordan. *T. Froukh[1,2], X. Zhu[2], V. Shashi[3], D. Goldstein[2], E. Heinzen[2].* 1) Biotechnology and Genetic Engineering, Philadelphia University, Amman, Jordan; 2) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY; 3) Duke University School of Medicine, Durham, NC.

Intellectual disability (ID) is primarily diagnosed based on IQ less than 70 and usually associated with global developmental delay. The prevalence of ID is 1 to 3% which tends to be higher in low-income countries and in inbred communities with high consanguinity. In Jordan, the genetic bases of ID are largely unstudied, however approximately 40% of marriages are consanguineous suggesting that recessive disease may be a major contributor. In this study, we used whole exome sequencing (WES) and homozygosity mapping to identify the genetic causes of ID in ten consanguineous families from Jordan with one or more intellectually disabled progeny. In a multiplex family, paternal translocation t(5;11) (p15.33-15.1; q23.3-q25) has led to two groups of aneuploid progeny due to malsegregation. While the two groups of the progeny are intellectually disabled, they have distinct phenotypes due to the distinct deletion/duplication. A *de novo* CNV deletion with 3.5Mb on 17p11.2, previously associated with Smith Magenis Syndrome was detected in another family with clinical features in the affected child consistent with this syndrome. Two intellectually disabled probands from two different families were found to have a homozygous missense variant in genes previously implicated with ID: (1) pA710V in *GUCY2D* associated with Leber congenital amaurosis 1, and (2) p.D312N *NAGLU* causing Mucopolysaccharidosis type IIIB; both were found to have similar phenotypic features associated with these disorders. We also highlight new candidate genes in six families with autosomal recessive inheritance pattern. This study is one of the first studies to look at the genetic bases of ID in Jordan, further demonstrates the power of consanguineous families to reveal the genetic basis of disease in small patient populations, and strongly motivates larger gene discovery endeavors in the Jordan population.

897F

Use of a cen (2) CY3 PNA probe together with an FITC telomere probe for detecting telomere size differences in T-lymphocyte interphases from older people with Down syndrome with and without mild cognitive impairment. *E.C. Jenkins[1], L. Ye[1], E. Marchi[1], S. Krinsky-McHale[1], W. Zigman[1], N. Schupf[1,2], W. Silverman[3].* 1) Human Genetics, NYS Institute for Basic Research in Developmental Disabilities, Staten Island, NY; 2) Taub Institute for Alzheimer's Disease and Aging Research, Columbia University, New York, NY; 3) Kennedy Krieger Institute and Johns Hopkins Hopkins University School of Medicine, Baltimore, MD.

Jenkins et al. (2006, 2008, 2010, 2012, and 2016) have shown that it is possible to detect reduced telomere number and size (quantified by light intensity as well as physical lengths) in short-term T-lymphocyte cultures from older people with Down syndrome (DS) with mild cognitive impairment (MCI-DS) or dementia. For all of these studies, as well as for the cases described below, dementia status was determined based on longitudinal assessments of cognition and function explicitly appropriate for this population and able to distinguish adult-onset impairment from pre-existing developmental disability. In the course of our studies using a MetaSystems image analyzer in combination with ISIS software and a Zeiss Axioskop 2 plus microscope, we obtained a centromere 2 FITC probe that provided a non-telomere standard. This enabled us to develop light intensity ratios from telomere intensities/cen2 intensities, using FITC probes with metaphase chromosome preparations. However, we could not distinguish the cen (2) FITC probe from telomere FITC probes in interphase. We have now obtained a cen (2) CY3 probe (gift from DAKO/Agilent), and can measure the FITC telomere light intensity (blue green) units/cen(2)CY3 (red) light intensity units ratio as well as total FITC fluorescence (as we had previously) in interphase . We used FITC/CY3 light intensity ratios to analyze telomere length in interphases in those with compared to those without MCI-DS. Four age- and sex-matched female pairings and four age and sex-matched male pairings all showed evidence of MCI-DS associated telomere shortening ($p<.0094 – p<.000000$). This suggests the usefulness of adding the CY3 probe in studies of telomere length in this and other populations.   This paper has no conflicts of interest relating to money or contents of the article. Grant Sponsor: NYS OPWDD Institute for Basic Research in Developmental Disabilities; Alzheimer's Association, Grant numbers: IIRG-07-60558, IIRG-96-077; NIH: P01-HD035897, R01-HD037425, R01-AG014673, P30-HD024061.

898W

Inheritance patterns may not always be as expected: Diagnostic exome sequencing (DES) uncovers alterations in X-linked genes in equal amounts in males and females with intellectual disability. *S. Tang, K. Hagman, T. Cain, H. Newman, Z. Powis.* Ambry Genetics, Aliso Viejo, CA.

Diagnostic exome sequencing (DES) uncovers a positive finding in a characterized gene in roughly 30% of patients with a broad range of underlying Mendelian disorders.   Among the first 2,000 patients undergoing DES, 1,364 (68.2%) patients were referred for intellectual disability (ID). Overall among patients with ID, a positive finding in a characterized gene was identified in 28.4% (388).   The rate of positive findings in characterized genes at the time of original analysis was higher in females (30.8%; 178/578) than in males (26.7%; 210/786) (p = 0.01, chi-square test). Interestingly and unexpectedly, the rate of positive findings in characterized X-linked genes was similar in males (4.3%; 34/786) vs females (4.7%; 27/578). All the heterozygous mutations on the X chromosome identified in females arose *de novo* (when both parents are available for testing) or are *bona fide* mutations (in which at least one parent was unavailable for testing). Recurrent causative XLID genes seen in females include *MECP2*, *HDAC8*, and *WDR45*. For males, a *de novo* occurrence accounts for 41.2% (14/34) of the detected mutations and the rest 58.8% (20/34) inherited the mutation from their mother. In the latter 20 male probands, there were no strong indications of an X-linked condition in any family.   It is possible that males with ID (and unlikely that females with ID) underwent testing for X-linked genes prior to DES and thus the most prevalent and best known XLID genes have already been excluded. Alternatively, these data may support the concept of a "female protective model" for neurodevelopmental disorders. For example, an increase in number of identifiable pathogenic CNVs and SNVs in female vs male probands with autism spectrum disorders has been observed, despite the increased prevalence in males (Jacquemont, *et al.* AJHG 2014).   Our data demonstrate that the chance to find an X-linked molecular etiology is similar in males and females undergoing DES and highlight the utility of DES to detect unanticipated inheritance patterns. The data have implications for genetic counseling; highlighting that pre-test counseling should include a review of all applicable inheritance patterns. .

899T

Novel compound heterozygous variations of *FKRP* **in patient with myopathy.** *H. Galehdari[1], G. Shariati[2,3], A.H. Saberi[2,3], M. Hamid[2,4].* 1) Departments of geneticsshahid chamran uinversity, ahvaz, iran, ahvaz, Iran; 2) Narges genetic diagnostic lab; 3) Ahvaz Jundishapour University of Medical Sciences; 4) Pasteur institute, Tehran, Iran.

Myopathy is a group of muscle disorders.   We applied whole exome sequencing for detection of genetic role for a patients with symptoms including muscle weakness, motor delay, elevated CPK and muscular dystrophy. At result screening of whole exome of peripheral blood mononuclear cells genome revealed two heterozygous variation g.10504A>T and g.10663T>C in FKRP. Both variations are disease causing according to the mutation taster analysis and are pathogenic depend on predictSNP analysis. Sanger sequencing and pedigree allele segregation approved these variations as potential causative candidates. This gene has role in posttranslational modification. Previous studies demonstrated mutations in this gene have been associated with congenital muscular dystrophy, mental retardation and cerebellar cysts.

900F

RNA-seq as a tool for diagnostically intractable cases of Duchenne Muscular Dystrophy. *E. Douine[1,2], R.T. Wang[1,2], A. Eskin[2,3], H. Lee[1,3], V. Arboleda[3], C.G. Bönnemann[4], N. Khanlou[3], C.M. McDonald[6], P.B. Shieh[2,5], S. Donkervoort[4], B.L. Wong[7], J.R. Nance[4], S.F. Nelson[1,2,3].* 1) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 2) Center for Duchenne Muscular Dystrophy, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 4) National Institutes of Health, National Institute of Neurological Disorders and Stroke, Neurogenetics Branch, Neuromuscular and Neurogenetic Disorders of Childhood Section, Bethesda, MD, USA; 5) Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 6) Department of Rehabilitation, University of California Davis Medical Center, Sacramento, CA, USA; 7) Division of Pediatric Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

   Duchenne muscular dystrophy (DMD) is an X-linked progressive muscle wasting disease. Roughly 1:5000 male births are affected, making it one of the most common pediatric genetic disorders. Disease progression begins with delayed walking and muscle weakness in early childhood. Boys become wheelchair bound around age 10. Death occurs in the 20's due to cardiorespiratory failure. DMD is caused by defects in the *DMD* gene which encodes the structural protein dystrophin. Commonly observed mutations include large exonic deletions (72%) and duplications (10%), indels (5%), splice site mutations (3%) and point mutations (9%). Multiplex ligation-dependent probe amplification (MLPA) and Sanger sequencing of *DMD* exons are the most common genetic tests and collectively find a cause for 85-95% of cases. However, roughly 5% of cases remain diagnostically intractable after conventional testing. We employed RNA-Seq on a series of clinically diagnosed DMD cases where standard testing failed to identify the molecular cause despite dystrophin negative muscle biopsy and characteristic clinical presentation. RNA-Seq identified two *DMD* structural variations—a large inversion on the X chromosome which disrupted the *DMD* transcript and a translocation between intron 44 of *DMD* and chromosome 15. We also found intronic mutations leading to pseudoexon inclusion and frame shifts in two other DMD cases. Lastly, one case of genomic duplication that included exon 44 was identified despite prior normal clinical testing which leads to an out-of-frame transcript. Detection of duplication events is challenging with existing methods and highlights strength of unbiased RNA-Seq analysis. All the above allowed the creation of family specific DNA diagnostics essential for carrier testing in females within each family. Further, the results of RNA-Seq are now highly relevant for personalized genetic therapies such as emerging exon skipping treatment for DMD as the exact mRNA produced is essential to the mechanism of action of the new drugs. For instance, in two boys with different partial exonic deletions where the consequence of their genomic deletions were unclear, we found strong evidence from RNA-Seq that an unpredicted splicing event created a frame-preserving transcript in one mildly affected boy with substantial dystrophin on muscle biopsy and the introduction of a stop codon in all transcripts in a second boy with no dystrophin on muscle biopsy. These data now have therapeutic implications.

901W

Genetic analysis of mitochondrial disorder in Japanese population. *M. Kohda[1], Y. Kishita[2], Y. Mizuno[2], A. Imai[3], A. Nakaya[3], T. Hirata[1], Y. Yatsuka[2], N. Borna[2], H. Harashima[4], K. Murayama[5], A. Ohtake[4], Y. Okazaki[1,2].* 1) Div Translational Res, Saitama Med Univ, Saitama, Japan; 2) Div. of Func. Genomics & Systems Med., Saitama Med. Univ., Saitama, Japan; 3) Dept. of Genome Info., Osaka Univ. Grad. School of Med., Osaka, Japan; 4) Dept. of Pediatrics, Saitama Med. Univ., Saitama, Japan; 5) Dept. of Metabolism, Chiba Children's Hospital, Chiba, Japan.

   Mitochondrial disorder has the highest incidence among congenital metabolic disorders characterized by biochemical respiratory chain complex deficiencies. It occurs at a rate of 1 in 5,000 births, and has phenotypic and genetic heterogeneity. More than 250 genes that cause mitochondrial disorder have been reported to date. However exact genetic diagnosis for patients still remained largely unknown. Recently, we published genomic analyses for 142 juvenile patients with mitochondrial respiratory chain complex deficiencies. We identified 37 novel mutations in known mitochondrial disease genes and 3 mitochondria-related genes (*MRPS23*, *QRSL1*, and *PNPLA4*) as novel causative genes, and also achieved genetic diagnoses in 49 of 142 patients (34.5%). The high proportion of novel mutations indicates the different ethnicity in Japanese population compared with those so far reported mainly by studies on Caucasian patients. Genetic causes for a large number of cases are still unknown. Therefore, to further clarify these unsolved cases, we are currently focusing on the analysis of exonic deletions as well as SNVs/INDELs/CNVs. By increasing the number of cases analyzed by exome sequencing, we will present our latest results and discuss about the contribution of these variants for the pathogenesis of mitochondrial disorder.

902T

Diagnosis of mitochondrial deletion syndromes by concurrent mitochondrial genome and whole exome sequencing. *A.M. Balog[1], H. Cui[1], J. Higgs[1], L. Carey[1], A. Dameron[1], K. Baldwin[2], R.A. Heidenreich[2], J. Cohen[3], A. Fatemi[3], R. Bai[1].* 1) GeneDx, Gaithersburg, MD; 2) University of New Mexico School of Medicine, Albuquerque, NM; 3) Kennedy Krieger Institute, Baltimore, MD.

   Background: The diagnosis of a mitochondrial DNA (mtDNA) deletion syndrome has important implications for surveillance, prognosis, and recurrence risk. Mitochondrial disorders represent a unique diagnostic challenge due to their variable clinical features and genetic heterogeneity. Due to the phenotypic overlap between primary mitochondrial disorders and other single gene disorders, whole exome sequencing (WES) in conjunction with whole mitochondrial genome sequencing (WMGS) has proved a valuable tool for evaluation of these disorders (PMID: 26633542). While methods have been developed to extract mtDNA reads from WES sequence data, these methods are unable to detect large mtDNA deletions associated with mtDNA deletion syndromes that can be detected by independent WMGS using next-generation sequencing of long range PCR amplified DNA with two sets of tail-to-tail primers (PMID: 24901348). **Patients and Methods:** In a retrospective study, 4,000 consecutive cases submitted for WES and WMGS were reviewed for the presence of a mtDNA deletion. WES and WMGS was performed by previously published methods (PMID: 26633542). **Results:** Out of 4000 cases, 5 patients harbored diagnostic mtDNA deletions and non-diagnostic WES results. One patient, presenting with growth retardation, ptosis, thrombocytopenia, and iron deficiency anemia with a family history significant for consanguinity, harbored a 5kb deletion at approximately 43% heteroplasmy in blood, consistent with a diagnosis of Pearson syndrome. Three patients, presenting with primarily non-specific neurodevelopmental phenotypes, including autism, seizures, and developmental delay, harbored deletions between 3.9 and 10.5kb at less than 15% heteroplasmy in blood consistent with a diagnosis of a mtDNA deletion syndrome. For the fifth patient who had a history of muscle weakness, exercise intolerance, and an abnormal muscle biopsy, a muscle biopsy was submitted for WMGS. This individual harbored a 12.8 kb deletion at less than 15% heteroplasmy . None of the five patients presented with chronic progressive external ophthalmoplegia, a symptom highly suspicious for mtDNA deletion syndromes. **Conclusion:** Concurrent WMGS and WES allowed for the diagnosis of 5 patients with atypical presentation of mtDNA deletion syndromes while excluding other potential diagnoses, supporting the benefit of this approach for evaluation of patients with complex presentations, even in the absence of mitochondrial disease-specific symptoms.

903F

***De novo DNM1L* variants associated with development delay, seizures and extended lifespan.** *H. Cui[1], A. Balog[1], J. Higgs[1], L. Carey[1], E. Heise[1], D. McKnight[1], J. Juusola[1], F. Zamora[1], S. Bale[1], M. McDonald[2], A. Irani[3], L. Escobar[3], M. Tarnopolsky[4], A. Kim[5], E. Leeth[5], J. Charrow[6], A. Qais[1], S. Suchy[1], R. Bai[1].* 1) GeneDx, Inc., Gaithersburg, MD; 2) Department of Pediatrics, Duke University, Durham, NC; 3) Medical Genetics & Neurodevelopmental Center, St. Vincent's Peyton Manning Children's Hospital, Indianapolis, IN; 4) Department of Pediatrics, McMaster University Medical Center, Ontario, CA; 5) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL.

   Dynamin-like protein 1, encoded by the *DNML1* gene, is a dynamin superfamily member that regulates mitochondrial fission and has a role in normal mitochondrial function. *DNM1L* is composed of a GTPase domain, a middle domain, and a GTPase effector domain (GED). *De novo* variants in *DNM1L* have been reported in association with autosomal dominant lethal encephalopathy due to defective mitochondrial and peroxisomal fission. A dominant-negative disease mechanism has been proposed. All reported missense variants are in the middle domain of *DNM1L*, the domain that may be responsible for dimerization. Compound heterozygous loss-of-function variants in *DNM1L* were reported in two siblings with profound hypotonia at birth, who died in early infancy, indicating that complete loss of *DNM1L* function is incompatible with life. With only a few *DNM1L* variants published to date, the phenotypic and mutational spectrum of pathogenic variants in the *DNM1L* gene needs further study. Our clinical laboratory has identified 4 likely pathogenic variants in *DNML1* from 5 unrelated patients referred for whole exome sequencing. All patients presented with developmental delay and seizures or abnormal movements. Variants from 4 patients for whom parental samples were available were *de novo*, and three variants were novel. An R403C variant, published in two unrelated patients, was detected twice in our patient cohort, suggesting this amino acid is a mutation hot spot. Three identified variants, A395G, R403C and C431Y, are located in the middle domain, which has a relatively low frequency of benign variants compared to the GTPase and GED domains according to ExAC database. A Y691C variant within the GED domain was also identified in a 21 year old patient with history of hypotonia and global developmental delay during infancy, suggesting the biological importance of other *DNM1L* domains is yet to be fully understood. Although the clinical features of patients in this cohort were consistent with previously published cases, most previously reported patients with *DNM1L*-related disorders died in early infancy, whereas at the time of testing, the mean age of patients in this cohort was 9.1 years. In summary, we have detected 3 novel *de novo* variants and a recurrent pathogenic variant in *DNM1L*. This cohort doubles the number of cases reported, and indicates that the life expectancy for this disorder may be significantly greater than reported previously.

904W

Can whole genome sequencing end the diagnostic odyssey for paediatric mitochondrial disorders? *L. Riley[1,2], M. Cowley[3], V. Gayevskiy[3], T. Roscioli[3], D. Thorburn[4,5], M. Bahlo[6,7], C. Sue[3,8], S. Balasubramaniam[9], J. Christodoulou[1,2,4,5,9,10].* 1) Genetic Metabolic Disorders Research Unit, The Children's Hospital at Westmead, Sydney, NSW, Australia; 2) Discipline of Paediatrics & Child Health, Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 3) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia; 4) Murdoch Childrens Research Institute and Victorian Clinical Genetics Services, Royal Children's Hospital, Melbourne, VIC, Australia; 5) Department of Paediatrics, University of Melbourne, Melbourne, VIC, Australia; 6) Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; 7) Department of Medical Biology, University of Melbourne, Melbourne, VIC, Australia; 8) Department of Neurogenetics, Kolling Institute of Medical Research, University of Sydney and Royal North Shore Hospital, Sydney, NSW, Australia; 9) Western Sydney Genetics Program, The Children's Hospital at Westmead, Sydney, NSW, Australia; 10) Discipline of Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, NSW, Australia.

Mitochondrial disorders are the most common inherited metabolic disease. They are a heterogeneous group of disorders caused by mutations in either the mitochondrial or nuclear genome. Diagnosis is often a lengthy process and usually relies on both clinical evidence and detection of a mitochondrial respiratory chain enzyme deficiency. We are investigating the utility of whole genome sequencing (WGS), which can detect both nuclear and mitochondrial variants, in diagnosis of paediatric mitochondrial disorders. We studied an Australian cohort of 41 patients with clinical features of mitochondrial disorders, from 4 state-based paediatric genetic metabolic service centres. Mitochondrial respiratory chain enzyme deficiencies had been reported in 25 cases and normal mitochondrial respiratory chain enzyme activities in 5 cases, the remaining cases have not had enzymology. DNA was extracted from blood samples from the patients, and their parents where available. WGS was performed at the Kinghorn Centre for Clinical Genomics, Sydney, Australia. WGS sequencing libraries were prepared using Illumina TruSeq NanoHT v2.5 sample preparation kits and sequenced on Illumina HiSeq X sequencers, via 2 × 150 bp reads, with 95% of the nuclear genome covered to > 15× depth. Alignment, annotation and identification of SNVs and indels were performed using VEP and the in-house Seave pipeline. Sanger sequencing was used to confirm variants. Data analysis has been performed on 29 families to date. Presumptive causal variants have been identified in 14 cases (48%) and involved the known mitochondrial disease genes *ACAD9* (2 unrelated cases), *ECHS1*, *GFM1*, *NDUFB11*, *RARS2*, *SERAC1* and mitochondrial-encoded *MT-ATP6*, *MT-CO2,* and *MT-TL1*. A putative novel mitochondrial disease gene has been identified in one case. The remaining cases all had evidence of a mitochondrial respiratory chain disorder however, variants were identified in disease genes with no previous evidence of causing a mitochondrial disorder (*ARX*, *G6PC*, *SKIV2L*, *SLC39A8*), but causing other symptoms consistent with the patient phenotypes. These 4 cases suggest that the primary diagnosis of a mitochondrial disorder may be incorrect or that in these cases the respiratory chain enzyme deficiency may be a secondary effect. Our preliminary data indicate that WGS rapidly and efficiently identifies variants in known mitochondrial disease genes of both nuclear and mitochondrial origin, with a positive detection rate approaching 50%.

905T

Development of a targeted next generation sequencing (TNGS) strategy for the genetic diagnosis of mitochondrial diseases in paediatric population. *G. Barcia[1,2], Z. Assouline[1], M. Rio[1,2], M. Bernardelli[1], M. Magen[1], S. Hanein[2], C. Fourrage[1,2], C. Bole-Feysot[2], P. Nitschke[2], J.P. Bonnefont[1,2], A. Munnich[1,2], A. Rotig[2], J. Steffann[1,2].* 1) Genetics Unit, Necker Hospital, Paris, France; 2) Imagine Institut of genetic diseases, Paris, France.

Purpose: Mitochondrial diseases (MD) are a clinically and genetically heterogeneous group of disorders caused by mutations of nuclear DNA (nDNA) or mitochondrial DNA (mtDNA) and resulting in mitochondrial respiratory chain dysfunction. The global incidence of MD is estimated 1/4000 live birth, corresponding to 200 new patients/year in France. Clinically, childhood MD differ from adult MD because they are characterized by a more severe phenotype and by a frequent brain involvement. In childhood MD, mutations in nDNA are more frequent than in adult MD. Mitochondrial DNA mutations account for less than 10% of paediatric MD patients. As a collaborative work of the Imagine Institute and the Clinical and Molecular Genetics Unit of the Necker Hospital, we developed a custom targeted next generation sequencing (TNGS) strategy for the molecular analysis of nuclear and mitochondrial genes involved in MD. Considering the technical complexity linked to heteroplasmy and tissue specificity of mtDNA mutations, we decided to develop two separate strategy for nDNA and mtDNA TNGS. Methods: In 2015-2016, we studied by TNGS 160 patients with a clinical, biochemical, metabolic and radiological work-up suggestive of MD. The sequencing of 212 genes involved in MD was performed by capture (SureSelectXT Target Enrichment Reagent Kit, Agilent, Illumina HiSeq2500). The sequencing of mtDNA was performed for 56 patients (long range PCR, Illumina MiSeq). Results: We identified de novo pathogenic or probably damaging nDNA variants in 58/160 patients (36%), including point mutations, large deletions (8 patients) and large duplications (2 patients) affecting genes involved in mtDNA transcription and protein synthesis (18/58 patients), mitochondrial DNA maintenance (13/58 patients) and assembly of the respiratory chain complexes (13/58). Mitochondrial DNA mutations were found only in 2/56 patients (3.5%). Conclusions: With a diagnostic yield of 36% in a cohort of 160 patients, nDNA TNGS is an efficient strategy for the genetic studies of MD. Our results suggest that, in a paediatric setting, nDNA TNGS can be used as a first step in the genetic diagnosis of MD and mitochondrial DNA sequencing can be performed as a second step in patients with a negative mtDNA result. .

906F

GeneMatcher facilitates the diagnosis of an autosomal-recessive mitochondrial disease caused by biallelic mutation of the tRNA isopentenyltransferase (*TRIT1*) gene. *. E.G. Farrow[1,2], K.D. Kernohan[3], D.A. Dyment[3,4], M. Pupavac[6], A. McBride[3], T. Hartley[3], E. Sell[5], J. Majewski[6], D.S. Rosenblatt[6], E. Shoubridge[6], A. Mhanni[7], T. Myers[1], J. Kussman[8], N. Safina[2,8], C.J. Saunders[2,9], K. Boycott[3], I. Thiffault[2,9], Care4Rare Consortium.* 1) Department of Pediatrics, Children's Mercy Hospital, Kansas City, MO; 2) University of Missouri-Kansas City School of Medicine, Kansas City, MO; 3) Care4Rare Consortium, Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; 4) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 5) Division of Neurology, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 6) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 7) Department of Paediatrics and Child Health, College of Medicine, Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada; 8) Department of Genetics, Children's Mercy Hospital, Kansas City, MO; 9) Department of Pathology, Children's Mercy Hospital, Kansas City, MO.

The myriad of important roles assumed by mitochondria in cellular function such as energy production, metabolism, signaling pathways and the regulation of apoptosis, are contributing factors of the clinical diversity of human mitochondrial diseases. From a clinical standpoint, mitochondrial disorders are often suspected on the basis of clinical presentations, but the diagnostic workup is often difficult, requiring complex and extensive clinical and laboratory evaluation. Typical clinical features often include, but are not limited to myopathy, diabetes mellitus, liver failure, ataxia, seizures, encephalopathy, sensorineural deafness and bone marrow failure. The minimum prevalence of isolated or combined, genetically defined OXPHOS defects is approximately 1:5000 live births, considering only the most recurrent mutations of mtDNA. Additionally, given the extensive locus and allelic heterogeneity present in mitochondriopathies, it is not surprising that over 1,300 nuclear genes are implicated in mitochondrial function and 265 disease genes identified to date. With the growing use of whole exome sequencing (WES), genetic diagnoses of both nuclear and mitochondrial mitochondriopathies, including novel disease genes is possible. However without functional data, the demonstration of causality in a novel gene candidate, in a single family, can be challenging. One efficient method to show evidence in favor of causality is to identify two or more unrelated individuals with the same candidate gene and sharing the similar phenotype. To facilitate connections of similarly affected individuals several matching platforms have been developed (MatchMaker Exchange, Genematcher). Here we describe three unrelated individuals "matched" by candidate gene, the mitochondrial tRNA isopentenyltransferase (*TRIT1*), with the use of GeneMatcher. A disorder characterized by microcephaly, profound developmental delay, hypotonia, epilepsy and brain anomalies. Functional studies revealed decreased levels of mitochondria complex III and complex IV proteins. Importantly, we find that TRIT1 is present in patient cells in the appropriate quantity and location, indicating the effects in patient cells are due to defective TRIT1 protein function. Our findings identify a human mitochondrial tRNA isopentenyltransferase defect syndrome linked to *TRIT1*, providing insight into the roles of defective modification of mitochondrial and cytosolic tRNAs in *TRIT1*-related diseases.

907W

Novallele™ HRM-based genotyping assays for detection of mitochondrial DNA mutations. *L. Jiang, L. Papazisi, R. Ndongo, F. Mokeyo, M. Cai, J. Huuskonen, I. Knight.* R&D, Canon US Life Sci, Inc, Rockville, MD.

Background: More than fourteen diseases and other disorders are identified as potentially related to mutations of mitochondrial DNA (mtDNA). Each human cell can contain thousands of copies of mtDNA, which are highly susceptible to mutations. While homoplasmic mtDNA can be normal or mutated, heteroplasmic mutations occur only in some copies of mtDNA. Even though methods exist for mtDNA mutation detection, such as next-generation sequencing and Sanger sequencing, there is still a need for a fast, sensitive, and cost-effective detection method. The goal of our research is to develop highly sensitive and specific assays using PCR followed by high-resolution melting (HRM) analysis. **Methods**: We designed and developed five new genotyping assays targeting mutations in the mitochondrial genome. We focused on mutations associated with Leber's hereditary optic neuropathy (LHON), mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), and myoclonic epilepsy with ragged red fiber (MERRF) disease. For each target, four assays were designed and the most robust assay was selected using gradient PCR. Testing was performed using an HRM-enabled thermocycler with both genomic DNA and synthetic constructs. One small-amplicon and four unlabeled-probe assays were selected. In order to determine assay sensitivity with regards to heteroplasmic mutations, two targets were studied by varying the amount of mutant DNA from 5–100% to mimic heteroplasmy. **Results**: Our assay selection process resulted in robust and specific PCR amplification. All assays yielded unambiguous genotyping. Fast results were achieved as the high copy number of mtDNA lowered the PCR amplification threshold (approximately 15 cycles) as compared to genomic DNA. The mutant DNA detection limit in the heterozygous samples was 10% for the small-amplicon assay and 15% for the unlabeled-probe assay. **Conclusion**: Our results demonstrate that HRM has potential over Sanger sequencing, which has a mutant DNA detection limit of 15–20%. Novallele genotyping assays provide a sensitive, fast, and cost-effective method to genotype various mtDNA mutations. HRM is useful for detecting heteroplasmic mutations and provides a viable alternative to other more-costly, labor-intensive technologies. The Novallele genotyping assays are for Research Use Only. Not for use in diagnostic procedures.

908T

Next-generation sequencing analysis for mitochondrial disorders in adult Japanese patients. *Y. Okamoto, J. Yuan, A. Yoshimra, Y. Hiramatsu, M. Ando, A. Hashiguchi, H. Takashima.* Dept Neurology & Geriatrics, Kagoshima Univ Grad Sch Medical & Dental Sci, Kagoshima City, Kagoshima, Japan.

Mitochondrial disorders have a reported incidence of 1 in 5000 live births and are a clinically and genetically heterogeneous group of diseases. A correct diagnosis is challenging mainly because of the absence of clear phenotype– genotype correlations. In addition, these diseases are attributed to the mutations in the 16.6-kb mitochondrial genome, with approximately 1500 genes encoded in the nuclear genome. The development of next-generation sequencing (NGS) has revolutionized diagnostic approaches, enabling faster, more sensitive, and more efficient genetic analyses. We developed a targeted NGS-based assay of the mitochondrial DNA (mtDNA) and 180 nuclear genes associated with mitochondrial disorders. We specifically designed and performed a three-step strategy, comprising clinical and pathological screening, long-range PCR using muscle DNA and whole mtDNA sequencing with Illumina MiSeq, and gene panel analysis following whole-exome sequencing (WES). We detected mtDNA deletions in approximately 45% of our cases. Nuclear gene mutations were identified in *RRM2B*, *POLG*, *C10orf2*, and *ABCB7* in five (14.3%) of 35 cases who underwent WES. Multiple cases had mild renal failure, external ophthalmoplegia, and neuropathy. A case with an *ABCB7* mutation had sideroblastic anemia, mental retardation, and mitochondrial myopathy, which is a rare symptom. Further studies should include investigating additional nuclear genes and pathological examinations to clarify the pathophysiology. Our results indicate that neuropathy might be an important symptom of adult mitochondrial disease caused by mutations in nuclear genes. We speculate that a noticeable number of mildly symptomatic patients with a mitochondrial disease are adults. This study shows that NGS can play a major role in the molecular diagnosis of mitochondrial disorders. .

909F

Detection and analysis of partial *USH2A* deletions in patients with hearing loss. *C.A. Austin-Tse[1], D.L. Mandelker[2], S.S. Amr[1,3].* 1) Laboratory for Molecular Medicine, Partners Personalized Medicine, Cambridge, MA; 2) Memorial Sloan Kettering Cancer Center, Department of Pathology, New York, NY; 3) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Usher syndrome (US) is an autosomal recessive disorder characterized by hearing loss and retinitis pigmentosa, which accounts for up to 50% of all deaf-blindness. While patients can benefit from early diagnosis, the most common form of the disease, Usher syndrome type II (USH2), presents as isolated hearing loss in childhood and is thus not immediately identifiable in the clinic. As a result, detection is largely dependent upon molecular diagnostics. The *USH2A* gene is the largest contributor to USH2: To date, more than 700 putative disease-causing SNVs and small indels in the *USH2A* gene have been reported in patients with US. However, in about 20% of US probands that undergo *USH2A* sequencing at our laboratory, only a single, heterozygous, disease-causing variant in *USH2A* is identified. This finding supports the presence of variants that are not detectable by clinical sequencing assays (which usually target the coding and splice site regions of the gene), such as variants in intronic and promoter regions, as well as deletions or duplications spanning 1 or more exons. To better understand the contribution of copy number variants (CNVs) to the genetic etiology of US, we screened for *USH2A* CNVs in 91 probands, who were divided into 1 of 3 groups: 1) individuals carrying only a single heterozygous clinically significant variant in the *USH2A* gene, 2) those with *USH2A* variants of uncertain significance, and 3) individuals with no *USH2A* variants who had clinical features highly suggestive of US. In total, 9 large deletions (including 3 previously reported and 6 novel deletions) were identified in 11 patients (12% of the cohort). The detection rate for large deletions was highest in individuals with a previously detected pathogenic variant (45%; 9/20 individuals in Group 1). The remaining 2 deletions were identified in individuals without previously identified *USH2A* variants (Group 3). Precise mapping of the deletion breakpoints provided evidence for a template switching mechanism underlying a recurrent exon 27 deletion. Although the remainder of the breakpoints were not enriched for homologies or repetitive elements, pooling of our data with previously reported *USH2A* CNVs revealed a concentration of breakpoints within 4 introns, raising the possibility of "fragile" regions within the gene. We hypothesize that deep sequencing of these introns in US probands may uncover a new class of *USH2A* variants that evade detection by standard assays: intragenic inversions.

910W

CDH23 and MYO7A mutations identified by targeted NGS approach in an Italian cohort with non-syndromic hearing loss. *F. Cesca[1], E. Bettella[1], R. Polli[1], E. Leonardi[1], B. Sicilian[2], D. Croatto[2], A. Murgia[1,3].* 1) Department of Salute Donna e Bambino, Laboratory of Molecular Genetics of Neurodevelopment. University of Padua, Padua, PD, Italy; 2) Medical Center of Phoniatrics, Casa di Cura Trieste, Padua, PD, Italy; 3) Department of Neuroscience. University of Padua, Padua, PD, Italy.

Hereditary sensorineural Non-Syndromic Hearing Loss (NSHL) is characterized by clinically indistinguishable phenotypes and genetic heterogeneity that until recently have hampered effective early etiological diagnoses. Next Generation Sequencing (NGS) technologies offer unprecedented diagnostic capacities and hold the promise of a significant increase in knowledge about the pathogenic role of many disease-genes. With the aim of developing advanced diagnostic tools to investigate the genetic bases of hearing impairment in a population of Italian individuals mainly of pediatric age, we designed a targeted NGS panel including 59 genes strongly associated, in Caucasians, with NSHL or with syndromes (i.e. Pendred and Usher) with onset as seemingly isolated deafness. In a cohort of 16 highly selected subjects, negative for *GJB2/GJB6* mutations, we have identified 3 cases (18%) with pathogenic mutations in genes involved both in NSHL and Usher's syndrome. In two sisters, aged 16 and 11, with profound pre-lingual hearing loss, we identified compound heterozygous *CDH23* (MIM #605516; NM_022124) mutations: a missense mutation c.6530C>A (p.P2177H) and a novel splice-site mutation c.8966-1G>C in intron 59-60, that cause the complete skipping of exon 60. In a second case two novel *CDH23* alterations, c.893T>C (p.L298P) and c.924-926del, and a novel *MYO7A* (MIM #276903; NM_000260) variant, c.808C>G (p.L270V) were detected in a 12 year-old boy with moderate hearing loss and referred post-lingual onset. In the third case we detect two *MYO7A* alterations, one of which already described as pathogenic, while the second one a novel variant, c.730C>T (p.R244C) in a 12 year-old girl with post-lingual moderate hearing loss, In all the three cases, the individuals do not have family history of deafness, nor retinal alterations or vestibular dysfunction and have a normal neurodevelopment, so we can exclude a syndromic form of hearing loss. Even though from a small cohort of individuals, our data are in agreement with previous reports about the relative frequency of *CDH23* mutations in NSHL. We show how a truncating mutation, in combination with a missense mutation of *CDH23* cause autosomal recessive NSHL (DFNB12 MIM #601386). We would like to stress the value of an NGS targeted gene panel as a first tier molecular genetic tool and the great importance of combining molecular and clinical data to reach a final correct diagnosis and provide appropriate counseling.

911T

Identification of POU3F4 defects in Chinese subjects with congenital hearing loss using MPS. *J. Cheng[1], Y. Zhao[2], Y. Lu[1], H. Yuan[1].* 1) Medical Genetics Center, Southwest Hospital, Chongqing, China; 2) Peking University People's Hospital, Beijing, China.

DFNX2 is the most common form of X- linked hearing loss, which caused by various types of mutations of *POU3F4*, accounting for up to 50% of X- linked hearing loss cases. In this study, the combination of targeted exome capture and massive parallel sequencing (MPS) with 685 hearing loss-related genes was performed to identify the underlying genes for 10000+ samples with hearing loss. We identified 49 distinct *POU3F4* mutations in 57 males with congenital hearing loss. Ten reported pathogenic mutations were identified in 17 cases. Among the 39 identified novel variants, there are 17 frameshift mutations, 5 nonsense mutations and 17 missense mutations. In 57 male subjects, 11 cases exhibited typical DFNX2 temporal bone anomalies and the others did not conduct CT scan. In a three-generation Chinese family with X-linked progressive nonsyndromic hearing loss, the proband, a six years old boy, is characterized with specific nonsyndromic low-frequency sensorineural hearing loss (LFSNHL) and showed typical IP III. A novel mutation (c.1074delC) in the *POU3F4* gene causes a frameshift extension mutation(p. C358fs) which located outside of POU domain and nuclear localization signal (NLS) at the C-terminus. This mutation was found to consistently cosegregate with the deafness phenotype in the family by Sanger sequencing. All the samples with novel mutations will be validated by CT. This study expends the genetic and phenotypic spectrum of DFNX2.

912F

Collaboration of different pathogenic variations can cause Retinitic Pigmentosa. *G. Shariati[1,2], H. Galehdari[3], A.H. Saberi[1], M. Hamid[4].* 1) Ahvaz Jundishapur University of Medical Sciences, Ahvaz, KHUZESTAN, Iran; 2) Narges Genetic Diagnostics lab; 3) department of Genetics, Shahid Chamran University, Ahvaz, Iran; 4) Institute Pasteur, Tehran, Iran.

Retinitis pigmentosa (RP) is a heterogeneous group of inherited degenerative eye diseases affect the retina's ability to respond to light subsequently causes severe vision impairment.Common symptoms of RPare difficulty seeing at night and a loss of peripheral vision.Various genes variations linked to the RP etiology. Here we applied exome sequencing to a patient with RP.At result we found two single nucleotide variation g.102775563C>T in PDZD7 and g.89971970C>G in GPR38. Although GPR98/PDZD7 mutation with digenic inheritance have been reported for Usher syndrome, type 2C, surprisingly Sanger sequencing and pedigree allele segregation checking approved these variations as potential causative candidates. Bioinformatics analysis using mutation taster showed these mutation are diseases causing. Notably according to the Exome Aggregation Consortium report allele frequencies of these variations respectively is 0.001318 and 0.0004415. Both variations (GPR98 - ENST00000405460- p.Ser1796Cys and PDZD7 - ENST00000433616-p. Glu102Lys) are pathogenic according to the SIFT and PolyPhen2 tools. Diagnosis can be challenging as variouscausative genes are related to non-syndromic RP and phenotypic symptoms significantly differ between patients. In conclusion we can emphasize exome sequencing as an opportunityfor obtaining a genetic diagnosis and characterize the comprehensive spectrum of RP.

913W

Heterozygous frame shift mutation and a three exon gain in *EYS* causing Retinitis Pigmentosa detected with whole exome sequencing. *L.H. Hoefsloot[1], V.J.M. Verhoeven[1], C.C.W. Klaver[2], T. Brands[1], M. van Tienhoven[1], A. de Klein[1], W.Y. Cheung[1], B. Eussen[1], A. Kievit[1], J. Knijnenburg[1].* 1) Dept Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; 2) Dept Ophthalmology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands.

Retinitis Pigmentosa (RP) is an inherited degenerative ocular disease that causes progressive vision impairment by retinal photoreceptor degeneration. The genetic disorder is estimated to show an autosomal recessive inheritance in 50 to 60% of patients, an autosomal dominant inheritance in 30 to 40% and an X-linked inheritance in 5 to 15%[1]. Mutations in more than 60 genes are currently known to cause a monogenic form of RP. A 74 year-old male patient, affected with RP with an onset at 16-17 years, was investigated using whole exome sequencing (WES) with filtering for a set of genes known to be involved in the genetic cause of vision disorders (333 genes). Analysis revealed a heterozygous pathogenic frame shift mutation in the *EYS* gene on chromosome 6q12. Biallelic homozygous or compound heterozygous mutations in this gene are known to cause autosomal recessive Retinitis Pigmentosa-25 (RP25) (OMIM #602772). No other pathogenic single nucleotide variation was found. Subsequent copy number variation analysis of the WES data using CoNVex[2] showed a three exon gain in *EYS*, which was confirmed using SNP-array analysis. Segregation studies in the daughter of the patient showed only the inheritance of the frame shift mutation, proving the biallelic presence of the two mutations in the patient. Although the presence of a three exon gain in the genome is not sufficient evidence to conclude pathogenicity of the event, the segregation studies support the contribution of the event to the phenotype. Characterisation of the position of the duplication, in combination with expression studies of the EYS gene in the patient will provide further evidence of its pathogenicity. [1]Hartong, D. T., Berson, E. L., Dryja, T. P. Retinitis pigmentosa. Lancet 2006; 368: 1795-1809. [2]Wright CF et al., Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 2015; 385:1305-14.

914T

Exome sequencing in a Turkish family revealed an autosomal dominant juvenile pulverulent cataract. *A.S. Dogan[1], A. Yesilyurt[2], B. Yurteri[2], O. Altintas[3], A. Karalezli[4], C. Gurdal[1]. 1) Ophthalmology, Diskapi Yildirim Beyazit Training and Research Hospital, Ankara, Turkey; 2) Genetics, Diskapi Yildirim Beyazit Training and Research Hospital, Ankara, Turkey; 3) Kocaeli University, Faculty of Medicine, Kocaeli, Turkey; 4) Mugla Sitki Kocman University, Training and Research Hospital, Mugla, Turkey.*

  Introduction: Cataract is the opacity in the crystalline lens. Juvenile forms appear from birth upto 16 years old. This study was aimed to identify the mutation in a Turkish family with juvenile autosomal dominant pulverulent cataract. Phenotypically identical cataracts may result from mutations at different genetic loci and may have different inheritance patterns as well as phenotypic variability may be seen in a single family. **Methods:** The family members underwent full ophthalmic examination and systemic evaluation. There were no ocular lesions other than the cataracts, and the patients were healthy intelligently, neurologically and metabolically. Pedigree of 3 generations revealed the presence of autosomal dominant inheritance with a complete penetrance pattern. The form of the cataracts was pulverulent, in a progressive behavior. Blood samples were collected from 5 members of this family and whole exome sequencing was carried out in proband. Sanger sequencing was performed for the rest of the family members to show segregation analysis as well as confirmation of NGS result. **Results:** Whole exome sequencing in the proband revealed the c.143G>A (p.R48H) heterozygous mutation in exon 2 of the *CRYGC* gene. Rests of the affected members of the family were revealed as bearing the same mutation in *CRYGC* gene by Sanger sequencing. **Conclusions:** In the present study, we described a family with pulverulent juvenile cataract that segregated the c.143G>A mutation (p.R48H) in the *CRYGC* gene. But, the mutation has conflicting interpretation whether it is pathogenic or a SNP. A few mutations so far have been described in the *CRYGC* gene leading to autosomal dominant cataract. Mutations in the *CRYGC* gene in literature have been shown as related to different types of cataract. Only two of them have been reported associated with pulverulent cataract. Ren et al. has shown a heterozygous 5-bp duplication in exon 2 of the CRYGC gene leading to zonular pulverulent cataract (CTRCT2; 604307). González-Huerta et al. have reported the same mutation associated with autosomal dominant pulverulent cataract. Mutation segregation in the affected family members as well as previous report by González-Huerta et al. suggested that p.R48H should be considered as a disease-causing mutation and not a SNP. **Acknowledgement:** Founded by Scientific Research Project Office of Diskapi Training and Research Hospital.

915F

Phenotype-driven analysis of whole genome sequencing to find missing alleles in retinal disease. *A.P. DeLuca[1,2], S. Hiatt[3], K. Bowling[3], J.L. Andorf[1,2], D. Gray[3], C. Finnila[3], M. Thompson[3], A.V. Drack[1,2], G. Cooper[3], T.E. Scheetz[1,2], V.C. Sheffield[1,4], E.M. Stone[1,2]. 1) Stephen A. Wynn Institute for Vision Research, University of Iowa, Iowa City, IA; 2) Dept of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA; 3) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 4) Stead Family Dept of Pediatrics, University of Iowa, Iowa City, IA.*

  There are well over one hundred genes that cause inherited retinal diseases. While comprehensive exome-based genetic testing has enabled diagnoses for many across this highly heterogeneous space, there remain cases where one or more disease alleles cannot be identified. We focused on a subset of these unsolved cases where a combination of genotypic and phenotypic information reduced the hypothesis to a single gene with a single missing allele. These cases include, for instance, dominant monogenic disorders such as *PRPH2*-associated pattern dystrophy, and recessive genetically heterogeneous disorders including *USH2A*-associated retinitis pigmentosa where a single disease-causing allele has already been identified. All cases were required to have a clinical phenotype consistent with the hypothesized gene. Whole Genome Sequencing (WGS) with 30x coverage was obtained in 27 such cases, including four where we suspected Copy Number Variants (CNVs) based on exome coverage data. Breakpoints for suspected CNVs were discovered in each of these four cases. Putative disease causing variants have been identified in 12/23 of the remaining cases. Of these 16 cases, 3 had exonic single nucleotide variants (*ACO2*, *OPA1*, *PRPH2*), 10 had CNVs that removed or added exons (*ACO2*, *CERKL*, *CHM*, *CNGB3*, *FRDMD7*, *GPR98*, *MAK*, *NMNAT1*, *PRPH2, USH2A*), 3 had structural variants disrupting regulatory sequences (*PAX6*, *TYR* x2). All identified variants were confirmed via Sanger sequencing. Segregation and phase were established by sequencing family members when available. Additional work is underway on the 11 unresolved cases to establish the pathogenicity of deep intronic or intergenic SNVs and/or small indels. With a current solve rate of 16/27 (59%), WGS has proven to be a very successful tool in identifying missing disease alleles in retinal disease patients where the available phenotypic and genotypic information can narrow the pre-test hypothesis. Additionally, the narrowed hypothesis allows additional scrutiny including even manual inspection of the sequence data that would not be possible with a genome-wide hypothesis. This additional scrutiny allowed the discovery of several variants that current bioinformatic tools failed to detect. Overall, this work emphasizes the increasing importance of high-quality clinical data as WGS becomes more readily available.

916W

Newborn deafness genetic screening of 142,417 neonates in Wuhan, China. *Z. Hao, D. Fu, Q. Zuo, Y. Ming, J. Yang, Q. Huang, H. Ring, B. Ring.* Huazhong University of science and technology, Wuhan, Hubei, China.

   Objective: Almost one third of the three million people in China suffering severe deafness are children, and 50% of these cases are believed to have genetic components to their etiology. Newborn hearing genetic screening can complement Universal Neonatal Hearing Screening for the diagnosis of congenital hearing loss, as well identifying children at risk for late-onset and progressive hearing impairment. The aim of this joint academic and Ministry of Health project was to prototype a newborn genetic screen in a community health setting on a city-wide level, and to ascertain the prevalence of variation in three genes that have been associated with non-syndromic hearing loss. Methods: With the participation of 143 local hospitals in the city of Wuhan, including both maternity and general hospitals, we screened 142,417 newborns born between May 2014 and Dec. 2015. The variants GJB2 235delC, SLC26A4 919-2A>G, and mitochondrial variants in the MT-RNR1 gene 1555A>G and 1494C>T were assayed using real time PCR. Newborns found to carry a variant were re-assayed by sequencing in duplicate.Results: We developed a high throughput genetic screening process and applied it to a total of 142,417 newborns delivered at 143 hospitals in Wuhan. About 98.7% of newborns at the contributing hospitals participated in this screen. The concordance between variants found in the genotyping screen and resequencing was >99.99%. The most frequent hearing loss associated allele detected in this population was the 235delC variant in GJB2 gene. In total, 4241 (2.98%) newborns were found to carry at least one allele of either GJB2 235delC, SLC26A4 919-2A>G or the two assayed MT-RNR1 variants. 3999 (2.8%) newborns were positive for at least one of the non-mitochondrial variants, and 242 (0.17%) newborns carried a mitochondrial DNA variant. In addition, sequencing found novel mutations in the GJB2 region.Conclusions: We performed the first large scale genetic screening in central China for alleles associated with hearing loss. This study shows that GJB2 mutation is a potential important contributor to prelingual deafness in the Wuhan newborns. Further collaboration with the participating hospitals will determine the specificity and sensitivity of the association of the studied variants with hearing loss at birth and arising in early childhood, allowing an estimation of the costs and benefits of delivering newborn hearing genetic screening in a large scale community setting.

917T

Novel *ACTG1* mutation in a Mexican patient with Baraitser-Winter syndrome. *T. Barragan, J.C. Zenteno, O. Chacon, D. Lozano, I. Hindi, B. Fuerte.* Genetics, Instituto de Oftalamologia Conde de Valenciana, Mexico City, Mexico, Mexico.

   INTRODUCTION   Baraitser-Winter syndrome is a rare autosomal dominant developmental disorder characterized by brain malformations, microcephaly, ridged metopic suture, congenital ptosis, high-arched eyebrows, hypertelorism, ocular coloboma, postnatal short stature and intellectual disability. Mutations in *ACTB* and *ACTG1* genes have been associated to this entity. In this work, non previously reported clinical features are described in a Mexican patient. A novel mutation in *ACTG1* is identified   METHODS AND MATERIALS   A 1-year old male child was referred for evaluation due to psychomotor delay and congenital eye and brain anomalies. He is the child of healthy, nonconsanguineous parents, and was delivered at term. General examination showed dysmorphic features that included microcephaly with ridged metopic suture, hypertelorism, bilateral ptosis, high-arched eyebrows and short nose with a broad bridge. On ocular examination, iris, retinal and optic nerve coloboma in both eyes, and an aparent extra ocular muscle insercion anomaly were observed; cerebral malformations were identified using CTscan. Conventional karyotype analysis was normal. *ACTB* and *ACTG1* gene analysis was performed using exon-by-exon PCR amplification and direct Sanger sequencing on genomic DNA from the patient.   RESULTS   This patient presented all cardinal features of Baraitser-Winter syndrome; moreover, we identified congenital anomalies which have not been previously reported. A novel *ACTG1* transition c.176 A>G: p.Q59R was identified in probands DNA. In silico analysis, its absence in population database (ExAC Browser, 1000 Genomes and EVS) and parental DNA, and familial segregation analysis, support the pathogenicity of this variant.   CONCLUSION   This is the first clinical and molecular analysis of a Baraitser – Winter syndrome patient from Latin America. Our patient presented typical features of this entity, and to our knowledge extra ocular muscle anomalies have not been described previously. The actin gene family consists of six different isoforms with a protein sequence similarity greater than 93%. *ACTB* and *ACTG1* differ by only four biochemically similar amino acids within the 10N-terminal residues. Interestingly, p.Q59R is described as a pathogenical variant in *ACTB* gene, which suggests that due to the great homology between these protein isoforms, this amino acid position could have important functions. Our results expand the clinical and mutational spectrum of Baraitser-Winter syndrome .

918F

Identification of a novel splice mutation in choroideremia patients involved in clinical trial. *R. Mah, A. Radziwon, S. Hume, I. MacDonald. University of Alberta, Edmonton, Alberta, Canada.*

  Purpose: Choroideremia (CHM), an X-linked disorder, causes eventual vision loss in affected male patients. Although there is no effective treatment for choroideremia, gene therapy trials are in progress in two locations. Currently patients are identified through clinical examination and sequencing and copy variant analysis of the only known gene, *CHM*. Patients with the appropriate phenotype but who lack a *CHM* pathogenic variant, can have immunoblot analysis done in a research setting as a lack of the corresponding REP-1 protein is also diagnostic for choroideremia. Because immunoblot analysis is not possible for most diagnostic laboratories future gene therapy patients will require a highly sensitive diagnostic test. To promote this we are investigating REP-1 negative patients who do not have a CHM coding variant. **Methods:** Primary fibroblasts cells from two patients were grown until fully confluent. Cells were washed with PBS and RNA extraction performed. cDNA was synthesized from these RNA extractions along with gDNA extracted from the same patients' blood samples. CHM coding region primers were designed in Geneious using the *Primer3* design tool. Product size was assessed used a 1% agarose gel. Sanger sequencing of both cDNA and gDNA was performed and sequences were then aligned to CHM gene using Seqpilot software. Splice predictions performed using Alamut Visual software. **Results:** To improve our test sensitivity, we chose to examine the RNA from two patients who were confirmed both clinically and by immunoblot analysis to have CHM. Scanning the exons of the cDNA it was noted that appropriately sized products were obtained from all exons up to exon 10. The larger fragment size for exon 10 suggested an insertion of approximately 100 bp. Sequencing of the cDNA revealed a variant, c.1245-521A>G. This variant was predicted by Alamut to cause a more favourable splice acceptor and as a result, 114 bases of intronic sequence is introduced into the mRNA transcript prior to exon 10. **Conclusions:** RNA transcript analysis of two unrelated individuals revealed a novel pathogenic splice variant, c.1245-521A>G. Because these individuals are unrelated; it is proposed that this is a hotspot for mutations to spontaneously occur. This 114 bp insertion is proposed to be sufficient to cause non-sense mediated RNA decay and therefore, no REP-1 protein is produced. Inclusion of this region in a diagnostic CHM test will considerably help the overall test sensitivity.

919W

PCR SNP assay for analysis of 22q11 region in Tunisian population. *O. Kaabi, R. Louati, N. Abdellaoui, N.B. Abdelmoula. Histology, Medical University, Sfax, Tunisia.*

  22q11.2 deletion syndrome is a genomic disorder with a broader clinical and genetic spectrum. Microalterations at the 22q11.2 locus include the most frequent 3 Mb deletion and the smaller nested 1.5 Mb deletion. Non-allelic recombination events between LCRs A to H are responsible of 22q11.2 genomic rearrangements. Molecular diagnosis of 22q11.2 microdeleletion is usually made by FISH, MLPA, chromosome microarray analysis or by qPCR. Since 2003, it has been shown that a three step PCR–RFLP analysis can be useful to exclude the presence of 22q11.2 microdeletion in the typically deleted 22q11 region of 1.5 Mb with less cost and high specificity (Huber et al. 2014) and with benefit of reducing the number of patients who must be screened for 22q11.2 microdeletion (from 100 to 1 pc). But this method was developed in the basis of the presence of polymorphism in consecutive selected SNP markers located in the 1,5 Mb 22q11.2 microdeleted region and with a high heterozygosity rate in Brazilian population. Heterozygosity excluded 22q11.2 deletion, whereas homozygosity cannot exclude deletion and leaded to subsequently analysis of 22q11 region by FISH or MLPA. Here, we perform this efficient PCR SNP assay for analysis of 22q11 region in a small cohort (n=10: deleted n=2 and not deleted n=8) of Tunisian patients having a known status of 22q11 deletion by karyotyping, FISH or/and MLPA. Interrogation of SNP database showed no information about allelic polymorphisms of the 3 first steps selected SNP (rs4819523, rs4680 and rs5748411) in North African population. PCR amplification was conducted using primers designed by Huber et al. PCR amplification reactions and conditions were optimized using in Silico PCR tools and gradient melting assays. One touch down PCR reaction was finally designed for a simultaneous amplification of the first step three markers. It was possible via Nebcutter to validate rs4819523 and rs5748411 digestion profiles, using respectively HaeIII and BsrI restriction enzymes. Digestion of rs4680 by BclI as described by Huber et al. was not valid and another enzyme CviAII was then considered. Fragments generated were visualized on 4 pc agarose gel. PCR-RFLP results were concordant with FISH/ MLPA status of deleted (homozygous) and not deleted (heterozygous) patients. Further population studies for determination of heterozygosity rate of 22q11 region markers are needed before setting up such interesting method of 22q11 deletion screening in our population.

920T

Novel mutations including a large deletion in the *ARSB* gene causing mucopolysaccharidosis type VI. *K. Suphapeetiporn[1,2], C. Ittiwut[1,2], S. Boonbuamas[1], C. Srichomthong[1,2], R. Ittiwut[1,2], V. Shotelersuk[1,2].* 1) Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; 2) Excellence Center for Medical Genetics, King Chulalongkorn Memorial Hospital, the Thai Red Cross Society, Bangkok, 10330, Thailand.

Mucopolysaccharidosis type VI (MPS VI; Maroteaux-Lamy syndrome), a rare autosomal recessive lysosomal storage disease, is caused by mutations in the *N*-acetylgalactosamine-4-sulfatase (*arylsulfatase B*, or *ARSB*) gene resulting in a deficiency of ARSB activity. Here we report on clinical and molecular features of four unrelated Thai patients with MPS VI. Two were products of consanguineous marriages. The diagnosis was confirmed by biochemical and genetic tests. We performed mutation analysis by PCR-sequencing the entire coding region of the *ARSB* gene. The causative mutations were detected in all cases. Of four different mutations identified, three have never been previously described which included two missense mutations (p.C155Y and p.R388T) and a deletion encompassing exons 2 and 3. Both missense mutations were absent in 110 unaffected ethnic-matched control chromosomes and an in-house database of 180 Thai exomes. The p.C155Y and p.R388T mutations were located in the highly conserved residues. Array CGH analysis combined with direct sequencing identified the breakpoints of the large 13,788 base-pair deletion. It is the largest deletion of *ARSB* described to date in patients with MPS VI. In conclusion, this study successfully characterized four unrelated Thai patients with MPS VI and identified causative mutations in all cases. Three novel mutations in *ARSB*; two missenses and one large deletion, were reported, expanding the mutational spectrum of *ARSB* causing MPS VI. This study is also the first to characterize the largest deletion encompassing exons 2 and 3 of the *ARSB* gene.

921F

A *de novo* terminal 4.5 Mb duplication of chromosome 16q24.1-q24.3 associated with atrial septal defect, dysmorphic features and severe speech and learning disabilities. *R. Singh[1], W.A. Khan[1], C. Simotas[2], A. Babu[1], B. Xiang[1], S.A. Scott[1], N. Cohen[1].* 1) Genetics and Genomic Sciences, Mount Sinai, New York, NY; 2) Department of Pediatrics, Elmhurst Hospital Center, New York, NY.

Distal trisomy 16q is a rare disorder associated with a spectrum of clinical features and limited postnatal survival. The variability in clinical presentation is due to both the size of the 16q duplication and the partial monosomy of an additional chromosome, since the majority of reported cases have been the result of an unbalanced translocation involving chromosome 16q. We report a child with an atrial septal defect, dysmorphic features including esotropia, microcephaly, a prominent lower lip, periorbital fullness, and shortened palpebral fissures. In addition, the patient had mild central hypotonia, developmental delay, severe speech delay, and learning difficulties. Microarray-based comparative genomic hybridization (aCGH) testing of the proband detected a previously unreported terminal 4.5 Mb duplication of chromosome 16q24.1-q24.3 that includes ~90 genes, 19 of which are known Mendelian disease genes. Interestingly, metaphase fluorescence *in situ* hybridization (FISH) using probes specific to the subtelomeric region of 16q and 22q revealed that the duplicated region was localized to the acrocentric arm of chromosome 22 as a result of an unbalanced 16;22 translocation. The derivative chromosome 22 could not be unequivocally identified by standard karyotyping, and additional FISH testing using AcroP (Cytocell) probes revealed that the chromosome 22p material was intact. Notably, parental analysis by aCGH and FISH determined that neither parent was a carrier of a balanced chromosomal rearrangement involving chromosomes 16 and 22, indicating that this unbalanced 16;22 translocation occurred *de novo*. In addition to the 4.5 Mb duplication of chromosome 16q24.1-q24.3, a *de novo* 665.2 kb deletion on chromosome 5q12.1 was detected in the proband by aCGH, which includes four genes (*KIF2A, DIMT1L, IPO11,* and *LRRC70*). Larger deletions that include *KIF2A* have been reported among patients with developmental delay; however, the identified 5q12.1 deletion was determined to be of uncertain clinical significance based on copy number variation coverage in the Database of Genomic Variation. Our patient shared some common dysmorphic features with reported partial trisomy 16q cases; however, the microcephaly, atrial septal defect, severe speech, and learning difficulties unique to this informative case implicate potential candidate genes within chromosome 16q24.1-q24.3 in the pathogenesis of these specific clinical features.

922W

Langer-Giedon syndrome with 8q23.1-q24.13 large deletion caused by 3-way translocation. *K. Yeon[1], E. Seo[1,2], H. Yoo[1,3], J. Lee[1], J. Lee[4].* 1) Medical Genetics Center, Asan Medical Center, Seoul, South Korea; 2) Department of Laboratory Medicine, Asan Medical Center, Seoul, South Korea; 3) Department of pediatrics, Asan Medical Center, Seoul, South Korea; 4) Asan Institute for Life Sciences, Asan Medical Center, Seoul, South Korea.

　　Langer-Giedion syndrome, also called trichorhinophalangeal syndrome type (TRPS2) or LGCR is a very uncommon autosomal dominant genetic disorder caused by the deletion or mutations of at least two genes on chromosome 8. The missing portion of the chromosome is 8q23-q24 that includes the genes TRPS1 and EXT1. It is a condition that causes bone abnormalities and distinctive facial features. The characteristic appearance of individuals with Langer-Giedion syndrome includes multiple exostoses, short stature, epiphyses, sparse scalp hair, rounded nose, philtrum, and a thin upper lip. Now we present a case of Langer-Giedion syndrome. Case was a 5-month-old girl with bending Lt.thumb, Hypolastic digits, brachycephaly, bulbous nose, prominent alae nasi, thick nasal septum, prominent ears, and missing uvula. Cytogenetic analysis was performed according to standard procedures. Karyotype was 46,XX,der(4)t(4;19)(q27;q11),der(8)t(4;8)(q27;q22.3),der(19)t(8;19)(q22.3;q11)del(8)(q23q24.1). This patient had 3-way translocation with 8q23-q24.1 interstitial deletion. Chromosome analysis of her parents was normal. We could know that in result, her abnormality of chromosome was de novo. Multiplex Ligation-Dependent Probe Amplification (MLPA) also was performed using MLPA probemix (P064, MRC Hollans) and P245 probemix. The patient was found to have a heterozygous deletion at 8q23-q24, rsa 8q24(P064)x1 and rsa 8q24(P245)x1, that include probes for TRPS1 and EXT1. The deletion was confirmed and visualized at the molecular level by array comparative genomic hybridization (array CGH). Array CGH showed a very large deletion of 17 Mb at 8q23.1-q24.13, arr 8q23.1q24.13(108,685,855-125,700,000)x1. Most patients of Langer-Giedion syndrome are caused by the interstitial deletion of 8 to 13 Mb. But this case was 17 Mb deletion at 8q23.1-q24.13 caused by 3-way translocation. This large deletion with Langer-Giedion syndrome is first reported.

923T

Distal trisomy 15q26.3-qter: Report of three familial cases and review of the literature. *S. Basalom, M.D. D'Agostino, S. Drury, L. Russell.* Medical Genetics, McGill University Health Center, Montreal, Canada.

　　Background: Pure duplication of distal 15q is a rare chromosome anomaly characterized by pre and postnatal overgrowth, mild to moderate intellectual deficiency and speech impairment, facial dysmorphisms, renal and cardiac anomalies and cranial synostosis. We report the clinical features of three patients in one family having pure duplication of distal 15q26.3-qter. **Clinical cases:** The proband, patient A, is an 8-year-old male. His birth weight was above the 85th %, length at 99.9th % and head circumference ~ 97th %. He presented at 5 years of age with severe expressive and receptive language delay and delayed motor skills. On physical exam the weight was ~ 85th %, height above the 85th % and head circumference ~ 75th %. Minor dysmorphic features included long face, pointed chin, tall/prominent forehead and deep-set eyes. Abdominal ultrasound was normal while echocardiography showed insignificant cardiac anomalies. The proband's father, patient B, reported symptoms of attention deficit hyperactivity disorder. His height was at the 90th % and head circumference at the 98th %. He also had mild scoliosis, myopia and astigmatism. The proband's paternal half-sister, patient C, reportedly had a mild expressive language delay that has not yet been formally assessed. Her birth weight was at the 75th %, birth length 97th % and head circumference 90th %; at 3 1/2 years her weight was again at the 75th % with height at the 85-98th % and head circumference at the 90th %. She was mildly, diffusely hypotonic and had a prominent forehead and frontal hair upsweep. Echocardiography and abdominal ultrasound were normal. The proband's chromosome microarray revealed a 3.084 Mb duplication of 15q26.3-qter (97004262-100088571). These results were confirmed by fluorescence in situ hybridization (FISH). The same duplication was found in the father (patient B) by FISH and in the sister (patient C) by aCGH and FISH. **Discussion:** We have identified three members of one family with pure trisomy 15q26.3-qter, a region that includes the *IGF1R*. To the best of our knowledge, these cases have the smallest region of distal trisomy 15q reported. Each of these cases exhibits overgrowth and/or macrocephaly, with variable dysmorphic features. We compared these cases to the cases of pure trisomy 15q reported by Chen et al, 2011, and Roggenbuck, et al, 2004. Overgrowth was noted only in those cases in which the trisomic region appeared to include *IGF1R*.

924F

***POLR1D* deletion causes Treacher-Collins syndrome 2 in a family with incomplete penetrance.** *M.J. Alshammari[1,2], S. Zeesman[1], M.J Nowaczyk[1].* 1) McMaster University, Hamilton, ON, Canada; 2) King Saud University, Riyadh, Saudi Arabia.

Treacher Collins syndrome is a mandibulofacial dysostosis that resultes from mutations in genes that contribute to ribosomal synthesis and function. *TCOF1* mutations are detected in the majority of the patients. Recently, *POLR1D* and *POLR1C* genes have been reported in *TCOF1* negative patients. with different manner of inheritance. We report a 30 year-old woman who was diagnosed with Treacher-Collins syndrome at age 18 months: she had micrognathia, zygomatic depression, partial cleft palate, bilateral small, dysplastic posteriorly rotated, low seat ears with poor development of superior pinna and chronic ears infection. She had hearing impairment that required hearing aids (Fig.1). She had normal appearing eyelids with no coloboma and no reduction of her eyelashes. Her growth was normal and she had good academic performance. Her mother had recurrent ears infection but otherwise normal. She had three apparently healthy sisters. However, her nephew had multiple ear infections, narrow ear canals and facial features suggesting a mild form of TCS (Fig. 2). *TCOF1* sequencing was normal. SNP oligonucleotide array analysis identified a 52-kb deletion within chromosomal region 13q12.2, extending from 28,192,790 to 28,245,365 that resulted in the deletion of the entire *POLR1D* gene and partial deletion of *LNX2* gene. This finding was further confirmed by quantitative PCR using primers targeted to chromosome region chr13:28,192,790-28,245,365. Quantitative PCR showed one copy of the loci examined that consistent with SNP oligonucleotide array. High density targeted array revealed a deletion of at least 47.1 kb (g.28195500_28242550) including *POLR1D* gene and this deletion extends from 5' of the *POLR1D* gene to 3' of the gene therefore all *POLR1D* exons are deleted. The same molecular result was found in the proband's mother, her phenotypically normal sister, and her affected nephew. This is the third report of *POLR1D* mutation causing an autosomal dominant inherited Treacher-Collins syndrome and it reinforces the concept of genetic heterogeneity and incomplete penetrance of Treacher-Collins syndrome. It emphasizes in the importance of testing apparently phenotypically normal TCS first degree relative for appropriate genetic counseling and to investigate further the possible factors that contribute in *POLR1D* expression.

925W

13q22.2q34 tetrasomy mosaicism due to an inverted duplication with a neocentromere. *T. Hattori[1], T. Togawa[2], Y. Togawa[1], H. Kawabe[1], T. Kato[3], H. Kurahashi[3], S. Saitoh[2], M. Kouwaki[1], N. Koyama[1].* 1) Department of Pediatrics, Toyohashi Municipal Hospital, Toyohashi, Japan; 2) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 3) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan.

Small supernumerary marker chromosomes (sSMC) comprise a heterogeneous group of structurally rearranged chromosomes origin of which is not determined by conventional cytogenetics. sSMCs lacking endogenous centromere regions are supposed to be maintained during mitosis through a neocentromere formation. Partial tetrasomy 13 due to inverted duplication with a neocentromere has been rarely reported in literature. Here, we report a patient with partial tetrasomy 13 mosaicism due to an inverted duplication of 13q22.2q34 with a neocentromere, which has never been reported. The patient, a Japanese male neonate, was born at 38th gestational age from a 23-year-old healthy mother. His birth weight, length, and head circumference were 3370g, 50cm, and 33cm, respectively. He showed hemangioma on the mid-forehead, low set ears, cleft lip and palate, postaxial polydactyly of the right hand and bilateral feet, laryngomalacia, and umbilical hernia. A brain MRI showed hypoplasia of cerebellar vermis and no sign of holoprosencephaly. G-banding chromosomal analysis from his peripheral blood leukocytes demonstrated 47,XY,+mar[2]/46,XY[28]. Both parents showed normal karyotypes, indicating that sSMC arose de novo. Multicolor fluorescent in situ hybridization (FISH) revealed that the sSMC was derived from chromosome 13. Because inverted duplication was suspected by the G-banding staining pattern of the sSMC, we further performed FISH analysis with dual probes detecting 13q14 and 13q34 on skin fibroblasts established from the patient. Forty of 50 analyzed cells showed four signals of 13q34, while no 13q14 signal was detected on the sSMC. Remaining 10 analyzed cells appeared normal having both 13q34 and 13q14 signals. Therefore, the sSMC did not have an endogenous centromere and was supposed to have a neocentromere. We performed Affymetrix's CytoScan™ HD SNP array to evaluate the range of duplication. Copy number elevation of approximately 39Mb corresponding to 13q22.2 to the q-terminal was identified. Accordingly, we confirmed the cytogenetic diagnosis of 47, XY,+inv dup(13)(q34→q22.2::q22.2→q34).arr[hg19]13q22.2q34(76,134,983-115,107,733)x4. To our knowledge, this is the first report of 13q22.2q34 tetrasomy due to an inverted duplication with a neocentromere. Clinical features of the patient are well corresponding to those with trisomy 13. Our findings can offer additional information to understand phenotype-genotype correlation in patients with partial tetrasomy 13.

926T

Partial trisomy 5q and partial monosomy 13q result in developmental delay with characteristic facial features. *L. Dupuis[1], M. Shago[2], R. Mendoza-Londono[1].* 1) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada.

We report a 12 year old female of Chinese descent with trisomy 5q33.3q35.3 and monosomy 13q34 derived from a paternal t(5;13)(q33.3;q34). Genomic microarray analysis revealed an estimated 23.2 Mb terminal duplication of chromosome region 5q33.3 to 5qter that involved approximatively 217 RefSeq genes and 27 OMIM genes. The terminal deletion of chromosome regions 13q34 to 13qter was estimated to be 1.765 Mb. This deletion resulted in the loss of one copy of 33 RefSeq genes including 5 OMIM genes (*F7, F10, PROZ, GRK1*, and *CHAMP1*).   A review of the literature revealed one case of a 2 year old female of Japanese descent with distal trisomy 5q and monosomy 13q from a maternal translocation (Shiihara *et al*., 2004). Phenotypic similarities between these two cases are striking. Craniofacial findings shared with patients with distal trisomy 5q include hypertelorism, flat midface, thin upper lip with a long simple philtrum, low set posteriorly rotated ears, and small mouth. Syndromic craniosynostosis and skull deformity have been linked to chromosome 5q35 and are due to duplication of the *MSX2* gene in the region. Interestingly, congenital heart disease has been associated with trisomy of 5q distal to 5q31. Our patient exhibited a small patent foramen ovale, whereas the previously reported patient had an ASD and PDA. Both of the patients had moderate developmental delay, particularly in areas of language and fine motor skills, but no major neurological deficits. Congenital heart defects have also been associated with 13q terminal deletions, including one patient with a deletion of 1.1Mb (*Yang et al*., 2013).   Recognition of this characteristic clinical phenotype may facilitate the diagnosis, understanding of the natural history and provision of adequate counseling to parents with similar balanced translocations.

927F

Copy number variation analysis identifies novel genomic regions associated with aberrant developmental phenotypes. *S. Chettiar[1], A. Patel[2], J. Dattani[3], D. Jhala[4], V. Sharma[5], U. Radhakrishna[6], M. Rao[4], N. Jain[1].* 1) Department of Life Science, Gujarat University, Ahmedabad, Gujarat, India; 2) GeneXplore Diagnostics and Research Centre, Ahmedabad, India; 3) NVBDP, Ministry of Health and Family Welfare, Government of India; 4) Department of Human Genetics, Gujarat University, Ahmedabad, India; 5) Department of Bioscience, Veer Narmad South Gujarat University, Surat, India; 6) Department of Obstetrics and Gynecology, Oakland University- William Beaumont School of Medicine, Royal Oak, MI, USA.

Copy-number variation (CNV) is the most common type of genomic structural variation in humans, and plays an important role in the etiology of many diseases such as cancers, psychiatric disorders, *congenital anomalies, r*ecurrent pregnancy loss and developmental disorders. These variations due to deletions and/or duplications of chromosomal segments alter the genome structure and also significantly affect gene expression levels. Twenty five amniotic fluid samples from individuals with a history of various developmental anomalies observed during radiological examination were analyzed. The average age of mothers was 30 years. Conventional cytogenetic analysis was performed, followed by high-resolution array Comparative Genomic Hybridization (aCGH) for the identification of copy number variations (CNVs). Using these combined techniques for the detection of structural variation, we discovered eight *genomic regions* associated with developmental disorders. In two cases with *echogenic intracardiac focus,* duplication of Chr19p13.11-p12*,* was detected and in three cases with short nasal bone, *retrognathia*, and *symmetrical intrauterine growth retardation,* a microdeletion on chromosome 4q35.2 and a duplication at chromosome 15q11.1-11.2 were observed. In addition, one case with atrioventricular septal defect showed a deletion at Chr8p11.22 region. Three different cases with bilateral *Ventriculomegaly* and facial dimorphism had a deletion at Chr3p26.2-26.1, ChrX22.33 and a duplication of Chr10q26.3. The present data suggest these copy number variations are one of multiple factors contributing to the development of *developmental* disorders in children. The current study emphasizes the utility of microarray based analysis in aberrant radiological observation.

928W

Identification of deletions during diagnostic massive parallel NGS gene panel sequencing. *S. Hinreiner[1], M.R. Hehr[1], N. Koch[1], A. Schweiger-Kabesch[2], I. Bader[3], B. Schmitt[4], S. Gebre-Medhin[5], S.M. Herbst[1], U. Hehr[1].* 1) Center for and Department of Human Genetics , Regensburg, Bayern, Germany; 2) Department of Pediatric Neurology, Klinik St. Hedwig, University Children's Hospital, Regensburg, Germany; 3) Division of Clinical Genetics, Department of Pediatrics, PMU Salzburg, Austria; 4) Department of Neurology, Children's Hospital, Zürich, Switzerland; 5) Department of Clinical Genetics, Skane University Hospital, Lund, Sweden.

   Introduction Massive parallel sequencing of multi gene panels (MGPS) allows fast and cost-effective genetic diagnostic testing including sequence analysis of genes with small contribution for a wide variety of phenotypes. The additional detection of copy number variations (CNVs) by MGPS within the analyzed genes can be expected to further improve the diagnostic yield, but so far is limited due to technical and bioinformatic aspects.   **Methods** 81 patients were sequenced with phenotype specific gene panels on a MiSeq platform (Illumina) and analyzed with our bioinformatic diagnostic workflow including a quantitative data assessment using our *in house* JAVA based bioinformatic script to search for gene or exon deletions. Detected deletions were confirmed by an independent method (e.g. MLPA, PCR amplification of the junction fragment), if available.   **Results** We here report details for five suspected deletions in six patients detected by our *in house* bioinformatic workflow, five of them currently independently confirmed: heterozygous gene deletions of *PAFAH1B1, Spastin* or *ARFGEF2,* respectively; two heterozygous *CFTR* deletions of exon 2 and 3 and one homozygous partial *SFTPB* deletion. The average coverage for the respective target reads was 84.16 for the controls (standard deviation 28.78) and 47.28 within the assumed heterozygous deletions. The complete gene deletions of *PAFAH1B1* and *Spastin* as well as the *CFTR* deletion of exon 2 and 3 could be confirmed by MLPA. For the partial *SFTPB* deletion both breakpoints could be precisely located within the read-out, allowing determining correct deletion size and design of primers to amplify the junction fragment.   **Discussion** Diagnostic multi gene panel sequencing after Nextera enrichment allows sufficient homogeneity of the obtained patient and control data per target to quantitatively search for constitutional deletions covering two or more adjacent targets. Combined data assessment considering the individual clinical data will not only further increase the diagnostic yield but can also be expected to further delineate the mutational spectrum for specific phenotypes by the simultaneous detection of clinically relevant sequence variants as well as copy number variations.

929T

Research reanalysis of clinical whole exome sequencing: Opportunities and discoveries. *E. Bhoj[1], D. Li[1], M. Harr[1], E. Chishom[2], S. Vergano[2], E. Zackai[1], H. Hakonarson[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Children's Hospital of the King's Daughters, Norfolk, VA.

   Clinical whole exome sequencing (WES) has entered the mainstream of genetic testing recently with thousands of patients testing in dozens of laboratories. There is some variability in the rates of achieving a diagnosis using exome sequencing, but regardless of testing laboratory, indication for testing, and analysis methodology, over half of the tests are non-diagnostic. Therefore, there are likely patients with a molecular diagnosis who are not being captured in their original exome sequencing analysis. These extra findings may be related to strengths and weakness of a particular analytic pipeline, or the result of undiscovered genes that contribute to the phenotype. Additionally, there is not a clear recommendation governing periodic reanalysis of the sequencing data, or a universal forum for collection of this valuable data. Therefore, clinicians are left with few options when a patient does not achieve a diagnosis through exome sequencing. At the Center for Applied Genomics at the Children's Hospital of Philadelphia we have an IRB-approved program to provide research-based reanalysis of whole exome/genome sequencing that was non-diagnostic at a clinical reference laboratory. After the families are consented, the bam and vcf files from the original sequencing are requested from the reference laboratory. We then analyze those data using a pipeline optimized both for known genes, and especially for gene discovery. Over thirty families from multiple references laboratories and clinical centers have been enrolled in the project to date. In our initial analysis we have found about a 20% diagnostic or strong candidate gene identification rate. In addition, we have been able to identify trends in those cases with a newly positive result, including application of phenotypic information to the analysis, and integration of multiple scientific databases during the analysis. We will highlight strategies for referring clinicians to communicate effectively with sequencing laboratories to improve diagnostic efficiency. Also, options to partner with laboratories that can provide follow up functional testing of novel variants will be discussed. Two informative cases from the project will be presented (with blinding of the original reference laboratories); one with the identification of a known gene and one with the discovery of a confirmed novel disease gene, TBCK-related intellectual disability.

930F

Miller-Dieker syndrome due to a 5.5 Mb 17p deletion from a Y-17 dicentric chromosome. *F.T. Bellucco[1], N. Nunes[1], M.E. Colovati[1], A.C. Malinverni[2], T.P. Caneloi[1], M.F. Soares[3], A.B. Perez[1], M.I. Melaragno[1].* 1) Department of Morphology and Genetics, UNIFESP - Universidade Federal de São Paulo, São Paulo, Brazil; 2) Department of Pathology, UNIFESP - Universidade Federal de São Paulo, São Paulo, Brazil; 3) Department of Diagnostic Imaging, UNIFESP - Universidade Federal de São Paulo, São Paulo, Brazil.

Miller–Dieker syndrome (MDS; OMIM 247200) is a contiguous gene deletion syndrome characterized by a severe type of lissencephaly, significant facial dysmorphism and, occasionally, other congenital anomalies, such as renal, gastrointestinal, and cardiac defects. We report a male infant with clinical phenotype of MDS including lissencephaly, corpus callosum hypoplasia, microcephaly, bitemporal hollowing, prominent forehead, short nose/upturned nares, broad nasal bridge, epicanthal folds, low-set ears, protuberant upper lip, high-arched palate, prenatal and postnatal growth deficiency, heart malformation, genital anomalies and transverse palmar creases. The patient presents 45 chromosomes and a derivative Y-17 chromosome. Array-CGH showed a *de novo* 5.5 Mb 17p13.3 deletion and no deletion in chromosome Y. C-banding showed the Yq distal heterochromatin attached to the 17p breakpoint. Fluorescent in situ hybridization (FISH) with *SRY* probe showed positive signal in the extremity of the dicentric chromosome revealing the fusion of Yq to 17p. The final cytogenomic result is: 45,X,dic(Y;17)(Ypter→Yq12::17p13.2→17qter). arr[hg19] 17p13.3p13.2(59,739-5,464,323)×1. The deletion involves more than 100 genes including the *PAFAH1B1* gene and the CRK and *YWHAE* genes, which contribute to cortical development and the severity of MDS. Almost all patients with MDS present *de novo* cytogenetic visible 17p13.3 deletions or microdeletions encompassing the *PAFAH1B1* gene along with several nearby genes including the *YWHAE* gene. Our patient presents a large deletion, a rare finding in the literature in MDS patients. Most patients with 17p deletion and unbalanced translocations show a combination of MDS and additional phenotypes because of the involvement of segments of various autosomes. Pure partial monosomy derived from balanced translocation is relatively rare in the literature. Our patient presents a Y-autosome translocation, a rare cytogenetic finding with frequency of 0.01%. At least 34 cases of 45,X male with Y-autosome translocation have been reported in the literature, none of them involving the chromosome 17. Only four cases with Y-autosome translocation associated with genetic syndromes were described: Jacobsen syndrome, DiGeorge syndrome, cri-du-chat and Dandy-Walker malformation with cerebral atresia. This is the first case of Miller-Dieker syndrome that showed a pure 17p deletion, derived from a t(Y;17) translocation.

931W

Molecular karyotyping in patients with SRS features: A powerful tool to increase diagnostic yield. *K. Eggermann[1], J. Sachwitz[1], R. Meyer[1], G. Fekete[2], M. Elbracht[1], T. Eggermann[1].* 1) Inst Human Genetics, RWTH Aachen, Aachen, Germany; 2) Dept. of Pediatrics, Semmelweis University, Budapest, Hungary.

Silver-Russell syndrome (SRS, RSS) is a growth retardation syndrome characterized by intrauterine and postnatal growth retardation, relative macrocephaly and protruding forehead, body asymmetry and feeding difficulties. Nearly 50% of cases with a characteristic phenotype show a hypomethylation of ICR1 in 11p15.5; maternal uniparental disomy of chromosome 7 (upd(7)mat) can be detected in 10% of patients. After exclusion of ICR1 hypomethylation and upd(7)mat, we analysed 44 individuals referred for SRS genetic testing by molecular karyotyping. Pathogenic variants were detected in six of them, including a duplication of the maternal Prader-Willi/ Angelman region (PWAS-CR) in 15q11.2-12, a duplication of NSD1 in 5q35 and a 14q32 microdeletion. The clinical picture of patients with 15q11.2-12 microduplications may vary considerably but among other features growth retardation, hypotonia, macrocephaly, epicanthus and clinodactyly are reported, all features are also described in SRS patients. NSD1 deletions lead to overgrowth syndromes and vice versa NSD1 duplications lead to growth retardation. The 14q32 deletion is typically associated with Temple syndrome (TS14) and this reflects the clinical overlap between TS14 and SRS. Determination of the molecular subtype is crucial for a directed counselling and therapy. Using molecular karyotyping, we identified pathogenic variants in >13% of unclassified patients referred for SRS testing. Among those with a characteristic SRS phenotype more than 16% showed pathogenic variants. We therefore conclude that in addition to 11p15.5 alterations and upd(7)mat, 6-7% of all SRS patients carry submicroscopic chromosome aberrations and we strongly suggest applying molecular karyotyping after exclusion of the more frequent aberrations.

932T

The presence of two rare genomic syndromes, 1q21.1 deletion and Xq28 duplication, segregating independently in a family with intellectual disability. *K. Ha*[1,2]*, H.G. Kim*[2,3]*.* 1) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 2) Section of Reproductive Endocrinology, Infertility & Genetics, Department of Obstetrics & Gynecology, Augusta University, Augusta, GA; 3) Department of Neuroscience and Regenerative Medicine, Augusta University, Augusta, GA.

1q21.1 microdeletion syndrome is a rare contiguous gene deletion disorder with *de novo* or autosomal dominant inheritance patterns and its phenotypic features include intellectual disability, distinctive facial dysmorphism, microcephaly, cardiac abnormalities, and cataracts. *MECP2* duplication syndrome is an X-linked recessive neurodevelopmental disorder characterized by intellectual disability, global developmental delay, and other neurologic complications including seizures later in life. Since these two syndromes are genetically unrelated, it has not been reported for these two different genetic syndromes segregating independently in a same family. Here we describe two siblings carrying a chromosome 1q21.1 microdeletion and chromosome Xq28 duplication, respectively. Using a comparative genomic hybridization (CGH) array, we identified a 1.24 Mb heterozygous deletion at 1q21.1 from position 146.5-147.78 Mb resulting in the loss of 12 genes in a girl with hypothyroidism, short stature, sensory integration disorder, and soft dysmorphic features including cupped ears and a unilateral ear pit. We also characterized a 508 kb duplication at Xq28 encompassing *MECP2* in the younger brother with hypotonia, poor speech, cognitive and motor impairment. The parental CGH and quantitative PCR (qPCR) analyses revealed that the 1q21.1 deletion in the elder sister is *de novo*, but the Xq28 duplication in the younger brother was originally inherited from maternal grandmother through the mother, both of whom as an asymptomatic carrier. RT-qPCR assays revealed that the affected brother has almost double amount of *MECP2* mRNA expression compared to other family members of both genders including maternal grandmother and mother who have the same Xq28 duplication with no phenotype, suggesting the X chromosome with an Xq28 duplication in the carrier females is preferentially silenced. We also discussed the possible involvement of other genes in the phenotypes of these two rare syndromes according to the relevant literature.

933F

Mosaicism identified by SNP array analysis in a large cohort of pediatric patients. *M. Luo*[1,2]*, J. Fan*[2]*, S. Mulchandani*[2]*, J. Chen*[2]*, B. Thiel*[2]*, M. Dulik*[2]*, E. Zackai*[3,4]*, N. Spinner*[1,2]*, L. Conlin*[1,2]*.* 1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Philadelphia, PA 19104; 2) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; 3) Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104; 4) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104.

Mosaicism denotes the presence of two or more populations of cells with different genomic contents. Abnormalities can arise from either a post-zygotic error or a pre-zygotic error corrected post-zygotically. Mosaicism complicates the clinical interpretation and following genetic counseling. We have previously demonstrated the clinical utility of single-nucleotide polymorphism (SNP) array analysis in the detection of mosaicism involving ≥5% of cells. In order to investigate the contribution of mosaicism in pediatric genetic diagnostics, we reviewed the SNP array data from patients received by our clinical laboratory for the diagnosis of suspected constitutional genetic disorders since May 2008. In this retrospective analysis of 11,659 patients, mosaic changes were detected in 195 cases (1.67%). Of these 195 cases, 161 carried the mosaic abnormalities associated with the patients' phenotypes, which also accounted for about 10% of all our positive cases. The remaining 34 cases were found to have mosaic findings with unknown significance. Among all the cases, seven cases were characterized as a genome wide mosaicism, caused by genome-wide uniparental disomy (UPD), chimerism or twin-twin transfusion. Aneuploidy with or without additional chromosomal abnormalities was the most common finding and was detected in 69 cases, with 42 cases involving only sex chromosomes, 26 involving only autosomes, and 1 affecting both. Single or multiple copy number variations affecting only part of a chromosome, either as an interstitial or a terminal change, were identified in 58 cases. Three cases involved mosaic unbalanced translocations. Partial chromosome loss of heterozygosity (LOH) suggesting segmental UPD was identified in 48 cases with the majority involving the short arm of chromosome 11. Whole chromosome UPD was seen in 7 cases, with two of them also associated with aneuploidy. In addition to these proband samples, three parents with germline mosaicism were identified. The possible structural abnormalities associated with the mosaic findings, the underlying mechanisms, and their tissue distribution were further investigated by additional clinical tests including chromosome and FISH studies. Our analysis provides an overview of the mosaic disorders detected in pediatric patients. This study demonstrates that mosaicism is much more frequent than previously recognized and the detection of mosaicism increases the diagnostic yield of chromosome array analysis.

934W

Molecular dissection of germline chromothripsis in a developmental context. *S. Middelkamp[1], S. van Heesch[1,2], K. Braat[1], J. de Ligt[1], M. van Iterson[3], M. Simonis[1], M. van Roosmalen[1], M. Kelder[4], E. Kruisselbrink[1], R. Hochstenbach[1], N. Verbeek[1], E. Ippel[1], Y. Adolfs[1], J. Pasterkamp[1], W. Kloosterman[1], E. Kuijk[1], E. Cuppen[1].* 1) University Medical Center Utrecht, Utrecht, The Netherlands; 2) Cardiovascular and Metabolic Sciences, Max-Delbrück-Center for Molecular Medicine (MDC) in the Helmholtz Association, Berlin, Germany; 3) Leiden University Medical Center, Leiden, The Netherlands; 4) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.

Chromothripsis causes complex genomic rearrangements that are likely to affect multiple genes and their regulatory contexts. The contribution of individual rearrangements and affected genes to the phenotypes of patients with complex germline genomic rearrangements is generally unknown. To dissect the impact of germline chromothripsis in a relevant developmental context, we performed trio-based RNA expression analysis on blood cells, induced pluripotent stem (iPS) cells and iPS-cell derived neuronal cells from a patient with *de novo* germline chromothripsis and both healthy parents. Sixty-seven genes are located within one megabase of the complex chromothripsis rearrangements involving seventeen breakpoints on four chromosomes. We find that three of these genes, *TWIST1*, *FOXP1* and *DPYD*, are both differentially expressed in the cells of the patient and associated with developmental disorders resembling the phenotype of the patient. Interestingly, the effect on *TWIST1* expression was exclusively detectable in the patient's iPS-derived neuronal cells, stressing the need for studying developmental disorders in the correct context. Analysis of the regulatory landscape around *TWIST1* by Hi-C indicates that deregulation of *TWIST1* expression is caused by gains and losses of patient- and cell type specific enhancer interactions near *TWIST1*. We demonstrate that a combination of patient-derived iPS cell differentiation and trio-based molecular profiling is a powerful approach to improve the interpretation of pathogenic complex genomic rearrangements. Here we have applied this approach to identify misexpression of *TWIST1, FOXP1* and *DPYD* as key contributors to the complex congenital phenotype resulting from germline chromothripsis rearrangements.

935T

***De novo* interstitial duplication 5q31.2q31.3 associated with microcephaly, growth restriction and global developmental delay.** *L. Schultz, E. McCready, C. Li.* McMaster Children's Hospital, Department of Genetics, 1200 Main St W, Hamilton, ON L8N3Z5.

We report a 19 months old female patient who presented with intrauterine growth restriction with a birth weight of 2.26kg at 37 weeks gestation and a postnatal history of failure to thrive. At 19 months, her weight was less than 5th percentile, height was at the 5th percentile and her head circumference was less than -3SD. She also had global developmental delay, unable to crawl or finger-feed herself and had no words. Physical examination also revealed subtle dysmorphic features that included a prominent forehead and nose with hypoplastic nares. A *de novo* duplication on 5q31.2q31.3 was found on microarray. The deletion is 2.98Mb in size and encompasses 50 RefSeq genes, including six that are catalogued in the Morbid OMIM: *KLHL3, MYOT, REEP2, SIL1, MATR3, TMEM173*. Interstitial duplication of 5q31.2q31.3 is rare. Larger duplications involving this region have been described [Giardino et al., 2004; Faguer et al., 2011] without a clear phenotypic pattern. We found only one interstitial duplication in 5q with similar breakpoints and size to our case [Rosenfeld et al., 2011]. Overlapping phenotype included microcephaly, developmental delay and a prominent forehead. However, congenital heart defect and ear anomaly are not present in our patient, and seizure/abnormal EEG has not been observed to date. By sharing our findings, we hope to contribute to the molecular and phenotypic delineation of a rare duplication 5q31.2q31.3.

936F

Lions and tigers and bears, oh my! Technically challenging variants are prevalent and require new approaches to NGS test development and validation. *S. Lincoln[1], R. Truty[1], J. Zook[2], B. Funke[3,4], C. Huang[5], J. Dickens[5], M. Ghent[1], S. Yang[1], A. Rosendorff[1], S. Aradhya[1], J. Park[6], M. Salit[2], R. Nussbaum[1].* 1) InVitae, San Francisco, CA; 2) National Institute for Standards and Technology, Gaithersburg, MD; 3) Laboratory for Molecular Medicine, Cambridge, MA; 4) Department of Pathology, Harvard Medical School, Boston, MA; 5) Seracare Life Sciences, Gaithersburg MD; 6) UT Southwestern Medical Center; Dallas, TX.

Most sequencing efforts currently focus on the detection of SNVs and small indels in accessible coding regions, although these methods can miss a significant fraction of variants with clinical importance. We examined clinical test results for over 30,000 patients across 1000 genes associated with cancer, cardiovascular, neurological and pediatric disorders. We found that technically challenging variants were a substantial fraction of the potentially actionable findings: Of roughly 5000 patients with a pathogenic variant, 2.9% had CNVs affecting only a single-exon,1.8% a large indel or complex sequence change, and 5.8% a variant in a region of high-GC, low complexity or poor short-read mapability. Even at >250x average depth, 3-4% of target bases in these genes had, by clinical testing standards, low (<20x) coverage in standard exome sequences, compared to 0.0% - 0.1% for coverage-optimized panels at similar depths. Also, 20% of target bases were outside the published Genome in a Bottle (GIAB) high-confidence regions, and very few examples of "hard" GIAB variants are present in these genes. Developing and validating tests that include these regions and variant types is challenging, in part because of the relative scarcity of positive and negative controls. Steps are being taken to improve this. Multiple assays including linked reads and long-read single molecule sequencing are being used to improve the GIAB reference, and the latest draft GIAB data increase the high-confidence regions from 78% to 89% of the genome. To increase the number of "hard" variants we developed sets of large plasmid inserts bearing specific, known variants engineered into the GIAB reference sequences. These were spiked into GIAB genomic DNAs at concentrations that mimic heterozygous variants and sequenced using standard NGS protocols. We compared the resulting BAM files against those from patients with the same variants and found them to be highly similar. Most currently published validation studies include few, if any, of the most challenging types of variants, which are indeed clinically important and are prevalent across a range of genetic disorders. We believe that clinical test validation standards should increasingly focus on this issue. To help labs do so, new GIAB data and GIAB samples with engineered plasmids will be available to the ASHG community by the 2016 Annual Meeting. Variant prevalence data will be available in ClinVar.

937W

Improving coverage of poorly sequenced regions in clinical exomes. *E. White, N. Chennagiri, E. Lopez, A. Frieden, J. Thompson.* Claritas Genomics, Cambridge, MA.

We have previously reported an orthogonal sequencing approach for clinical whole exome sequencing in which results of two next-generation sequencing platforms are combined for rapid variant confirmation. This both reduces the Sanger sequencing confirmation burden by ~90% and increases overall assay sensitivity since each platform uniquely sequences thousands of exons. In the current orthogonal approach, we sequence the Agilent Clinical Research Exome (CRE) libraries on the Illumina NextSeq and combine variants with those identified from AmpliSeq Exome libraries sequenced on the Ion Torrent Proton. Although the orthogonal platform increases exome variant sensitivity, there are still poorly covered regions that remain and may result in missed pathogenic variants. To minimize this problem, we have designed new sets of primers for low coverage AmpliSeq amplicons and amplified these independently at lower multiplicity than the highly multiplexed standard amplicons. Independent pools are designed for up to hundreds of genes in a phenotypically-driven manner. These are used as a supplement to the standard amplicons and sequenced together with them. We find that many of the low coverage regions are enhanced to the point that variants can be called and sensitivity in those regions is substantially improved so that patients receive a higher quality analysis.

938T

Rapid clinical WES in critically ill infants. *Y. Yang[1,2], M. Walkiewicz[1,2], F. Xia[1,2], R. Xiao[1,2], W. Bi[1,2], P. Liu[1,2], C. Qu[2], Y. Ding[3], J. Chandarana[2], P. Stankiewicz[1,2], A.M. Breman[1,2], J. Smith[1,2], S. Lalani[1,2], J. Smith[1,2], A. Patel[1,2], S.W. Cheung[1,2], N. Narayanan[2], A. Braxton[2], P. Ward[1,2], J.R. Lupski[1,3,5], A.L. Beaudet[1,2], S.E. Plon[1,4], R.A. Gibbs[1,3], D.M. Muzny[1,3], C.M. Eng[1,2].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) The Human Genome Sequencing Center, Baylor Col Med, Houston, TX; 4) Texas Children's Hospital, Houston, TX; 5) Department of Pediatrics, Baylor Col Med, Houston, TX.

Clinical whole exome sequencing (WES) has been integrated into healthcare in recent years. Initially, WES was used primarily on patients who had previously tested negative for specific genes, karyotype, and/or microarray studies, and the turnaround time (TAT) was usually several months. With the improvements of NGS methodologies and analysis tools, the TAT of WES can be reduced to about two weeks. The short TAT enables new clinical applications especially in the molecular diagnosis of prenatal patients or patients who are at high risk for potentially life-threatening conditions and are therefore considered critically ill. Here, we present the rapid clinical WES of 61 patients who were under 12 months of age. The patients were referred by multiple institutions; most of them had been admitted into intensive care units. Twenty-two out of the 61 cases had both WES and chromosomal microarray (CMA) tests ordered concurrently. The CMA/WES ordering status for the remaining cases is unknown. Overall, approximately 56% (34/61) of the patients received a molecular diagnosis from rapid WES. Twelve of the 22 (55%) patients with concurrent CMA/WES received a molecular diagnosis including 1 diagnosis of a large heterozygous deletion identified by both tests. No additional diagnoses were made by CMA. The molecular diagnosis rate from our data is consistent with the approximately 50% reported from other studies on critically ill young patients. The collective data to date suggest that rapid exome sequencing should become a standard of care and the first-tier NGS test for NICU or PICU patients whose phenotypic features suggest an underlying genetic etiology. However, accessibility to clinical exome sequencing has remained challenging due to low reimbursement and authorization rates from insurance plans. It is therefore critical to further evaluate the clinical utility of rapid WES and establish practice guidelines leading to efficient and cost effective molecular diagnoses for NICU and PICU patients. .

939F

A novel mutation of *ROBO3* in horizontal gaze palsy with progressive scoliosis. *S. Tug Bozdogan[1], E. Dinc[2], A. Sari[2], A. Ozgur[3], A. Bisgin[1].* 1) Medical Genetics, Cukurova University Medical School, Adana, Turkey; 2) Mersin University Medical School, Ophthalmology; 3) Mersin University Medical School, Radiology.

INTRODUCTION Horizontal gaze palsy with progressive scoliosis (HGPPS), is a rare autosomal recessive disorder with oculomotor and general disturbances in innervation. The *ROBO3* gene mutation was described as a responsible gene of HGPPS. Here we present a case of HGPPS with a novel mutation of *ROBO3.* **CASE REPORT** A 22-year-old female patient was admitted to clinic with dysmorphic features and strabismus in both eyes. She was the fifth child of 8 siblings of first cousins. Her two elder sisters have a history of basic esotropia surgery. Her motor development was delayed, thoracolumbar scoliosis was diagnosed at the age of 2 years and had been progressive. Her physical examination revealed short stature, short neck, pectus carinatum with left thoracic scoliosis extending from T6 to L1 with a quite prominent hump to the right. MR imaging of the brain depicted a hypoplastic pons with a cleft, a hypoplastic medulla oblongata with butterfly morphology and the fourth ventricle with a tent-shaped roof. There was severely limited abduction and adduction in both eyes; vertical movements were unimpaired. Because the patient presented with clinical features suggestive of HGPPS, the coding regions of *ROBO3* was subjected to sequence analysis. She carried a homozygous p.R842* (c.2524C>T) mutation in exon 16 of *ROBO3.* However the mutation has not been reported previously, in-silico analyses revealed as a disease making mutation. Only the mother accepted to involve in the genetic testing, that resulted as a carrier of this novel mutation; heterozygous p.R842* (c.2524C>T). **DISCUSSION** HGPPS is a rare congenital disorder characterized by defective horizontal eye movements and progressive scoliosis during childhood and adolescence. Brain stem malformation and defective crossing of brain stem neuronal pathways are also seen in HGPPS. Our patient presented with severe left thoracic scoliosis extending from T6 to L1 vertebrae, exotropia and defective horizontal eye movements with normal vertical movements. *ROBO3* gene play a role in the regulation of hindbrain axonal midline crossing and helps cell migration. To date, 33 different mutations of *ROBO3* have been described and related to HGPPS. In our case, we determined a novel homozygous mutation p.R842* (c.2524C>T) causing a premature stop codon which is a disease making mutation. Much remains to be learned about this condition despite the advances in our understanding.

940W

Analysis of chromosome 22q11.2 copy number variations by multiplex ligation-dependent probe amplification. *T. Alarcon Martinez[1], S. Kabacam[2], A.C. Ceylan[3], O.P. Simsek Kiper[2], M. Alikasifoglu[3], K. Boduroglu[2], G.E. Utine[2].* 1) Department of Pediatrics, Hacettepe University, Ankara, Turkey; 2) Division of Pediatric Genetics, Hacettepe University, Ankara, Turkey; 3) Department of Medical Genetics, Hacettepe University, Ankara, Turkey.

Introduction: Genomic copy number variations (CNVs) at 22q11.2 are associated with several syndromes, characterized by different combinations of clinical manifestations. The traditional molecular diagnostic procedure for CNVs of the region is chromosomal analysis coupled with fluorescence *in situ* hybridization (FISH). However, atypical deletions and duplications may be missed by these methods. Multiplex ligation-dependent probe amplification (MLPA) was herein investigated as an alternative to FISH for molecular diagnosis and more precise delineation of CNVs at 22q11.2. Methods: MLPA analysis was performed in a total of 83 patients, in search for higher resolution molecular results and also for possible genotype-phenotype correlations. Of these, 30 patients had FISH-confirmed 22q11.2 deletion syndrome (22q11.2DS), and 53 patients had clinical diagnoses previously not supported by FISH. Results: MLPA confirmed the diagnosis of 22q11.2DS in 30 patients with positive FISH. Among 30 patients, 27 (90%) showed typical 3-Mb deletion, while 2 (6.6%) displayed proximal 1.5-Mb deletion. Patients with 1.5-Mb deletion were monozygotic twins, demonstrating only velopharyngeal insufficiency. One patient (3.4%) had a deletion encompassing cat-eye region and LCRA, while her sister exhibited duplication in the same region. Parental karyotyping results showed 46, XX,t(13;22)(q12.1;q12.1), balanced translocation in their mother. MLPA analysis of 53 patients with negative FISH results revealed mosaic 3-Mb deletion in one, *TBX1* duplication in one and *TBX1* deletion in three patients. *TBX1* deletions are confirmed by real time PCR, whereas *TBX1* duplication is also detected by comparative genomic hybridization. Conclusions: MLPA is more sensitive than FISH in molecular diagnosis and delineation of CNVs at 22q11.2, with a yield in the current study of 9.4% (5/53) in patients with previously negative results. Moreover, patients with duplications and atypical deletions showed different clinical characteristics than 22q11.2DS.

941T

Nonrecurrent *PMP22-RAI1* contiguous gene deletions arise from replication-based mechanisms and result in Smith-Magenis syndrome with evident peripheral neuropathy. *J. Neira[1], B. Yuan[1], S. Gu[1], T. Harel[1], P. Liu[1], I. Briceño[2,3,4], S. Elsea[1], A. Gómez Gutiérrez[2,3], L. Potocki[1,7], J.R. Lupski[1,5,6,7].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 2) Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia; 3) Instituto de Referencia Andino, Bogotá, Colombia; 4) Facultad de Medicina, Universidad de La Sabana, Chía, Colombia; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA; 6) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 7) Texas Children's Hospital, Houston, Texas, 77030, USA.

Hereditary neuropathy with liability to pressure palsies (HNPP) and Smith-Magenis syndrome (SMS) are genomic disorders associated with deletion copy number variants (CNVs) involving chromosome 17p12 and 17p11.2, respectively. Nonallelic homologous recombination (NAHR)-mediated recurrent deletions are responsible for the majority of HNPP and SMS cases; the rearrangement products encompass the key dosage sensitive genes *PMP22* and *RAI1*, respectively, and result in haploinsufficiency for these genes. Less frequently, nonrecurrent genomic rearrangements occur at this locus. Contiguous gene duplications encompassing both *PMP22* and *RAI1*, i.e. *PMP22-RAI1* duplications, have been investigated, and replication-based mechanisms rather than NAHR have been proposed for these rearrangements. In the current study, we report molecular and clinical characterizations of six subjects with the reciprocal phenomenon of deletions spanning both genes, i.e. *PMP22-RAI1* deletions. Molecular studies utilizing high-resolution array comparative genomic hybridization and breakpoint junction sequencing identified mutational signatures that were suggestive of replication-based mechanisms. Systematic clinical studies revealed features consistent with SMS, including features of intellectual disability, speech and gross motor delays, behavioral problems and ocular abnormalities. Five out of six subjects presented clinical signs and/or objective electrophysiologic studies of peripheral neuropathy. Clinical profiling may improve the clinical management of this unique group of subjects, as the peripheral neuropathy can be more severe or of earlier onset as compared to SMS patients having the common recurrent deletion. Moreover, the current study, in combination with the previous report of *PMP22-RAI1* duplications, contributes to the understanding of rare complex phenotypes involving multiple dosage-sensitive genes from a genetic mechanistic standpoint. .

942F

Comprehensive assay to detect DNA sequence variants in the muco-polysaccharidoses related genes using next generation sequence. *A. Brusius-Facchin[1,2], M. Siebert[3], D. Leão[3], D. Malaga[1,2], G. Pasqualim[2,4], U. Matte[2,3,4,5], R. Giugliani[1,2,5], S. Leistner-Segal[1,6].* 1) Hospital de Clinicas de Porto Aelgre, Porto Alegre, Brazil; 2) Post-Graduation Program in Genetics and Molecular Biology (PPGBM) – Universidade Federal do Rio Grande do Sul (UFRGS), Brazil;; 3) Unidade de Análises Moleculares e de Proteínas (UAMP) – Centro de Pesquisa Experimental, HCPA, Brazil; 4) Gene Therapy Center, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 5) Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 6) 6.Post-Graduation Program in Medicine: Medical Sciences (PPGCM) – Universidade Federal do Rio Grande do Sul (UFRGS), Brazil.

The mucopolysaccharidoses (MPS) a group of rare genetic disorders caused by mutations in any of the eleven genes that encode enzymes responsible for the stepwise degradation of glycosaminoglycans (GAGs). The abnormal accumulation of these substrate inside the lysosomes of most cells, induce progressive cellular damage and multiple organ failure, with consequent reduction of the quality of life and life expectancy. Over the past 30 years Sanger sequencing technology was the gold standard approach to identify sequence alterations in the target region analyzing individual genes, usually exon-by-exon. Currently, the Next Generation Sequence, a fast tool with high capacity of sequencing, allows the mutation detection of multiple genes and patients at same time improving the molecular diagnosis. In order to evaluate the NGS technology for MPS diagnosis, eleven genes related to MPS diseases were divided into three different custom panels according to the overall clinical presentations. DNA samples from 70 patients with previous biochemical diagnosis of MPS and already genotyped with Sanger sequencing, were analyzed by Ion Torrent semiconductor technology. Using this new approach, we could identify different types of pathogenic mutations distributed across all exons of MPS related genes. We were able to identify 96% of all gene variants previously identified by Sanger, showing high sensitivity in detecting several types of mutations. As no new variations were detected, the NGS method presented 100% specificity. This comprehensive assay for the identification of the disease causing mutations in the MPS is an attractive option for the genotyping of MPS patients. Indeed, with the incorporation of NGS to clinical laboratories, we could consider that the present diagnostic flow could even be inverted, with biochemical tests becoming confirmatory methods to prove the functional consequences of a genetic abnormality detected by NGS.

943W

Combined genetic analyses with targeted next-generation sequencing, multiplex ligation-dependent probe amplification and microarray comparative genomic hybridization for Alagille syndrome and biliary atresia. *K. Ohashi[1], T. Togawa[1], T. Sugiura[1], K. Ito[1], T. Endo[1], K. Aoyama[1], Y. Negishi[1], T. Kudo[2], R. Ito[2], S. Saitoh[1].* 1) Department of Neonatology and Pediatrics, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Hepatology, National Medical Center for Children and Mothers, National Center for Child Health and Development, Tokyo, Japan.

Introduction Alagille syndrome (ALGS) is an autosomal dominant multisystem disorder, and is caused by mutations in *JAG1* or rarely in *NOTCH2*. Biliary atresia (BA) is the representative differential diagnosis of ALGS. We aimed to evaluate combined genetic analyses with targeted next-generation sequencing (NGS), multiplex ligation probe amplification (MLPA) and microarray comparative genomic hybridization (CGH) in subjects with ALGS, atypical ALGS and BA. **Method** We recruited subjects with ALGS, atypical ALGS, and BA. All subjects were tested by targeted NGS analysis including *JAG1* and *NOTCH2* using the AmpliSeq and Ion Torrent Personal Genome Machine (Ion PGM, ThermoFisher Scientific) system. If no mutations were detected in *JAG1* or *NOTCH2*, we performed MLPA analysis covering entire exons of *JAG1*. MLPA was also applied to confirm the possible CNVs called by NGS. We further performed microarray CGH analysis to subjects with whole exon deletion detected by the MLPA analysis. **Result** We analyzed 56 subjects (28 with ALGS, 6 with atypical ALGS, and 22 with BA), and detected mutations in *JAG1* or *NOTCH2* in 23/28 (82.1%) subjects with ALGS, in 4/6 (66.7%) subjects with atypical ALGS, and in 1/22 (4.5%) subjects with BA. The mutation type was as follow; 3 missense mutation, 5 nonsense mutation, 2 splice site mutation, 8 small insertion or deletion, 1 single exon deletion, 3 whole exons deletions in *JAG1* and 1 missense mutation in *NOTCH2* in subjects with ALGS, 1 missense mutation, 1 nonsense mutation, 1 small deletion, 1 single exon deletion in *JAG1* in subjects with atypical ALGS, and 1 missense mutation in *JAG1* in a subject with BA. Distribution of *JAG1* mutations was missense mutation 18.5%, nonsense mutation 22.2%, splice site mutation 7.4%, small indels 33.3%, single exon deletion 7.4%, and whole exons deletion 11.1%. The targeted NGS analysis detected 26/28 (92.9%) of mutations, and MLPA analysis of *JAG1* detected 2/28 (7.1%) of mutations. **Conclusion** Combined molecular genetic analyses, using targeted NGS, MLPA of *JAG1*, and microarray-CGH can achieve efficient diagnostic yields for the subjects with typical and atypical ALGS as well as subjects with BA phenotypically overlapping with ALGS.

944T

Supplemental CNV analysis in diagnostic NGS gene panel data. *J.J. Saris, M.A. van Slegtenhorst, R. Minkelen, M. Wilke, A.M.W. van den Ouweland, H.T. Brüggenwirth, L.H. Hoefsloot.* Clinical Genetics, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands.

Standard diagnostic laboratory flows based on PCR and Sanger sequencing often include MLPA analysis when available and relevant. Introducing NGS gene panels in a clinical setting promises a higher diagnostic yield due to scanning more genes, while reducing costs, but is usually designed only for sequence variant detection. Normalized Depth-of-Coverage (DOC) calculation can indicate exon copy number variation (CNV) and might replace MLPA. Our lab has established a NGS sequencing flow using JSI SeqNext software (www.jsi-medisys.de). Recently we validated the effectiveness and usability of a DOC-CNV submodule using data from genes located on chromosome X and positive controls. Type and number of reference files (1-on-1 and n=12), signal-to-noise ratio and detection cut-off were evaluated in SeqNext and independent via Z-score analysis. Aim: Retrospective analysis of 27 gene panel experiments (324 samples), comparison to validation results and prospective analysis of future cardiomyopathy experiments (±12) for exon-CNV and wet lab confirmation of calls ≥ 2 exons and <5% frequency, using an internal database, including a filter of "noisy" ROI. Results: In 324 samples (incl. 116 cardiomyopathy) DOC based gender determination using chromosome X located genes, calling was concordant with the recorded clinical gender. Seven multi-exon (a.o. NEXN-e2e11 & B9D2-e2e4) and 10 whole gene CNV (a.o. NPHP1, ACTB) were detected, of which 8 confirmed (Known True Positives), 7 new variants of which confirmation is ongoing, and 2 polymorphic CNV based of occurrence frequency. Manual curation of single exon CNV (124 events) indicates deep intronic sequence variation near certain exons influencing accuracy of DOC results on capturing and/or read mapping level, e.g. 5' ANKRD1-exon4 (MAF ±30%) with Mendelian appearance. Conclusion: Exon CNV detection using DOC-analysis in SeqNext promises to be an possible replacement of MLPA in diagnostics. Validation, update on results and lessons learned will be presented.

945F

Acute Myeloid Leukemia-associated *DNMT3A* **p.Arg882His mutation in a patient with Tatton-Brown-Rahman Overgrowth Syndrome as a constitutional mutation.** *R. Kosaki[1], H. Terashima[2], M. Kubota[2], K. Kosaki[3].* 1) Division of Medical Genetics, National Center for Child Health and Development, Tokyo, Japan; 2) Division of Neurology,National Center for Child Health and Development, Tokyo, Japan; 3) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

DNA methylation plays a critical role in both embryonic development and tumorigenesis and is mediated through various DNA methyltransferases. Constitutional mutations in the *de novo* DNA methyltransferase *DNMT3A* cause a recently identified Tatton-Brown-Rahman overgrowth syndrome (TBRS). Somatically acquired mutations in *DNMT3A* are causally associated with acute myeloid leukemia (AML), and p.Arg882His represents the most prevalent hotspot. So far, no patients with TBRS have been reported to have subsequently developed AML. Here, we report a live birth and the survival of a female infant with the TBRS phenotype who had a heterozygous constitutional *DNMT3A* mutation at the AML somatic mutation hotspot p.Arg882His in her DNA from peripheral blood and buccal tissue. Her characteristic features at birth included hypotonia, round face, narrow palpebral fissures, ventricular septal defect, umbilical hernia, sacral cyst, Chiari type I anomaly, and intellectual disability. At the age of 6 years, she exhibited overgrowth(+3SD). The documentation of the same variant (*DNMT3A* p.Arg882His) as both the constitutional mutation associated with TBRS and the somatic mutation hotspot of AML supports a notion that mutations responsible for TBRS and those for AML share the same mode of action. Surveillance for hematologic abnormalities may be considered in patients with TBRS. .

946W

A new syndrome associated with compound heterozygous changes in
NCOR2 **gene.** *C. Li, M. Szybowska, G. Cowing.* Dept of Genetics, McMaster Children's Hospital, Hamilton, ON, Canada.

We report the clinical phenotype of a case with multiple congenital anomalies and profound intellectual disability in association with compound heterozygous changes in *NCOR2*, c.3322C>T (p.L1108F) and c.4621C>G (p.L1541V). The patient presented antenatally with increased nuchal, two-vessel cord, polyhydramnios and hydronephrosis. Postnatally she was noted to have glaucoma, cloudy cornea, kyphoscoliosis, hip dysplasia, pulmonary lymphangiectasia, interstitial glycogenosis and multiple dysmorphic features. These include brachycephaly, well-defined arched brows, shallow orbits with wide palpebral fissures and long lashes. She has a bulbous nose with hypoplastic nares. The mouth is small with thin lips and the tongue is protruding which in infancy led to obstruction of the airway and difficulty with intubation. She has small and tightly spaced teeth. The facial profile is flat. The ears are crumpled and flat against the skull. She has thin extremities with narrow hands and feet and slender fingers and toes. Reduced range of motion is noted of the hips, knees and elbows. Camptodactyly of the fingers and overlapping toes with hypoplastic nails are also noted. There is generalized hypotrophy of the muscles. Brain MRI showed thinning of the corpus callosum, periventricular white matter volume loss and exvacuo dilation of the ventricles. Spine MRI showed tethered cord and query segmentation defects. At her last assessment at 10 years of age, she was not able to sit unsupported and had only few signs for communication. Her parents each carry one of the genetic changes in the NCOR2 gene. The family history is non-contributory otherwise. The *NCOR2* gene encodes SMRT (silencing mediator for retinoid and thyroid hormone receptors). NCOR-family plays central roles in the timing of embryo and stem cell differentiation [Short et al., 2015]. SMRT is a transcription co-suppressor that, in concert with MeCP2 protein (involved in Rett syndrome), allows non-liganded nuclear receptors to remain inactive [Pei et al., 2011]. Studies in mice showed *Ncor2* to be essential for normal neural and cardiac development. Aberrant expression of SMRT in Xenopus also resulted in abnormal head development [Malatre et al., 2006]. However, pathogenic mutations have not been reported in the *NCOR2* gene in association with a constitutional disease condition in humans. We hope to share our findings and draw attention to the potential role of *NCOR2* in human disease conditions.

947T

The NIH Undiagnosed Diseases Program: A mystery solved at long last.
A.P. Liebendorfer[1], G. Nesterova[1], C. Wahl[1], B.P. Brooks[2], W.M. `[2], C.C. Chan[2], J. Balow[3], D. Adams[1], W. Gahl[1], T. Markello[1]. 1) National Human Genome Research Institute, Bethesda, MD; 2) National Eye Institute, Bethesda, MD; 3) National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

Introduction: The NIH Undiagnosed Diseases Program (UDP) enrolls participants who have illnesses that remain undiagnosed despite extensive medical workup. In 2008, a 32 y/o woman presented to the UDP with h/o seizures, weakness in her lower extremities with abnormal gait, exudative retinopathy leading to right eye enucleation, multiple intestinal bleedings without ulceration and eventually chronic renal failure, starting at age 8-10. MRIs revealed multiple cavernous hemangiomas. The patient started peritoneal dialysis at age 27 due to ESRD after renal injury with preeclampsia. The patient died from intractable intestinal bleeding, acute renal failure and pneumonia seven months after evaluation. Overall the presentation was multisystemic involving GI, skin, bone, eye and CNS. **Methods:** Evaluation included neurological, dermatological, ophthalmological, hematological, and radiological exams and autopsy. Discussions focused on evidence of diffuse basement membrane defects. Her exome was analyzed in 2016 because population and disease variant databases had matured; the biologic mother also became available for study. **Results:** Postmortem, the patient exhibited anterior scalp poliosis, depigmented macules on her arms and chest, patches of hyperpigmentation on the abdomen and left thigh, and hard nodules bilaterally affecting the distal two thirds of her legs. Retinal telangiectasia was found. Calcification was noted intracranially and in her lower extremities, and osteopenia was noted in much of her lower body. SNP chip analysis suggested a de novo or compound heterozygous model for disease. Exome analysis revealed two phased variants in *CTC1*: c.2959C>T, p.R9087W and c.1603T>C, p.C535R. *CTC1* is associated with Coats Plus Syndrome (OMIM 612199). **Conclusions:** The case illustrates the rapid changes in genetic diagnostics and discovery using genome-scale sequencing as a diagnostic tool. When our patient was first seen, no exome sequencing was available and connection between *CTC1* and Coats disease was unknown. The case also provides an extensive description of advanced Coats Plus Syndrome and further characterizes the phenotype of this rare disorder. This research is supported by the Intramural Research Program of the National Human Genome Research Institute and the Common Fund of the National Institutes of Health.

948F

Detection of copy number variants by whole exome sequencing. *J.M. Meck, L.B. Henderson, G. Douglas, O. Cano, G. Li, J. Scuffins, G. Richard, K. Retterer, J. Juusola.* GeneDx, Gaithersburg, MD.

Whole exome sequencing (WES) is typically utilized to identify sequence variants in the human exome, while genomic copy number variants (CNVs) are detected by chromosomal microarray (CMA). Our clinical laboratory developed an algorithm for identifying CNVs including multi-genic, whole, or partial gene deletions (dels)/duplications (dups) directly from the WES next generation sequencing (NGS) data [1]. Overall, the positive rate of WES in our cohort of over 11,000 affected individuals was 26.4%. Almost 5% of the positive findings were pathogenic CNVs confirmed by array, MLPA, qPCR or PCR. We reviewed 146 pathogenic CNVs identified from WES NGS data in 143 probands (including 3 with presumed unbalanced translocations) to determine their type, size, inheritance, and associated disease spectrum. Reported CNVs included 127 (87%) dels and 19 (13%) dups ranging from a single exon (10 cases) to 16.3 Mb. Seventy-seven (54%) probands had whole or partial del/dup of a single gene, while 66 (46%) had multi-genic aberrations. Most (35/66; 53%) multi-genic CNVs were associated with recurrent microdel/dup syndromes, such as 16p11.2 del or dup (n=5), classic and distal 22q11.2 dels (n=4), 15q13.3 del (n=2), 17p12 dup (n=2) and monosomy 1p36 (n=2). Although most reported CNVs act in a dominant fashion, 1/3 were either homozygous (22; 15%) or heterozygous CNVs (26; 18%) in *trans* with a pathogenic sequence variant in a gene associated with an autosomal recessive (AR) disorder. Five AR disorders with at least one pathogenic CNV were detected recurrently: Cohen syndrome (*VPS13B*, n=4); juvenile neuronal ceroid lipofuscinosis (*CLN3*, n=3); *TBCK*-related disorder (*TBCK*, n=2); 3-methylglutaconic aciduria with deafness, encephalopathy, and Leigh-like syndrome (*SERAC1*, n=2); and myopathy with extrapyramidal signs (*MICU1*, n=2). Analysis of inheritance pattern showed that 59 (41%) CNVs occurred *de novo*, 54 (38%) were inherited, and 30 (21%) had no or uninformative parental testing. Prior or concurrent CMA was performed in at least 66/143 (46%) cases, with results known to us in 59. In 42/59 (71%) cases the pathogenic CNV was not detected by CMA; most of these (38/42; 90%) had intragenic CNVs. In summary, our data demonstrate the significant contribution of pathogenic CNVs to the mutational spectrum identifiable by WES, and highlight the utility of WES NGS for identifying single gene and exon-level CNVs not detected by routine CMA testing. [1] Retterer et al. PMID 25356966.

949W

Multiple rare actionable mutations identified by whole exome or genome sequencing of patients with craniosynostosis. *K.A. Miller[1], S.R.F. Twigg[1], S.J. McGowan[2], J.M. Phipps[1], A.L. Fenwick[1], D. Johnson[3], P. Noons[4], K.E.M Rees[5], E.A. Tidey[5], J. Craft[6], J. Taylor[7], J.C. Taylor[8], J.A.C. Goos[9], S.M.A Swagemakers[9], I.M.J Mathijssen[8], H. Lord[7], T. Lester[7], D. Cilliers[10], J.A Hurst[5], J. Morton[11], E. Sweeney[12], L.C. Wilson[5], A.O.M. Wilkie[1,3].* 1) Clinical Genetics Group, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 2) Computational Biology Research Group, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 3) Craniofacial Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; 4) Department of Craniofacial Surgery, Birmingham Children's Hospital NHS Foundation Trust, United Kingdom; 5) North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; 6) Oxford Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 7) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; 8) Erasmus Medical Centre, University Medical Centre Rotterdam, Department of Plastic and Reconstructive Surgery and Hand Surgery, Rotterdam, the Netherlands; 9) Erasmus Medical Centre, University Medical Centre Rotterdam, Department of Bioinformatics, Rotterdam, the Netherlands; 10) Department of Clinical Genetics, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; 11) Birmingham Women's Hospital NHS Foundation Trust, Birmingham, United Kingdom; 12) Department of Clinical Genetics, Liverpool Women's NHS Foundation Trust, Liverpool, United Kingdom.

Accurate molecular diagnosis is critical for the clinical management, counselling and prognosis of individuals with suspected monogenic diseases, particularly where early diagnosis and intervention would drastically influence decision making by the clinician or family. Rare causes of disease are often overlooked following traditional genetic testing routes, as many conditions have a highly variable clinical presentation, present as part of a phenotypic spectrum or are difficult to distinguish from other, very similar, disorders. Use of whole exome or genome sequencing (WES/WGS) technologies circumvents many of these problems, since these are agnostic to the underlying genetic cause. Here we evaluated the use of WES (n=37) or WGS (n=3) in a consecutive series of 40 patients with craniosynostosis (premature fusion of the cranial sutures), who had been identified as priority cases by either clinical geneticists with particular expertise in craniosynostosis (n= 22) or by specialist clinical laboratory scientists (n= 18), and in whom standard diagnostic testing had failed to reveal an underlying cause. To date we have identified a likely causative mutation in 15 families (37.5%). All associated genes were mutated in single families, except for *IL11RA* (2 families), a recently recognised recessive cause of craniosynostosis that has now been added to our diagnostic panel. We classified the other positive diagnoses as follows: known craniosynostosis gene with atypical presentation (*EFNB1*, *TWIST1*); genes for which mutations are only rarely associated with craniosynostosis (*FBN1*, *KRAS*, *MSX2*, *STAT3*); genes that represent very recently recognised associations with craniosynostosis (*CDC45*, *HUWE1*, *ZIC1*); and known disease genes for which the causal relationship with craniosynostosis is currently uncertain (*AHDC1*, *NTRK2*). In two further families, likely novel disease genes are currently undergoing functional validation. In 5 of the 15 positive cases, the (previously unanticipated) molecular diagnosis had immediate, actionable consequences for either genetic or medical management of the families (mutations in *EFNB1*, *FBN1*, *KRAS*, *STAT3*, *NTRK2*), which we will discuss. In summary our findings illustrate the molecular diversity of causes of craniosynostosis and the considerable added value provided by WES/WGS for the precise diagnosis of patients suspected to have a genetic cause, where the underlying mutation would have otherwise eluded routine clinical testing.

950T

Secondary actionable findings identified by whole-exome sequencing: Frequency and potential consecutive medical recommendations and organisation: Study on 653 consecutive tests. *C. Thauvin-Robinet[1], J. Thevenon[1], S. Nambot[1], P. Kuentz[1], A.-L. Bruel[1], A. Chassagne[1], E. Cretin[1], O. Putois[1], A. Pelissier[1], C. Peyron[1], E. Gautier[1], J. Skrzypski[2], D. Lehalle[1], N. Jean-Marçais[1], P. Callier[1], A.-L. Mosca-Boidron[1], C. Poé[1], T. Jouan[1], M. Chevarin[1], M. Lefebvre[1], E. Tisserant[1], C. Binquet[3], J. Deleuze[4], Y. Duffourd[1], L. Faivre[1].* 1) Clinical genetics centre, FHU-TRANSLAD, Dijon, France; 2) Centre Georges François Leclerc Dijon, France; 3) CIC-EC, Dijon, France; 4) Centre national de génotypage, Évry, France.

In whole exome sequencing (WES) integrated into the practice of medicine, there is potential for the recognition and reporting of incidental or secondary findings unrelated to the indication for ordering the sequencing but of medical value for patient care. In 2013, the American College of Medical Genetics and Genomics (ACMG) examined the issue of incidental findings in WES/WGS, and introduced recommendations to search for, evaluate and report medically actionable variants in a minimal set of 56 genes. In order to evaluate the implications in organization of care, we have conducted a retrospective study evaluating the frequency of variants in a list of actionable genes in a cohort of 653 consecutive probands with multiple congenital anomalies explored by WES. The 174 genes of interest comprised the list of the ACMG, extended to further cancer predisposition genes (n=154), and 3 genes of the most frequent autosomal recessive diseases in France leading to the screening of heterozygote in relatives (*CFTR*, *HFE*-p.C282Y, *CYP21A2*-classical form). Variants in 110 genes implicated in X-linked intellectual deficiency (XLID) were also searched to detect female carriers for genetic counselling issues. We looked for all the variants with known Clinvar status and considered only pathogenic and likely pathogenic variants following the interpretation recommendations of the ACMG and Association for Molecular Pathology. We identified 19 cases (2,9%) with autosomal dominant heterozygous pathogenic variant (0,9%, in a gene responsible for a cancer predisposition, 1,4% in a hereditary cardiac disease, 0.6% in polycystic kidney disease) (11/19 variants in the ACMG actionable 56 genes), 1 DMD *carrier*, no XLID carrier, 21 *CFTR* heterozygous cases (3,2%), 55 HFE-pC282Y heterozygous cases (8,4%) and 1 *CYP21A2* heterozygous case (0,1%). *SMN1* deletion and the majority of *CYP21A2*-classical form variants could not be detected. If returned to patients (14,9%), these results should induce 653 pre-test and 97 post-test genetic consultations, as well as 11 cardiologic, 7 nephrology, 5 gynecologic, 4 dermatologic, 2 gastroenterological, and endocrine evaluations. Additional family evaluations should also be anticipated. The study is in progress to also include the additional impact of 17 genes with actionable pharmacogenetic alleles and genes leading to coagulation defects.

951F

Detection of low level mosaic mutations in a gene panel for segmental overgrowth. *M. Alders, M.A. Haagmans, O.R. Mook, M.M.A.M. Mannens.* Department of Clinical Genetics, AMC, Amsterdam, the Netherlands.

Somatic mutations in genes of the PI3K/AKT/mTOR pathway causing segmental overgrowth may be present in a very low percentage. This can hinder the identification of causal mutations. Sanger sequencing, but also standard NGS diagnostic pipelines are able to identify mosaic mutations with a detection limit of approximately 10%. Targeted analysis of specific sites is usually used for more sensitive detection of mutations. With the growing number of known mutations there is a need for a high sensitive mutation scanning method to analyze the complete coding sequences of the genes involved rather than analysis of hotspots. We have developed a NGS based gene panel test for segmental overgrowth (AKT1, AKT3, MTOR, PIK3CA, PIK3R2, PTEN, TSC1 and TSC2) especially for the detection of low level mosaicism. Samples are sequenced with high coverage (average 1500-2000 reads, minimum 500 reads) and an analysis pipeline was developed that is able to detect mosaic mutations as low as 1%.

952W

The CAUSES Research Clinic: A pediatric sequencing initiative in British Columbia. *C. du Souich[1,2,3], A.M. Elliott[1,2,3], C. van Karnebeek[2,4], S. Adam[3], J. Gillis[4], J.C. Mwenifumbo[2], T.N. Nelson[2,5], A. Lehman[1,2,3], J.M. Friedman[1,2,3].* 1) Provincial Medical Genetics Program, BC Women's Hospital and Health Centre, Vancouver, BC, Canada; 2) Child and Family Research Institute, Vancouver, BC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Division of Biochemical Diseases, Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada; 5) Department of Pathology and Laboratory Medicine, BC Children's Hospital, Vancouver, BC, Canada.

   Background: The CAUSES study, a 3 year translational research initiative to sequence 500 BC pediatric patients with suspected genetic disorders, began in June 2015. Inclusion criteria and a clinical genomic consultation service were developed to select patients suitable for enrollment. Patients are referred from various pediatric subspecialties throughout the province. Genetic counselling is a critical component throughout the study. Telehealth is available for families from remote areas and interpreters for non-English speaking families. Genetic counseling and health economic research are embedded in the program. *Methods:* A comprehensive list of issues and implications of sequencing is discussed and informed consent is obtained by a genetic counselor prior to testing. Exome sequencing (ES; Agilent All Exon V5+UTR) is performed on the proband and both parents using an Illumina platform. A customized bioinformatics pipeline is used to align sequencing reads to the reference human genome (Bowtie 2) and call good quality SNVs & indels (samtools/bcftools). Ensembl transcript set annotation identifies functional variants, which are then filtered against public & in-house databases. A list of rare variants compatible with Mendelian inheritance patterns is annotated with diagnostically relevant information, and delineated by genes in the OMIM compendium as disease-causing (clinome) or not (non-clinome). The annotated variant list produced is reviewed by a panel that includes clinical geneticists, genetic counselors, bioinformaticians, clinical molecular geneticists and the referring physician. Putative causal variants are confirmed by the clinical laboratory, including Sanger sequencing and clinical variant interpretation, prior to return of results. Genetic counseling is provided to families with result interpretation. *Findings:* To date, 82 families have undergone ES and bioinformatic analysis. Thirty-four patients received a probable or definite diagnosis (40%). The majority were *de novo* autosomal dominant disorders. Three patients had variants in *EP300*, 3 in *ARID1B* and 2 in *DDX3X*. Treatment/management has been impacted in several patients. *Conclusions:* The CAUSES clinic demonstrates the clinical utility of ES and the importance of utilizing trios in patients with undiagnosed genetic disorders. It also shows the importance of a comprehensive approach to genome sequencing, which includes a strong genetic counseling component.

953T

Clinical features of 28 patients with presumptive diagnosis of *PIK3CA*-**related overgrowth spectrum (PROS) in a tertiary hospital in Brazil.** *C.H. Chung, R.S. Honjo, D.R. Bertola, C.A. Kim.* Unidade de Genética, Instituto da Criança - Hospital das Clínicas da FMUSP, São Paulo, Brazil.

   PIK3CA-related overgrowth spectrum (PROS) is a group of segmental overgrowth disorders usually caused by somatic mosaic mutation in *PIK3CA*. It has been associated to Megalencephaly-capillary malformation (MCAP) syndrome, Hemimegalencephaly, CLOVE syndrome (Congenital Lipomatous Overgrowth, Vascular Malformations, and Epidermal Nevi), fibroadipose hyperplasia, Klippel-Trenaunay syndrome and isolated macrodactyly. We report the clinical findings of 28 patients with presumptive diagnosis of PROS, from the Genetics Unit of Instituto da Criança - HCFMUSP (São Paulo, Brazil). 15/28 were initially diagnosed as Klippel-Trenaunay syndrome; 3/28 as Klippel-Trenaunay-Weber syndrome; 3/28 as CLOVE syndrome; 2/28 as Isolated macrodactyly; 2/28 as Proteus syndrome; 3/28 were suspected of Proteus syndrome or Klippel-Trenaunay syndrome. Their clinical manifestations were: hemangioma (82.1%), body asymmetry (67.8%), vascular malformations (46.4%), macrodactyly (39.3%), learning disability (21.4%), seizures (17.8%), renal abnormality (17.8%), high stature (14.3%), macrocephaly (14.3%), lipomas (14.3%), skeletal involvement (14.3%), eye involvement (14.3%), CNS malformation (10.7%), thromboembolic events (7.1%), short stature (3.6%) and polydactyly (3.6%). Prenatal ultrasound detected alterations in 17.8% of the cases. Even though a few patients were initially evaluated for clinical suspicion of Proteus syndrome, none fulfilled the diagnostic criteria for this syndrome. Due to the overlapping phenotypes in segmental overgrowth disorders, the molecular testing in the affected tissues is crucial to establish the correct diagnosis. It may also contribute to a better understanding of the clinical characteristics and complications of this group and, maybe in the future, open a pathway to therapeutic strategies.

954F

Next generation sequencing illuminates genetic heterogeneity in infantile cholestatic disorders. *S.M. Herbst[1], S. Hinreiner[1], C. Posovszky[3], F. Jochum[4], T. Rödl[1], J.A. Schroeder[6], T.F. Barth[5], M. Melter[2], J. Vermehren[2], U. Hehr[1].* 1) Center for and Department of Human Genetics, Regensburg, Bayern, Germany; 2) KUNO - Children´s University Hospital Regensburg, Regensburg, Germany; 3) University Medical Center Ulm, Department of Pediatrics and Adolescent Medicine, Ulm; 4) Department of Pediatrics and Adolescent Medicine, Evangelisches Waldkrankenhaus Spandau, Berlin; 5) University of Ulm, Institute of Pathology, Ulm; 6) Department of Pathology, University of Regensburg.

   Introduction A wide variety of genetic and non-genetic conditions may result in infantile cholestasis. The differential diagnosis is especially challenging due to their similar clinical presentation including biochemical parameters and often non-specific liver histology. More than 90 genes are associated with infantile cholestasis, ranging from monogenic forms to multisytemic disorders, which can primarily manifest as hepatic disease in infancy with or without cholestasis, thus preventing routine genetic workup by Sanger sequencing. However, rapid identification of an underlying genetic disease is essential (1) to discriminate patients who might benefit from liver transplantation in the long term, (2) to facilitate optimal treatment algorithms and (3) to prevent unnecessary and potentially harmful diagnostic procedures. **Method** Massively parallel sequencing of 93 genes associated with cholestatic disease in infancy was prospectively performed for 12 clinically well characterized patients in whom common causes of infantile cholestasis had already been excluded. **Results** Seven novel and nine known pathogenic mutations were identified in our patient cohort. The identified underlying diseases included for example one case of progressive familial intrahepatic cholestasis type 2 (PFIC2) (*ABCB11*: p.Gly628Trpfs*3, c.611+1G>A), one case of Niemann-Pick Disease Type C1 (*NPC1*: p.Glu391Lys, p.Arg116*), one case of atypical PFIC (*LBR*: p.Arg372Cys, *ABCB11*: p.Val1112Phe) and one case of autosomal recessive polycystic kidney disease (ARPKD) (*PKHD1*: p.Thr777Met, p.Tyr2260Cys). Three exemplary clinical cases of infants with cholestasis are presented and discussed in the context of their genetic, histopathological and electron microscopic findings (ARPKD, PFIC and NPC). **Conclusion** Our data confirm the power of massive parallel sequencing by targeted multi gene panels for infantile cholestasis. This testing strategy overcomes the complexity of the phenotype-based, candidate gene approach. Furthermore, our data highlight the critical impact of integrating clinical, histopathological and genetic data during the process of multi gene panel testing to ultimately pinpoint rare genetic diagnoses.

955W

Partial uniparental disomy results in homozygous 2p21 deletion in a male newborn with hypotonia-cystinuria syndrome. *L. Mao[1], A. Janssen[1], Y. Li[1,2], C. Dvorak[1], H. Andersson[1,2], T. Chen[1,2].* 1) Hayward Genetics Center, Tulane University School of Medicine, New Orleans, LA; 2) Department of Pediatrics, Tulane University School of Medicine, New Orleans, LA.

   Hypotonia-cystinuria syndrome (2p21 deletion syndrome, OMIM606407) is a contiguous gene syndrome caused by a homozygous deletion on chromosome 2p21 with clinical features including hypotonia, cystinuria, neonatal seizures, developmental delay, facial dysmorphism, and lactic acidosis. This syndrome has been reported from 11 families with 20 patients up to date. All of these patients were from families with consanguineous history. In this report, a male newborn was born by induced vaginal delivery at 38 weeks gestation to a 21-year-old mother, G2P1. The infant had low calcium related to low parathyroid hormone and severe hypotonia with poor feeding. Physical exam revealed no remarkable facial dysmorphology except for micrognathia, mildly flattened ears, and mildly abnormal palmer creases. Biochemical profile indicated that excretion of arginine and ornithine in urine was increased but cysteine level was normal. This patient may be too young to manifest other reported clinical features. There is no history of consanguinity in the family. The parents each have a healthy child with another partner. Cytogenomic analysis identified a homozygous 91 Kb microdeletion located within a 4.09 Mb region of copy-neutral loss of heterozygosity at 2p21 by CGH+SNP array. This alteration results in partial deletion of SLC3A1 and CAMKMT, as well as full deletion of the PREPL gene, consistent with the diagnosis of 2p21 deletion syndrome. Results from PCR amplification using primers against exon 7 and exon 10 of SLC3A1 and exon 2 of CAMKMT confirmed the loss of both copies of these two genes in this patient. Parental genetic analysis is pending to discover the origin of the deletion. To our best knowledge, this is the first case of 2p21 deletion syndrome identified from a family without consanguineous history. The homozygous deletion is very likely due to the location of the deletion within a partial uniparental disomy region.

956T

Evaluation of the parental origin of the chromosomes by using SNP genotype data from CGH+SNP microarray: Two cases with multiple congenital abnormalities. *K. Wakui[1,2], K. Takano[1,2], T. Kosho[1,2], T. Yamaguchi[1], Y. Takahashi[2], R. Kawamura[1], Y. Fukushima[1,2].* 1) Medical Genetics, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 2) Clinical and Molecular Genetics, Shinshu University Hospital, Matsumoto, Nagano, Japan.

Cytogenomic microarray is useful to detect pathogenic copy number changes in the subjects with birth defects, and copy-number-neutral absence of heterozygosity (AOH) and uniparental isodisomy can also be detected by SNP array. In addition, the data of SNP genotype can be used effectively to evaluate the parental origin of the chromosomes in some cases as shown below. Platform of the analysis: CGX™-SNP array (108K CGH + 60K SNP, PerkinElmer) and CytoGenomics software (Agilent) Case 1: No pathogenic CNVs, but a chromosome 20 has two AOH regions. One is 6.5-6.8Mb from pter, and the other is 10.9-11.6 Mb from qter. The SNP genotype data showed this subject was maternal uniparental disomy; uniparental isodisomy in both ends and uniparental heterodisomy in the middle of the region of chromosomes 20. Case 2: Gain CNVs in whole chromosome 14 in blood cells, but the average value of log2 ratio was 0.25. Metaphase FISH analyses showed that the subject had about 12% of mosaic trisomy 14. But fibroblasts of the subject had only around 1% of mosaic trisomy 14. We compared the SNP genotype data between the blood cells and fibroblasts. The results were that the chromosomes 14 in the cells of disomy 14 were biparental, and the additional chromosome 14 in the cells of trisomy 14 were paternal.

957F

Detection of copy number variation using shallow whole genome sequencing (CNVseq) as a cost-effective alternative to genomic microarrays. *B. Menten, M. De Smet, T. Sante, S. Vergult, A. Dheedene.* Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.

Genomic Copy Number Variations (CNVs) are an important cause of several human genetic disorders. Conventional karyotyping has been the gold standard for many years, but has been replaced by genomic microarrays, enabling a higher throughput and more important, a much higher resolution. More recently, next generation sequencing has been introduced in genetic diagnostics by means of targeted resequencing such as (clinical) exome sequencing. Despite the big advantage of whole exome sequencing (WES) to interrogate patient genomes at the nucleotide level to identify disease-associated single nucleotide variants, additional testing is still warranted to unequivocally detect, or rule out, CNVs. Although it has been proven that microarrays perform well for detecting submicroscopic CNVs, throughput is still rather limited, consumable cost is high, and resolution is fixed depending on the probe density of the microarray used. We therefor investigated whether genomic microarrays can be replaced by shallow whole genome sequencing for the detection of submicroscopic copy number variants (CNVseq) in a diagnostic setting. First, a statistical model was built to predict the theoretical amount of sequence reads necessary for a certain resolution. Subsequently, this model was evaluated by shallow whole genome sequencing at several read depths and bin sizes for several patient samples with clinically relevant genomic aberrations. We defined the number of reads necessary to achieve the same resolution as with commonly used microarrays and a thorough comparison between genomic microarrays and CNVseq was made. Moreover, we show that CNVseq results in a higher dynamic range and is more sensitive for the detection of low-grade mosaicism. Since library preparation is completely automatable and many samples can be pooled on one sequencing run, the hands-on time diminishes. Together with the plummeting cost of sequencing, this results in a more cost-effective workflow and a higher throughput. Furthermore, resolution is merely dependent on sequencing depth and is not limited to the probes present on the array, hence CNVseq enables more flexibility compared to genomic microarrays.

958W

A case of maternal UPD 20 in an infant with severe feeding difficulties, postnatal growth delay and mild developmental delay. *J.E. Watt[1], V.B. Bickley[1], K. Doudney[1], A.M. Kidd[1], G. Poke[2], M.D. Robertson[1], R.I. King[1].* 1) Genetics, Canterbury Health Laboratories, PO Box 151, Christchurch , New Zealand; 2) Genetic Health Service NZ, Wellington , New Zealand.

Maternal uniparental disomy for chromosome 20 [UPD(20)mat] is an emerging imprinting disorder. Characterisation of the phenotype has been hindered by case reports in patients with complete or partial trisomy 20 mosaicism, with only nine cases of isolated UPD(20)mat reported to date. The emerging phenotype is characterised by failure to thrive, profound difficulty feeding often requiring gastric tube intervention and growth retardation. Here we present a further case of isolated UPD(20)mat. Our case, a female infant, presented with postnatal growth delay, poor feeding, mild developmental delay - particularly gross motor skills, plagiocephaly and eczema. Birth weight was on the 25th centile, and poor feeding from birth resulted in a drop in weight to the 0.4th centile by 4 weeks. At 6 months of age a gastrostomy tube was inserted to improve nutrition. Chromosomal microarray analysis undertaken at 4 months showed no significant copy number change. However SNP analysis identified long continuous stretches of homozygosity (LCSH) confined to the terminal arms of chromosome 20. This pattern typifies UPD(20) resulting from a meiosis I nondisjunction error. Fluorescent in situ hybridisation (FISH), performed on cultured peripheral blood, excluded trisomy 20. Analysis of parental SNP data confirmed the UPD(20) to be of maternal origin. This case adds to the phenotypic data on this emerging disorder and is consistent with previously reported cases, with profound feeding difficulty leading to postnatal growth retardation being the predominant feature. .

959T

Opportunity and challenge of applying clinical exome for critical newborn patients: 1,000 clinical exome cases from a neonatal clinic. *H. Wang[1], L. Yang[1], B. Wu[1], Y. Lu[1], B. Sun[1], Z. Li[1], B. Liu[1], Z. Wei[2], Q. Zhao[2], H. Sun[2], F. Xia[3], W. Zhou[1].* 1) Pediatrics Research Institute, Children's Hospital of Fudan University, Shanghai, Shanghai, China; 2) Wuxi NEXTCODE ,288 Fute Zhong Road, Waigaoqiao Free Trade Zone, Shanghai 200131, China; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX, USA.

Background Birth defects are estimated to affect 3-6% of newborns worldwide, and are frequently caused by genetic factors. Accurate molecular diagnosis holds the key to prevent and treat congenital diseases. However, routine prenatal/newborn screens can only detect a small fraction of newborn disorders. On the other hand, many genetic diseases only manifest partially, or show higher variability at infant stage, which leaves little clue for clinician to cherry-pick a short list of genes to test. A more comprehensive approach, such as exome or genome sequencing, is desirable to cover all genetic possibilities. Methods At Children's Hospital of Fudan University, we continuously evaluated ~12,000 patients with critical conditions referred to our neonatal clinic by a systemic protocol. About 20% of them are suspected to have underlying genetic causes. If the patient with possible genetic disorder did not display distinctive clinical features, we will recommend the case for exome sequencing test. For each exome, the total targeted region is 50MB, and average coverage is more than 100X with 97% of targets covered by 20X or more. Parental studies were performed on candidate variants detected by exome test, if the parents are available. Results More than 1,000 of probands have been tested by clinic exome sequencing. Neuromuscular disorders are the largest group, and account for 43% of cases. 30% have structural anomalies. 14% of them have complex endocrine or immunology disorders. Due to the sample availability and economic issues, 27% of the cases are tested as Trio exome, and 64% of them have both parents available as familial controls. In total, about 46% of cases received diagnosis or possible diagnosis, and more than 75% have notable findings related. For cases received positive findings, 18% had improved treatment plan; 96% of them did further consultation of family planning on prenatal and preimplantation options. Conclusion Clinic exome is a valuable diagnostic tool in neonatal clinics, with optimal results from trio exome and prepared families. Patient management and family planning can greatly benefit from exome testing results. Pre-test counseling is challenging to convince both parents involved, especially for families with less-educated backgrounds.

960F

Evaluation of non-invasive prenatal testing for fetal aneuploidies in Korean population. *D.J. Kim[1], D.E. Lee[1], H.J. Kim[1], S.Y. Kim[1], S.Y. Park[1], J. Park[3], S. Ji[3], T. Yun[3], B.C. Lee[3], H.M Ryu[1,2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, South Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Dankook University College of Medicine, Seoul, Korea; 3) Molecular Diagnostics Team, IVD Business Unit, New business Division, SK telecom, Seongnam, Korea.

Background: Non-invasive Prenatal Testing (NIPT) usingcell-free fetal DNA from maternal plasma for fetal aneuploidies is expanding worldwide. In this study, we developed and validated a NIPT technology using cell-free fetal DNAfor fetal aneuploidy detection in Korean population. **Methods:** This preliminary study was performed at Cheil General Hospital and Women's Healthcare Center, Seoul, Korea. Only pregnant women who underwent both NIPT and amniocentesis were included in this study. NIPT results were compared with those of karyotype analysis. **Results:** Among 202 cases, 19 cases of fetal aneuploidies, including T21 (n=14), T18 (n=4), and T13 (n=1), were identified by NIPT and confirmed by karyotyping of amniotic fluid cells. There were no false-positiveand false-negativeresults. In the 101 male pregnancies, the median fetal DNA fraction was 10.65% and there was a trend towards an increasing fetal fraction from 15.0 to 22.5 weeks' gestation. The majority (98%) of pregnancies had a fetal DNA fraction>4%, which is generally the limit for accurate aneuploidy detection by NIPT. **Conclusion:** NIPT is highly reliable and accurate whenapplied to maternal DNA samples collected frompregnant women in the second trimester. Further large prospective studies are needed to adequately *assess* the performance of NIPT in first trimester pregnancies.

961W

Non-invasive prenatal detection of fetal aneuploidies by targeted semiconductor sequencing: A technical feasibility study. *Y. Xi[1,2], A. Arbabi[3,4], A. McNaughton[5], A. Hamilton[6], D. Hull[7], H. Perras[6], T. Chiu[6], S. Morrison[6], C. Goldsmith[6], E. Creede[6], C. Honeywell[6], M. Cloutier[6], N. Macchio[7], C. Kiss[7], X. Liu[5,8], S. Crocker[5,8], G.A. Davies[5,8], M. Brudno[3,4], C.M. Armour[6,9].* 1) Molecular Diagnostics Laboratory South, Alberta Health Services, Calgary, AB, Canada; 2) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada; 3) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 4) Center for Computational Medicine, Hospital for Sick Children, Toronto, ON, Canada; 5) Queen's University Genomics Laboratory at Ongwanada, Kingston, ON, Canada; 6) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada; 7) Kingston General Hospital, Kingston, ON, Canada; 8) Queen's University, Kingston, ON, Canada; 9) Regional Genetics Unit, Children's Hospital of Eastern, Ottawa, ON, Canada.

With the discovery of fetal cell-free DNA (cfDNA) in maternal plasma and the development of massively parallel sequencing techniques, non-invasive prenatal testing (NIPT) for common chromosomal aneuploidies has become possible and has been widely tested for clinical use with high sensitivity and specificity. However, it remains a relatively expensive test compared to traditional serum screening methods and expensive in many parts of the world. Thus there remains significant interest in developing other approaches that may utilize less costly sequencing approaches. Here, we designed a customized AmpliSeq panel, including 1067 primer pairs targeting specific regions on chromosome 21, 18, 13 and other chromosomes. Maternal plasma cfDNAs from a total of 235 pregnancies at high risk, including 30 affected with Trisomy 21 (T21) were extracted, and sequenced with Ion Torrent Proton sequencer using the custom panel. Analysis of the data and applying the derived risk score yields a bimodal distribution with the affected samples clustering near 1.0 and the unaffected near 0. For a risk score cut-off of 0.345, above which all would be considered 'high risk', all 30 T21-positive pregnancies were correctly predicted to be affected and 200 of the 207 non-T21 samples were also correctly predicted. At this cut-off there were 7 false positives (7/207; 3.3%) and no false negatives. With the described targeted sequencing approach on the semiconductor platform using a custom designed library and probabilistic statistical approach, we have demonstrated a feasible method to assess for fetal T21. Such an approach could be further developed for the common aneuploidies, potentially providing a cost and resource effective alternative to current methods.

962T

Copy number variation in Thai individuals with autism spectrum disorder. *N. Jinawath[1,2], A. Hnoonual[3], T. Tim-Aroon[4], K. Rojnueangnit[5], T. Hansakunachai[6], T. Sombuntham[7], R. Roongpraiwan[7], J. Worachotekamjorn[8], J. Chuthapisith[7], D. Wattanasirichaigoon[4], N. Ruangdaraganon[7], P. Limprasert[9].* 1) Program in Translational Medicine, Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Integrative Computational Bioscience Center, Mahidol University Salaya Campus, Nakhon Pathom, Thailand; 3) Graduate Program in Biomedical Sciences, Prince of Songkla University, Songkhla, Thailand; 4) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine, Thammasat University, Pathumthani, Thailand; 6) Division of Child Development, Department of Pediatrics, Faculty of Medicine, Thammasat University, Pathumthani, Thailand; 7) Division of Developmental-Behavioral Pediatrics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 8) Division of Child Development, Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand; 9) Division of Human Genetics, Department of Pathology, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand.

Autism spectrum disorder (ASD) is a group of common neurodevelopmental disorders defined by deficits in social skills, behavioral regulation, and communication. Although the cause of ASD is still largely unknown, copy number variations (CNVs) significantly contribute to the genetic causes of ASD. However, the contribution of these polymorphisms in non-European populations with ASD has not been well characterized. In this study, we aim to identify novel ASD-associated CNVs and to evaluate the diagnostic yield of a high-resolution genome-wide SNP-based microarray in a cohort of Thai patients with ASD of unknown cause. Illumina CytoSNP-850K BeadChip, which contains approximately 850,000 probes spanning the whole genome, was used to identify CNVs in the 114 patients who met the DSM-IV or DSM-5 criteria for ASD. The microarray data were analyzed using GenomeStudio and BlueFuse Multi v4.1 software. Multiple large curated CNV databases including Thai CNV database, which contains CNVs of the 3,017 general Thai individuals were utilized for interpretation of clinical significance. We detected pathogenic CNVs and variants of uncertain clinical significance (VUS) in 29 of 114 patients (25.4%), including five *de novo* CNVs, eighteen inherited CNVs, and six of unknown inheritance. Of the 29 patients, pathogenic CNVs were identified in 7 patients (6.1%) and VUS were found in 22 patients (19.3%). Of the VUS group, 7 CNVs were likely pathogenic and 15 CNVs were likely benign. Among the five *de novo* CNVs, one was a large deletion associated with the known 9q21.13 microdeletion syndrome, while the other four were rare duplications of 100-300 Kb in size. These gains contained both known ASD-associated genes such as *DPP10* and *CHL1*, and genes that are highly expressed in brain but have not yet been linked to ASD, namely *CNDP1* and *SERINC2*. In conclusion, this is the first SNP-based microarray study in a large cohort of Thai patients with ASD. Our findings provide supportive evidence that chromosomal microarray should be considered as a first tier genetic analysis for Thai patients with ASD of unknown cause. Additionally, CNV analysis using ancestry-matched controls in diverse population is a valuable tool for identification of novel ASD-associated genes. The CNVs detected in our study suggest a number of ASD candidate genes that warrant further investigation.

963F

MMP-8 and TIMP-1 expression in pressure ulcer management through negative pressure wound therapy in spinal cord injury patients. *R. Srivastava, M. Dwivedi, A. Bhagat, S. Raj.* Orthopaedic Surgery, KG Medical College, CSM Medical University, Lucknow, UP, India.

Introduction: Pressure ulcer (PU) is devastating co morbidity in SCI subjects. Our aim was to assess the concentrations of matrix metalloproteinase-8 (MMP-8) and tissue inhibitor of metalloproteinase-1 (TIMP-1) between PU treated with negative pressure wound therapy (NPWT) and PU treated with standard wound care and correlates it with healing outcomes. **Methods:** 44 patients of stage 3 & 4 pressure ulcers were randomized into two treatment groups. Group 1 (n=21) received NPWT and Group 2 (n=23) received conventional method of dressing. The tissues from base of PU were taken at week 0, 3, 6 and week 9 for analysis of MMP-8 and TIMP-1 by enzyme linked immune sorbent assay (ELISA). **Results:** Significantly reduced level of MMP-8 was observed in NPWT group at week 3 (p=0.014), week 6 (p=0.0007), week 9 (p=0.0001) as compared to standard wound care. Significant higher value of TIMP-1 was observed at week 6 (p=0.03) and week 9 (p=0.003) in NPWT group as compared to standard wound care. Significantly reduced length of PU in NPWT group was observed at week 6 (p=0.01), which further reduced more at week 9 (p=0.0001) as compared to standard wound care. Similarly significant reduction of width (p=0.0006) and depth (p=0.001) of PU was observed in NPWT at week 9 as compared to standard wound care group. Correlation analysis revealed that reduction of length, width and depth were positive correlated with MMP-8 and negatively correlated with TIMP-1. **Conclusion:** Negative pressure wound therapy leads to early wound healing by decreasing the concentration of MMP-8 and may increased the concentration of TIMP-1 as compared with standard wound care. New treatment strategies for chronic pressure ulcer healing could be directed towards molecular factors including maintenance of matrix metalloproteinases.

964W

Successful detection of pathogenic copy number variations in patients with neonatal/infantile cholestasis by next generation sequencing with Ion Torrent Personal Genome Machine system. *T. Togawa, T. Sugiura, K. Ito, T. Endo, K. Aoyama, Y. Negishi, K. Ohashi, S. Saitoh.* Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

Copy number variation (CNV) substantially underlies the pathogenesis of genetic disorders, and has become of much greater interest in a clinical setting. Array comparative genomic hybridization and multiplex ligation-dependent probe amplification (MLPA) analysis are the standard methods of detecting CNVs. Recently, next generation sequencing (NGS) technologies provide an alternative method of detecting CNVs by comparing the number of sequence depth. We developed a diagnostic panel of causative genes for neonatal/infantile intrahepatic cholestasis using the AmpliSeq and Ion Torrent Personal Genome Machine (Ion PGM, ThermoFisher Scientific) system. The panel included 18 genes, and number of amplicons, total targeted bases and exons were 551, 53.426kb, 355, respectively. Sequenced reads were analyzed through Ion Reporter software (IR, ThermoFisher Scientific) version 4.0/5.0. The IR could call CNVs by computational prediction according to two scores; the confidence score (CS) and the precision score (PS). Candidates of pathogenic CNV were collected as follows; 1) CS or PS indicated >10 points and called CNVs were expected to be potentially pathogenic, or 2) CS or PS < 10 points, however, called CNVs were strongly suspected of disease-causing. We analyzed 331 patients who were suspected of having a colestatic syndrome, such as Alagille syndrome (ALGS), progressive familial intrahepatic cholestasis type 1 (PFIC1), or Dubin-Johnson syndrome (DJS). The IR called 6 candidates of pathogenic CNVs; 3 *JAG1* CNVs in 3 patients with ALGS, 2 *ATP8B1* CNVs in 2 patients with PFIC, or 1 *ABCC2* CNV in 1 patient with no definitive etiology. In the candidates of *JAG1*, the IR predicted heterozygous whole gene deletions encompassing the length of 34 kb, and CS/ PS showed 296/296, 294/294, and 213/213 in respective patient. Regarding *ATP8B1*, each of two patients was predicted harboring a heterozygous deletion of exon 2 to 6 or exon 13, and CS/PS showed 48/48 or 3.2/3.2, respectively. In *ABCC2*, one homozygous deletion of exon 7 was predicted as CS/PS indicated 7.4/23. All candidates CNVs were successfully confirmed by MLPA analysis for *JAG1* using the SALSA MLPA probemix P184-C2 JAG1 (MRC-Holland), by quantitative polymerase chain reaction by LightCycler Nano system (Roche Diagnostics) for *ATP8B1*, or by Sanger sequencing for *ABCC2* detecting the break point. In conclusion, IR analysis using sequence depth by Ion PGM system successfully.

965T

Accurate detection of copy number variants in *ADME* **genes using PharmacoScan™ Assay.** *R. Varma, D. Black, J. Gollub, C. Bruckner, J. Law, M. Mittmann, T. Webster, A. Roter, M. Shapero, J. Schmidt.* Affymetrix, part of Thermo Fisher Scientific, Santa Clara, CA.

Copy number variations (CNVs) in genes relevant to drug absorption, distribution, metabolism, and excretion (ADME) have been characterized using several technologies. CNV detection using a standard single-nucleotide polymorphism (SNP) genotyping assay is efficient and advantageous because the genotypes of these genes are of utmost importance. We report here on the development of a CNV detection method for five ADME genes using PharmacoScan Assay, a cost-effective, high-throughput pharmacogenetics analysis solution. Copy number calls are compared to TaqMan® copy number assays, published data, or both, when available. We focus on CNV calling in five genes: CYP2A6, CYP2D6, GSTM1, GSTT1, and UGT2B17. CYP2A6 and CYP2D6 are particularly difficult to characterize because of the presence of closely related pseudogenes, which may form hybrids with their respective genes. Copy number states called include gains and homozygous and hemizygous losses. Both SNP and non-polymorphic probes are used for copy number detection and were selected based on sequence homology and response to copy number. Data analysis steps include signal summarization and normalization; calculation of log-ratios using normal diploid reference; correction of signals based on GC content and other covariate factors; median log-ratio calculation; and CNV calling based on thresholds for each region. Several hundred samples from the HapMap and 1000 Genome projects were analyzed using PharmacoScan Assay, and CNVs were called in eight regions: GSTM1, GSTT1, UGT2B17, two CYP2D6 sub-regions, and three CYP2A6 sub-regions. The predicted copy number states were compared to TaqMan results, and to sequencing results reported by the 1000 Genomes project. Results show high concordance in copy number calls for all eight regions between PharmacoScan Assay and other technologies. For CYP2D6 and CYP2A6, many samples were identified with differing copy number calls in the sub-regions, demonstrating the ability to distinguish copy number changes that are likely due to complex recombination events. In order to obtain equivalent results using TaqMan, multiple assays probing different locations within the gene are required. In contrast, the method presented here has the advantage of using a single assay to call CNVs in multiple sub-regions while also providing comprehensive genotypic information in genes of interest.

966F

Mechanistic analysis of interchromosomal insertional translocation.
T. Kato[1], Y. Ouchi[1,2], H. Inagaki[1,2], Y. Makita[3], S. Mizuno[4], H. Kurahashi[1,2]. 1) Divison of Molecular Genetics, Fujita Health University, Toyoake, Japan; 2) Genome and Transcriptome Analysis Center, Fujita Health University, Aichi, Japan; 3) Education Center, Asahikawa Medical University, Hokkaido, Japan; 4) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi, Japan.

   Interchromosomal insertional translocation is relatively rare chromosomal structural aberration, since they require at least three breaks followed by a segment from one chromosome is inserted into as interstitial region of different chromosome. Although several mechanisms such as double-strand-repairs via non-allelic homologous recombination or non-homologous end joining, and replication-based pathway have been implicated as mechanisms of chromosomal structural aberrations, the mechanisms of the insertional translocation is still unknown. In this study, we analyzed breakpoints and rearranged fragments by means of FISH, SNP array and next generation sequencing, to gain insight into their mechanisms of formation. To characterize the breakpoint diversity and insertion site complexities, we applied whole genome sequencing and mate-pair sequencing using next generation sequencer. All insertion cases are not simple structural aberration mediated by three breakages and three repairs. Inserted fragments are consisted of multiple pieces of genomic segments with a disorderly array and copy number alternations. All of the junctions have short microhomology or short insertional nucleotides. Polymorphism analysis revealed that insertion site sequences are of the same parental origin with normal counterpart homologous chromosome. One case showed that an insertion event is intermingled, i.e. maternal insertion site is inserted into another paternal chromosome. These data imply that the interchromosomal insertional translocation results from a meiotic nondisjunction event followed by a trisomic rescue at early post-zygotic stage with anaphase lag. Micronucleus that is formed from anaphase lagging chromosomes might undergo chromoanasynthes-like rearrangements. Pulverized chromosomes might be then inserted into another chromosome. It is likely that insertional translocation might be formed by several independent processes across pre- and post-zygotic events.

967W

Nomenclature matters: A collaboration to bridge the gap between ISCN and HGVS. *J. McGowan-Jordan[1,2], A. Simons[3], J. den Dunnen[4].* 1) Genetics, CHEO, Ottawa, ON, Canada; 2) University of Ottawa, Canada; 3) Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 4) Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

   Clear descriptions of clinical and experimental findings are essential for consistent communication amongst individual scientists, clinicians, and groups. In laboratory genetic studies this has happened in two distinct ways. For decades, the Standing Committee of the International System for Human Cytogenetic Nomenclature (ISCN) has established a standard scheme for the description of numerical and structural chromosome abnormalities. Detection of aberrations began with methods describing unbanded, then banded chromosomes, and later using technological advances including FISH, microarray and region-specific assays. When the molecular basis of various human diseases began to be identified, the need for a standard system of describing variants at a base-pair level became evident. The Human Genome Variation Society (HGVS) established such guidelines and continues to update and improve the originally proposed scheme, enabling the description of variants based on different reference sequences and covering the DNA, RNA and protein levels.   With widespread uptake of advanced sequencing technologies large rearrangements, well beyond those typically identified by more classical sequencing, are being identified. Nomenclature to describe these changes, both within and between chromosomes, was outside the scope of the existing ISCN and HGVS schemes and required the establishment of a scheme that would be useful for both the Cytogenetics and Molecular Genetics communities. In a unique and historic collaboration, the ISCN Standing Committee and HGVS Working Group on Sequence Variant Description and Nomenclature have worked together to draft a scheme that will meet the needs of both communities and provide consistent description of large intra- and inter-chromosomal rearrangements. This scheme, published as a new chapter in ISCN (newly renamed as An International System for Human Cytogenomic Nomenclature) 2016, combines aspects of both HGVS and ISCN, so that information on both the chromosomal and base-pair level is readily evident. Elements and examples of the combined scheme will be presented to illustrate its utility.

968T

Molecular genetic analysis of 30 families with Joubert syndrome and related disorders. *T. Suzuki[1,2], N. Miyake[1], Y. Tsurusaki[3], M. Nakashima[1], H. Saitsu[1,4], S. Takeda[2], N. Matsumoto[1].* 1) Depertment of Human Genetics, Yokohama City University Graduate School of Medicine, yokohama, Japan; 2) Department of Obstetrics and Gynecology, Juntendo University Faculty of Medicine, Tokyo, Japan; 3) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 4) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

Joubert Syndrome and Related Disorders (JSRD, MIM 213300) are rare recessive disorders characterized by hypoplasia/aplasia of cerebellar vermis and a deep posterior interpeduncular fossa, thickened and elongated superior cerebellar peduncles known as "molar tooth sign" on brain MRI. JSRD is genetically highly heterogeneous and at least 28 disease mutant genes have been found in JSRD so far. To elucidate genetic causes in JSRD, we performed whole exome sequencing (WES) in 24 JSRD families newly recruited. Including six families which we previously reported, genetic and clinical aspect of 30 JSRD families (27 Japanese families, two Omani families, and one family of half-Japanese and half-Paraguayan mother and Paraguayan father) were analyzed. We run the exon-captured samples on Genome Analyzer IIx or HiSeq (Illumina platform). Candidate pathogenic variants were validated by Sanger sequencing, and the mutation segregation was checked when familial samples were also available. In addition, we analyzed copy number variations (CNV) using the eXome Hidden Markov Model and Nord program which are WES-based CNV detection tools. As a result, we identified the causative mutations in 25 out of 30 families (83.3%). We did not detect any pathogenic CNVs. Among eight genes which mutations were identified in 27 Japanese families, mutations of *TMEM67* (7/27, 25.9%) and *CEP290* (6/27, 22.2%) were most frequent. Interestingly, nine of 12 disease alleles of *CEP290* were c.6012-12T>A (75.0%), while it has never been reported in other ethnic studies. Additionally, we found compound heterozygous mutations with *C5orf42* in two Japanese families. Among of them, three mutations (c.3577C>T, c.3557delA, and c.3599C>T) in *C5orf42* were previously reported in Japanese families, but not in other ethnic groups. Therefore, c.6012-12T>A in *CEP290* and c.3577C>T, c.3557delA, and c.3599C>T in *C5orf42* could be founder alleles in a Japanese population. Furthermore, in two families (2/30, 6.7%), we found concomitant mutations consisting of compound heterozygous mutations in *TMEM67* and *RPGRIP1L* in one Japanese family, and homozygous mutations in *TMEM67* and *BBS1* in an Omani family. *BBS1* is a causative gene with Bardet-Biedl syndrome. Since they showed severe and/or complex clinical features, the dual effects of different mutant genes are suggested.

969F

Detection of constitutional copy number variation in saliva DNA using the Affymetrix Cytoscan HD array. *J. Reiner[1], L. Karger[1], G. Mamtora[2], R. Duttagupta[2], E. Fung[2], L. Mehta[1], L. Edelmann[1], S.A. Scott[1].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Cytogenetics Research and Development, Thermo Fisher Scientific, Santa Clara, CA.

Chromosomal microarray (CMA) analysis is considered a first-tier test for the detection of constitutional abnormalities in individuals with idiopathic developmental conditions and/or congenital anomalies. Currently, postnatal clinical CMA testing is most commonly performed using DNA extracted from lymphocytes in whole blood. However, the collection of saliva samples for downstream molecular testing can offer several advantages over blood, including a non-invasive collection procedure and increased sample stability. Previous studies using saliva or buccal samples for single nucleotide polymorphism (SNP) genotyping microarrays have reported mixed results with respect to SNP calling and copy number variation (CNV) detection, likely due to the presence of variable concentrations of bacterial DNA among the saliva samples. As such, further investigation is warranted before implementing saliva DNA for clinical CMA testing. In order to test the effect of high bacterial content on CMA performance, we tested DNA from six saliva samples collected using the Oragene Dx kit (OGD-500; DNA Genotek) with a broad range of PCR-quantified bacterial DNA content (3.1-20.8%) on the Affymetrix CytoScan HD array. No significant differences in mapd values were detected between samples, suggesting that bacterial mass does not adversely affect Cytoscan HD array quality metrics. We also compared the performance of matched blood and saliva DNA from 10 healthy adult controls and an additional 10 postnatal cases with clinically-significant copy number aberrations previously detected by clinical CMA analysis on blood DNA. Analysis of the aggregate blood and saliva samples showed comparable mapd and snpQC values, as well as a similar number of absence of heterozygosity (AOH) calls. Despite a trend towards increased CNV call rate in saliva, concordance between the matched blood and saliva samples was greater than 80% for all CNV and AOH calls. Importantly, the clinical utility of using saliva DNA for CMA testing with the Cytoscan HD array was further validated by the detection of all clinically-significant copy number aberrations in saliva that were originally identified in blood DNA.

970W

Triaging complex cases for genomic tests in South Australia: A multi-disciplinary team approach. *P.A. Kaub, K.S. Kassahn, J.M. Fletcher.* Genetics & Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia.

New genetic testing platforms offer much promise, but are still a relatively limited resource in clinical diagnostic settings in Australia. Our clinical diagnostic service uses a multi-disciplinary team (MDT) forum, where medical scientists and bioinformaticians performing genetic testing and analysis directly interact with primary clinicians and pathologists to determine the most appropriate testing for complex clinical cases. Routine genetic tests are sent directly to relevant laboratories. For complex cases however, we have developed a systematic approach for case presentations and decision criteria to assist determination of the most appropriate testing for each case, based on both clinical and laboratory information. We consider the management implications, evidence supporting genetic testing indication and utility, test availability, service capacity and in-house expertise, as well as expected turnaround time. A recommendation for each case is formed after open and robust discussion between the MDT meeting members. Outcomes include recommendations for phenotype-targeted next generation sequencing (NGS) gene panels, clinical exome, whole exome or whole-genome sequencing, array-based testing and MLPA. Where there are issues of capacity or lack of expertise, samples have been referred to research colleagues or external laboratories. In just under two years, this consensus-based, open discussion platform has reviewed approximately 100 complex clinical cases. The majority have been Paediatric patients, with the greatest number referred from Clinical Geneticists, however Immunologists, Neurologists, Cardiologists, Metabolic, Renal and General Physicians have also been represented. Common phenotypes include neurological, syndromic, dysmorphic/anatomic, developmental/intellectual delay or metabolic conditions. This process is scaleable as each new panel moves from the complex to standard pathway. The MDT meeting format has been key to our successful implementation of NGS and ability to meet increased demand through optimised test selection that is phenotype-driven. Scientists, researchers, clinicians & bioinformaticians are all essential members of our MDT meeting format, maximising the utility of diagnostic services and assisting best clinical outcome for patients.

971T

Assessing the clinical genomic load and outcome within the eMERGE III network. *M. Leduc[1], D. Muzny[2], K. Walker[2], E. Venner[2], E. Kudalkar[3], F. Funke[3,4], A. Gordon[5], G. Jarvik[5], H. Rehm[3,4], Y. Yang[1,6], C. Eng[1,6], R. Gibbs[2] on behalf of the Electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Laboratory for Molecular Medicine at Partners Health-Care Personalized Medicine, Cambridge, MA; 4) Department of Pathology, Harvard Medical School/Massachusetts General Hospital, Cambridge, MA; 5) Dept. of Medicine (Medical Genetics), University of Washington, Seattle, WA; 6) Baylor Miraca Genetics Laboratories, Houston, TX.

The eMERGE (Electronic Medical Records and Genomics) network is a NHGRI funded national consortium that aims to advance medical care by implementing genomic medicine and integrating Electronic Medical Records (EMR) systems. In eMERGE III, we will resequence 109 genes in a CAP and CLIA environment in addition to testing ~1,500 unique single nucleotide variants. The test will use a CARE seq DNA capture panel to test 25,000 DNA samples from a cohort submitted by the 9 eMERGE clinical sites, including symptomatic and asymptomatic individuals. The 109 eMERGE genes include the 56 ACMG medically actionable genes and 53 genes of interest selected by the clinical sites, 92 of which are of clinical significance and 66 of which are medically actionable. Two core clinical laboratories will perform the sequencing, clinical assessement of the variants and provide an individualized clinical report with pathogenic (P) and likely pathogenic (LP) variants. Carrier status will not be reported. The clinical report will be linked to the patient's EMR while the vcf datafile will be transmitted to the clinical sites for further research. As a preliminary analysis, we used our internal clinical exome (WES) databases to assess the performance of our DNA capture panel reagent and predict the number of positive clinical reports to likely be required for 'healthy' asymptomatic individuals. Based on data from 647 test samples, only 14 exons had no or low coverage using the criteria of 10% of samples with coverage <20X, illustrating the overall high quality of the capture reagent. We then catalogued all known P and LP variants detected and/or curated by our laboratory within the 92 eMERGE genes and 1,500 variants (n=5,663 P/LP variants). We investigated 6,000 WES patients for P/LP variants and identified 1,835 P/LP variants in 1,591 patients (28%). Since the clinical WES population is pediatric and symptomatic, we removed the variants related to the patient's phenotype and interrogated the medically actionable genes and variants. We identified 166 patients with either one P/LP heterozygous variants in autosomal dominant disorders (n=163) or a hemi and homozygous variants in autosomal recessive and X-linked disorders (n=3). While additional variant curation and harmonization of this categorization is underway between the two core laboratories, we estimated that approximately 3% of asymptomatic participants enrolled in the eMERGE consortium will require a positive reportable finding.

972F

The Wessex Clinical Exome Pilot: A comprehensive service evaluation of focussed exome sequencing in clinical diagnostics. *R. Pengelly[1], D. Ward[2], E. Baple[3], S. Thomas[2], D. Hunt[3], C. Mattocks[2], S. Ennis[1], Wessex Clinical Exome Pilot.* 1) Human Genetics & Genomic Medicine, University of Southampton, Southampton, United Kingdom; 2) Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, UK; 3) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK.

Next-generation sequencing (NGS) has revolutionised the practice of medical genetic research, allowing the massively parallel sequencing of DNA, and is now increasingly applied to clinical diagnostics. Focussed-exome sequencing (FES), the sequencing of the exonic regions of ~4,500 genes previously implicated in human disease (the 'morbidome') offers an intermediate option between the sequencing of all genes and small targeted gene panels. The Wessex Clinical Exome Pilot (WCEP) is a partnership between academia and healthcare, involving the University of Southampton, University Hospital Southampton NHS Foundation Trust and Salisbury District Hospital NHS Foundation Trust. The WCEP aims to perform a comprehensive service evaluation of FES in routine clinical diagnostics. A through evaluation of FES kits was performed and the Illumina TruSight One focussed exome kit will be applied to a cohort of individuals with molecular diagnoses previously determined by routine clinical diagnostic methods. Furthermore, we will investigate ~150 diverse cases for which clinical targeted gene panels were being performed in parallel. We will comprehensively evaluate these cohorts for operational parameters including diagnostic rate, turn around time and cost, as well as determining the robustness of the technology in clinical diagnostics and improving our bioinformatics tools to allow maximal utility of these data and conducting further research utilising this dataset. The development during the WCEP will lay strong foundations for ongoing commercial clinical diagnostic FES testing in the Wessex region, leading to improved economy and better patient outcomes for those undergoing genetic testing.

973W

Evaluation of the truXTRAC™ DBS DNA Kit for extraction of DNA from dried blood spot samples for use in NGS library preparation. *K. Ramachandran[1], R. Daber[1,2], J. Jackson[1], L. Vu[1], J. Adigun[1].* 1) GeneDx, 205 Perry Parkway, Gaithersburg, MD 20877; 2) Bio-Reference Laboratories, Inc, 481 Edward H. Ross Drive Elmwood Park, NJ 07407.

Dried blood spots (DBS) provide an easy and inexpensive way to collect and store peripheral blood specimens from infants, children and adults. The ability to run molecular or clinical assays on this type of specimen allows for easier specimen shipments and less invasive procedures for patients. This is also a convenient method for the long term room temperature storage of materials suitable for molecular assays and also minimizes storage and archival space. The truXTRAC DBS DNA Kit from Covaris® is designed for controlled and efficient extraction of NGS-grade DNA from DBS samples using Adaptive Focused Acoustics (AFA™). According to the manufacturer, Covaris® AFA enables sample rehydration while providing simultaneous cell lysis and DNA shearing resulting in high-yield and high-quality DNA. Here we report the successful use of the truXTRAC DBS DNA Kit to extract DNA from a single 3mm punch of a DBS cards for generation of NGS libraries. Hybridization based capture was then performed on the libraries to assay for germline mutations in an inherited cancer panel. Downstream data analysis showed that data quality and variant calling for libraries prepared from DBS DNA were indistinguishable from those prepared from DNA extracted from 1ml of whole blood. This new protocol provides more evidence for the efficacy of DNA extracted from DBS cards in the molecular diagnostics of genetic disorders.

974T

Toward objective interpretation of sequence variants in clinical testing. *J. Weber[1], Z. Ye[2], Z. Wang[1].* 1) PreventionGenetics, Marshfield, WI; 2) Marshfield Clinic Research Foundation, Marshfield, WI.

Interpretation of sequence variants is certainly one the greatest current challenges in clinical genetics. To date, variant interpretation has been a highly subjective exercise. In an effort to make interpretation more objective, we have utilized several simple, quantitative measures of pathogenicity. These measures include allele frequency differences between cases and controls, and the fraction of times a particular variant occurs together with a second plausible pathogenic variant in patients affected with recessive disease. We demonstrate the utility of these approaches with several examples including autosomal dominant Malignant Hyperthermia Susceptibiltiy via *RYR1* variants and Autosomal Recessive Polycystic Kidney Disease via *PKHD1* variants. Our approaches have limitations and potential pitfalls and therefore must be used with caution. These approaches also require quantitative definitions of Pathogenic, Likely Pathogenic, Likely Benign and Benign. Nevertheless, these approaches are important first steps in transforming sequence interpretation from an art to a science.

975F

A pathogenicity/phenotype correlation dual-scoring system for precise clinical interpretation of genetic variants: In the setting of expanded sequencing-based carrier screen for 20,000 individuals. *F. Xia, L. Meng, J. Zhang, R. Xiao, C. Eng, A. Beaudet.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background Population-based carrier testing is in the process of transitioning from targeted mutation analysis on a small number of genes to full sequencing of all exons for hundreds of targeted genes. This represents a major challenge where large numbers of variants need to be interpreted very rapidly, for both pathogenicity and phenotype correlation.   Methods Since 2015, Baylor Miraca Genetics Laboratories launched the GeneAware® comprehensive carrier screen test. The main component is sequencing the coding regions of a 159-gene panel associated with autosomal recessive and X-linked conditions. As the candidate pool, we collected over 30,000 variants from locus specific mutation databases, including HGMD®, ClinVar. For accurate interpretation, we developed a scoring system giving each variant both pathogenicity ranks resembling the new ACMG 5-category system, and phenotype scores including four phenotype categories: classic severe (1), atypical mild (2), subclinical condition (3) and to-be-determined (4). The evidence for scoring is classified into types as clinical cases, population genetics, and functinoal data and theoretical predictions. Variants scored as pathogenic/likely pathogenic are reported.   Results To date, over 20,000 clinical cases have been analyzed, and >10,000 variants have been curated and assigned with both pathogenicity and phenotype correlation scores. Among the variants scored as pathogenic or likely pathogenic, 95.7% of them have classic phenotype score. Atypical mild and subclinical are 3.9% and 0.4%, respectively. However, in 1000 genome and ExAC databases, the atypical and subclinical variants are 250% more frequent than the severe ones. The observation is consistent with that mild alleles may survive better during natural selection. For variants associated with classic severe phenotype, evidence from clinical cases and functional data are consistent; population variant frequencies are in general supportive. For variants of mild and subclinical categories, clinical cases become dominantly important, while other types of evidence show ambiguity.   Conclusion The pathogenicity/phenotype correlation dual scoring system allows us to give precise interpretation of clinical testing results. Given the wide and expanding phenotypic spectrum of genetic disorders, professional curation in combination of informatics tools is crucial to achieve the optimal results for complex genetic testing like panel, exome and whole genome sequencing.

976W

Variant classifications are highly concordant in ClinVar, but with variability in genes from different disease areas. *S. Yang, S. Lincoln, K. Nykamp, T. Hambuch, Y. Kobayashi, S. Topper, R. Nussbaum.* Invitae, San Francisco, CA.

As the number of laboratories offering genetic tests grows, the potential for inconsistent variant classifications increases. Public databases of clinically classified variants afford us, for the first time, the ability to systematically evaluate this issue. We evaluated pathogenicity assessments in the ClinVar April 2016 XML file. We included all unique variants from genes Invitae currently offer with at least 2 classifications submitted by established clinical laboratories including data from Myriad Genetics submitted via the Sharing Clinical Reports Project (SCRP). 24,455 total classifications of 10,000 unique variants were available for comparison. The largest data sets exist for hereditary cancer (48%) and cardiovascular (33%) genes, with smaller data sets available in: neurology, excluding epilepsy (7%), epilepsy (7%) and inherited metabolic disorders (4%). We first focused on differences between putatively disease causing variants classified as Pathogenic (P) or Likely Pathogenic (LP), as opposed to those which are not clinically actionable: Variant of Uncertain Significance (VUS), and Likely Benign (LB), or Benign (B) variants. Counting each unique variant as a data point, concordance among different labs was high: variants in cancer genes had the highest agreement (98.9%) while inherited metabolic disorder variants had the lowest (92.7%). However, when the concordance is defined with finer detail as pathogenic (P/LP), vs. uncertain (VUS) vs. benign (B/LB), we see lower agreement. Variants in cancer, neurological and inherited metabolic disorder genes maintain over 90% concordance; while cardiology and epilepsy are the least concordant at 85.7% and 76.1% respectively. The major diagnostic labs we investigated generally agree on the clinical significance of potentially actionable (P) vs. not actionable (NP) findings. Classification of variants in genes in different disease areas, however, does vary in their concordance. This may arise from differences in the depth and sophistication of knowledge about these diseases, the complexity and diversity of genetic causes of the disease, the number of labs participating in that area and other factors. The small number of differences is important to resolve collaboratively through peer-review of such data, and serves as an important form of laboratory quality control. The learning from such comparisons helps to improve standards and knowledge among all clinical genetics professionals.

977T

Segmental genomic deletions and splicing mutations causing tuberous sclerosis complex (TSC). *L. Almeida[1], J.P.G. Almeida[2], L. Masulk[3], M. Richartz[3], N. Miranda[3], I.M.W. Silva[1], S. Chiavegatto[4], M.R. Passos-Bueno[1], S.A. Antoniuk[3], S. Rosemberg[2], L.A. Haddad[1].* 1) Department of Genetics and Evolutionary Biology, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; 2) Department of Pediatrics, Division of Child Neurology, Faculdade de Medicina da Santa Casa de São Paulo, São Paulo, Brazil; 3) Department of Pediatrics, Universidade Federal do Paraná, Curitiba, Brazil; 4) Department of Pharmacology, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil.

Tuberous sclerosis complex (TSC) is an autosomal dominant hamartomatosis with variable expression, affecting notably the brain, kidneys, heart, skin, lungs, and retina. Mutations in either tumor suppressor genes *TSC1* (9q34) or *TSC2* (16p13.3) are responsible for TSC. The proteins expressed by those genes, respectively hamartin and tuberin, form a heterodimer that inhibits the mammalian target of rapamycin complex 1 (mTORC1), controlling cell growth and proliferation. Comprehensive screening for *TSC1* and *TSC2* DNA variants are well established, and disclose in average 70% of pathogenic (loss-of-function) DNA variants in the *TSC2* and 30% in the *TSC1* genes. Different clinical and research groups worldwide have reported that 75-90% of TSC-causing mutations lie in the *TSC1* and *TSC2* coding regions (frameshifting, nonsense and missense DNA variants). Increasing evidence indicates that DNA variants in non-coding sequence and copy number variations (CNV) may be responsible for *TSC1* and *TSC2* gene inactivation, including patients previously classified with no-mutation identified (NMI) (Nellist, M *et al.*, 2015, *BMC medical genetics*). Here we report on segmental genomic deletions and splicing mutations detected altogether in 24% of TSC patients from a Brazilian cohort. Forty-one patients with definite clinical diagnosis of TSC had leukocyte DNA previously analyzed by Sanger sequencing (Almeida, L. *et al.*, 2015, *Am Soc Hum Genet*). Deletions identified by multiplex ligation-dependent probe amplification (MLPA) analyses were submitted to quantitative PCR (qPCR), and *TSC1* and *TSC2*-targetted exome sequencing. Data from the latter approach were analyzed using a set of statistical and computational tools named XHMM (Frommer *et al.*, 2012; *Am J Hum Genet*). A high rate of mutation detection (97.6%) was observed for this group of patients (N=41) that had been referred to our laboratory from Neurology services from two tertiary referral hospitals in Brazil. Three patients (7%) had segmental deletions (two in *TSC1*, and one in *TSC2*); one of them with a mosaic pattern confirmed by different techniques. The *TSC2* gene presented splicing mutations in seven patients (17%). No splicing mutation was detected in the *TSC1* gene. In addition, the only patient that remained NMI had a single nucleotide variant in one *TSC2* intron, which needs to be functionally evaluated.

978F

High-throughput copy number variation detection in the application of clinical carrier testing. *L. Meng[1], C. Shaw[1], T. Chiang[3], W. Jin[2], T. Luong[2], T. Zhang[2], Z. Chen[1,2], F. Xia[1,2], J. Zhang[1,2], C.M. Christine[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Human Genome Center, Baylor College of Medicine, Houston, TX.

Statement of purpose: We developed a novel 500 kb clinical carrier testing panel which includes full gene next-generation sequencing for 152 genes, triplet repeat analysis for FMR1, and analysis for recurrent CNVs in 15 genes. As of May 2016 we have reported over 10,000 clinical samples for this test. We continually optimize our pipeline to improve data quality and overall efficiency. Here, we explored the possibility of using NGS data to perform CNV analysis for non-homologous genes as part of the clinical carrier testing panel. Methods used: Normalized read depth from NGS data are computed for each target exon in the given sample and copy number was calculated by reference it to the median read depth within the capture midpool. To compare the NGS CNV pipeline with the current system, Fluidigm qPCR assay, we performed CNV analysis by both methods and compared the results for *CFTR* and *DMD* from 5000 de-identified clinical samples. Summary of results: The Fluidigm qPCR assay were designed to target for *CFTR* exon2 and 3 only, and detected only 1/5000 samples with exon2 deletion. By expanding the detection region to full gene by NGS, CNV detection rate was increased to 4/5000 (1 case with exon19 deletion, 1 case with exon 19-21 deletion, and 1 case with complex deletion with exon 4-8 and exon 12-21), without significantly affecting the false positive rate (0.21% by qPCR assay versus 0.45% by NGS). Since Fluidigm qPCR assay covers a great majority of the *DMD* gene, both methods successfully detected samples with CNVs over the *DMD* region (eight cases with copy number loss and three cases with copy number gain). But it should be noted that NGS analysis still has higher coverage (79 exons from the major isoform and 6 exons from other important isoforms) than Fluidigm qPCR assay (54 exons). Even with our best effort for analysis improving, false positive samples remained as high as 12.16% for Fluidigm qPCR assay, while only 0.57% samples were called false positive by standard NGS algorism. A great majority of false positive deletions in qPCR testing are caused by rare SNV located in the primer/probe region resulting in allele drop out, while these SNPs have little effect in NGS-based CNV results. In summary, compared to qPCR-based testing, NGS-based CNV analysis is superior in coverage, detection rate, and specificity, and may replace dedicated CNV assays in clinical carrier testing.

979W

The clinical application of exome CNV detection in a large cohort of patients with various genetic disorders. *R. Pfundt[1], M. del Rosario[1], L.E.L.M. Vissers[1], M.P. Kwint[1], I.M. Janssen[1], H.G. Yntema[1], M.R. Nelen[1], D. Lugtenberg[1], E. Kamsteeg[1], N. Wieskamp[1], A.P.A. Stegmann[2], S.J.C. Stevens[2], R.J.T. Rodenburg[3], A. Simons[1], A.R. Mensenkamp[1], T. Rinne[1], C. Gilissen[1], H. Scheffer[1,2], N. de Leeuw[1], J.A. Veltman[1,2], J.Y. Hehir-Kwa[1].* 1) Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, Gelderland, Netherlands; 2) Department of Clinical Genetics, GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands; 3) . Nijmegen Center for Mitochondrial Disorders, Department of Pediatrics, Radboud university medical centre, Nijmegen, The Netherlands.

Copy number variation (CNV) is a common source of genomic variation and an important genetic cause of disease. Microarray based CNV analysis has become a first tier diagnostic test for patients with intellectual disability, with a diagnostic yield of 10-20%. However, for most other genetic disorders the role of CNVs is less clear and genetic studies are generally limited to the study of single nucleotide variations (SNVs) and other small variants. With the introduction of exome and genome sequencing it is now possible to detect both SNVs and CNVs in a single test. Here, we have performed exome based CNV screening on data from 2,603 patients affected by a range of 14 different genetic disorders for which exome sequencing was performed in a diagnostic setting. Using read depth analysis we identified 131 clinically relevant CNVs ranging in size from 729bp – 15.3 Mb. This results in over 50 conclusive diagnoses, with an overall average diagnostic yield of ~2%. CNVs were found exerting both dominant and recessive effects, as well as CNVs unmasking a recessive mutation that results in pathogenic compound heterozygous events. This study shows that CNVs play an important role in a broad range of genetic disorders and illustrates how these CNVs can be readily detected from exome sequencing, without the need for additional genetic laboratory testing. Combining SNV and CNV analysis maximizes the diagnostic yield of whole exome sequencing, making it very suited as a first tier genetic test for many disorders. This brings us closer to single test genomics.

980T

What have genes got to do with it? Analyzing gene content across interpreted CNVs in the Clinical Genome Resource Structural Variation Interpretation Working Group. *D.I. Ritter[1], E. Riggs[2], E. Anderson*[3], A. Cherry[4], S. Kantarci[5], H. Kearney[6], C.P. Lorentz[6], J.M. Meck[7], A. Patel[8], S.E. Plon[1], G. Raca[9], S. South[10], E. Thorland*[6], R. Vanzo[11], D. Pineda-Alvarez**[12], S. Aradhya**[13], C.L. Martin[2] on behalf of the Clinical Genome Resource, Co-chairs of ClinGen *Dosage WG & **SV Interpretation WG.* 1) Baylor College of Medicine, Texas Children's Hospital, Houston, TX; 2) Geisinger Health System, Danville, PA; 3) ARUP Laboratories, University of Utah, Salt Lake City, UT; 4) Departments of Pathology and of Pediatrics, Stanford University Medical Center; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; 6) Mayo Clinic, Rochester, MN; 7) GeneDx, Gaithersburg, MD; 8) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX; 9) Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Keck School of Medicine, University of Southern California; 10) 23andMe, Mountain View, CA; 11) Lineagen Inc, Salt Lake City, UT,; 12) Courtagen Life Sciences Inc, Woburn, MA; 13) Invitae, San Francisco, CA.

Structural variation (SV) through copy number variants (CNVs) is an important disease mechanism implicated in phenotypes ranging from developmental delay to cancers. The Clinical Genome Resource (ClinGen) SV Interpretation Working Group represents a team of investigators from multiple institutions and diagnostic laboratories collaborating with the American College of Medical Genetics and Genomics (ACMG) to update and revise clinical CNV interpretation guidelines. To understand the impact of gene content in interpreting CNVs for pathogenicity, we solicited CNVs that were evaluated in the last 5 years from participating diagnostic laboratories. CNVs requested included deletions of >200kb and <5MB or duplications >500kb to <10MB, with inheritance information and interpreted as: "pathogenic", "likely-pathogenic", "VOUS" (variant of unknown significance), "likely-benign" or "benign." Additionally, we supplied a questionnaire to obtain general reporting practices. We annotated CNVs with RefSeq genes (hg19) and converted all relevant information into a standardized 4-field BED format. We have completed one round of analysis on a subset (~500 CNVs) and expect to collect ~5K interpreted CNVs from reporting laboratories. We found that the median gene content was significantly higher (using a pairwise t-test) between pathogenic and VOUS CNV deletions (p<2e-16; 20 and 8 genes, respectively) with a similar trend for CNV duplications (p<2e-16; 42 and 9 genes, respectively), although the minimum number of genes in a pathogenic CNV could be one in cases of a well-known disease gene. However, likely-pathogenic and VOUS CNVs were not as easily separable by gene content. Ongoing analysis of the larger set will exclude CNVs containing well-documented disease genes by intersection with three databases of gene dosage or truncating variant tolerance: ClinGen Dosage Sensitivity Map, DECIPHER Haploinsufficiency Index and ExAC Loss of Function Intolerance. Inheritance information, such as gene content of *de-novo* vs inherited CNVs will also be analyzed. This work is revealing how current diagnostic laboratories are using gene content in CNV interpretation. The results will inform suggested gene content windows for CNV interpretation, particularly critical for cases of CNVs lacking clearly causal genes. We will incorporate this work into the larger ClinGen SV Interpretation Working Group effort to systematically update and revise the ACMG guidelines for CNV interpretation.

981F

CentoMD®, a comprehensive genotype-phenotype database for rare diseases. *G.E. Oprea[1], D. Trujillano[1], Y. Schmitz[1], A.M. Bertoli-Avella[1], R. Abou Jamra[1,2], O. Brandau[1], P. Bauer[1], A. Rolfs[1,3].* 1) Centogene AG, Rostock, Germany; 2) Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany; 3) Albrecht-Kossel-Institute for Neuroregeneration, Medical University Rostock, Rostock, Germany.

Our ability to discover genetic variants in a patient runs far ahead of our ability to interpret them medically. Databases with accurate and curated descriptions of the causal relationship between the variants and the phenotype are valuable since these are critical tools in clinical genetic diagnostics. Here we introduce CentoMD®, a comprehensive genotype-phenotype database with the main focus on rare diseases. CentoMD® is browser based software that enables access to a comprehensive, curated, and growing repository of genetic and clinical information. Its main goals are to aid medical professionals in the evaluation of genetic variants and to enhance the validity of the genetic analytical workflow. CentoMD® correlates the clinical information of consented patients and probands of different racial and ethnic backgrounds (>100 countries) with a large dataset of genetic variants and, when available, biomarker information. More than 74,000 genetically screened individuals are documented in CentoMD®, resulting in more than 200 million variant detections. Approximately 56% of the clinically relevant and uncertain variants in CentoMD are novel. Since every variant retrieval is systematically logged, we have implemented an automated re-contacting tool informing users whenever a variant classification has changed to allow them updating their clinical reports if necessary. Taken together, these unique features, fully embedded in a CLIA/CAP accredited quality management system, allow for excellent data quality and thus enhance patient safety in genetic testing.

982W

Data-driven Sanger confirmation guidelines for next generation sequencing calls. *S.H. Askree[1,2], C. da Silva[1,2], L. Bean[1,2], C. Collins[1,2], A. Tanner[1,2], J. Alexander[1,2], Z. Dai[1,2], A. Ankala[1,2], M. Hegde[1,2]*. 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) EGL Genetic Diagnostic, Decatur, GA.

Sanger confirmation in accurate variant reporting of next generation sequencing (NGS) data has burdened molecular diagnostics laboratories. In order to minimize unnecessary cost and labor, while not compromising on accuracy, we sought to generate guidelines for careful selection of variant calls that truly need confirmation. At EGL genetic laboratory, 4011 variants were targeted for confirmation via Sanger sequencing in the first year of NGS testing alone. All data was generated using HiSeq 2500 in rapid run mode and variants were called with NextGENe software and annotation processed by EGL bioinformatics pipeline. Out of the 2973 that were single nucleotide variants, 85% (2512) confirmed to be true calls, while 15% (461) were determined to be false calls. We first grouped the NGS calls according to percent reads that constitute the non-reference call. Next, data was analyzed for coverage, NextGENe mutation score, presence of pseudo-exon and the frequency of observation in our internal data as well as population databases. It was quickly evident that calls with 35-75% non-reference reads that fulfil a defined quality criteria for coverage and score, were all confirmed as true heterozygous variants by Sanger sequencing. Similarly, calls >90% non-reference reads and above a definable quality criteria, confirmed as true homozygous variants. Zygosity confirmation is needed for all calls with 75-90% non-reference reads. For calls with <35% non-reference reads, Sanger sequencing is necessary, especially considering possibility of mosaicism. The guidelines presented here include quality criteria for each group of calls stratified for percent reads. Since the implementation of these guidelines we have been able to decrease Sanger confirmation by 74%. In addition, we have been able to test the system with a number of variants and have not found an outlier. In summary, NGS data accuracy is more reliable than originally feared and variant calls pipeline can be set up to minimize Sanger confirmation.

983T

***APAF1* and *CASP9* implicated in recurrent folate-resistant neural tube defects.** *C.J. Spellicy[1], C. Bupp[2], P. Mester[3], T. Reynolds[1], C. Schwartz[1], E. Alexov[4], J. Norris[1], R. Stevenson[1], M.J. Friez[1]*. 1) Greenwood Genetic Center, Greenwood, SC; 2) Spectrum Health, Grand Rapids, MI; 3) University of South Carolina School of Medicine, Columbia, SC; 4) Clemson University, Clemson, SC.

Neural tube defects (NTDs) are among the most serious of birth defects, contributing significantly to prenatal and postnatal morbidity and mortality. Although a number of risk factors have been identified and a risk-lowering intervention (including maternal intake of folic acid in the periconceptional period) is available, NTDs continue to occur at a baseline rate of about 5 cases per 10,000 pregnancies. Approximately 200 genetic mouse models of NTDs have been produced, implicating genes in several pathways as risk factors for NTDs. However, specific Mendelian causes of NTDs in humans have remained elusive. Using whole exome sequencing (WES) and trio analysis, we have identified potentially deleterious variants in key apoptotic genes in two families with recurrent folate-resistant NTDs. Sequencing revealed that affected fetuses in Family 1 have compound heterozygous missense changes in the *APAF1* gene, and affected fetuses in Family 2 have compound heterozygous changes in the *CASP9* gene. The phenotype of the fetuses with NTDs is consistent between these two families, which included open NTDs (lumbar spina bifida in two fetuses in Family 1, and craniorachischisis in two fetuses in Family 2) and unusual rib anomalies (rib bifurcation and irregular placement in the two fetuses in Family 1, and abnormal rib morphology and number in one fetus in Family 2 with craniorachischisis). The open NTD phenotypes observed in Family 1 and Family 2 resemble those observed in knockout mouse models of *Apaf1* and *Casp9*, which typically exhibit exencephaly. Other knockout mouse models of apoptotic pathway genes display phenotypes including rib anomalies and NTDs, indicating the importance of the apoptotic gene family in development of these structures. *In silico* analyses of the missense alterations in these families suggest a potential loss-of-function and/or loss of stability compared to the wild type protein. This is the first report of an association of NTDs in humans with mutations in apoptosis genes.

984F

Monitoring for low-level contamination of extracted DNA. *S. Taylor, R. Tomaszewski, K. Baptista, K. Sprysak, L. Vicen, M. Somerville, S. Hume.* Molecular Diagnostic Laboratory, Genetic Laboratory Services, Alberta Health Services, Edmonton, AB, Canada.

The use of robotic extraction systems has greatly improved sample through-put in molecular diagnostics laboratories. A custom TECAN EVO-CHEMAGEN Robotic DNA Extraction system is used in our laboratory to extract DNA from whole blood. While initial validation of this system suggested cross contamination between specimens was not a concern we decided, as part of our lab quality assurance program, to monitor for such an event during all extractions. As extracted DNA is used for a variety of purposes from simple PCR based assays such as HFE genotyping to NGS assays for low level genetic variants it was important to establish a measure of allowable contamination and a value above which would require corrective action. To this end, dilutions from 0 to 2000X of 15 ng/ul male and female DNAs used routinely as controls in our assays were run in triplicate for a quantitative PCR assay of three loci (HIRA, TBX and COMT) within the DiGeorge Chromosome Region (DGCR) on chromosome 22q11.2. A standard curve of DNA concentration was plotted against the cycle change from the undiluted sample. The amplification efficiency of the 22q11 deletion assay is calculated at 88%. Our findings suggest that a 10 cycle difference, roughly equivalent to a dilution of 300X, would not interfere with interpretation of assay results. Corrective action and a repeat extraction would be initiated if the cycle difference was less. This level of contamination was confirmed to be acceptable on a range of assays, including, quantitative PCR for 22q11 deletions, allelic discrimination for HFE, allele specific PCR for Y chromosome deletions and the AmpFLSTR® Identifiler® PCR kit. We have implemented as part of our standard extraction protocol the inclusion of a 3 ml sample of water in each batch of samples extracted. The water samples are monitored post extraction using the PCR assay for 22q11 deletions. To date amplification has only been detectable in a small portion of extraction contamination controls. We recommend that all molecular diagnostic laboratories monitor for extraction contamination using a quantitative assay and define through assay validation, acceptable contamination limits.

985W

Comparative assessment of Sanger sequencing and targeted next generation sequencing (NGS) for molecular diagnoses of patients with alterations in gonadal axis. *M.N. Nishi[1], L.R. Montenegro[1], M.F.A. Funari[1], M.M. França[1], A.M. Narcizo[2], S.M. Oba-Shinjo[3], A.C. Latronico[1], B.B. Mendonca[1].* 1) Laboratório de Hormônios e Genética Molecular LIM42, HCFMUSP, Sao Paulo, Brazil; 2) SELA, FMUSP, Sao Paulo, Brazil; 3) Laboratório de Investigação em Neurologia, LIM15, FMUSP, Sao Paulo, Brazil.

Introduction: NGS has been widely used for targeted sequencing of candidate genes for diagnosis and research. More than 100 genes are associated with the etiology of isolated hypogonadotropic hypogonadism (IHH) and disorders of sex development (DSD). A comparison of cost and time of execution by Sanger method and targeted NGS can help the researchers choose which methodology best applies to the financial resources and urgency of the study. **Objective:** To compare the cost and execution time between Sanger sequencing and targeted NGS panel for molecular diagnoses of patients with gonadal axis disorders. **Material and Methods:** Genomic DNA was isolated from peripheral blood leukocytes from all patients using salting out method. Sanger sequencing was performed using 60 bp flanking region primers for each exon of candidate genes, 11 for IHH and 11 for 46,XY DSD). A total of 431 patients were evaluated (283 IHH and 148 DSD) by Sanger method. A group of these patients without molecular defects were assessed by NGS (96 IHH and 17 DSD patients). This study used a custom SureSelectXT DNA target enrichment panel (Agilent Technologies Inc) designed to capture 100 known candidate genes involved in gonadal axis. All exons, the 25 base pairs of intronic flanking region and 5' and 3' untranslated region of each gene were sequenced. Sequence capture was performed according to the SureSelectXT Target Enrichment Protocol using Agilent Bravo Automated Liquid Handling Platform. Sequencing was performed on a NextSeq 500 (Illumina, Inc). Allelic variants were considered as causative molecular defect if they had a minor population allele frequency <0,001, segregated with the phenotype in an appropriate inheritance pattern and were found damaging by several prediction sites. **Results:** The cost of NGS, considering only the reagents was around US$370 per patient. To study only the coding region of all genes present on the target panel by Sanger sequencing, it will be necessary around 866 PCRs with an estimated cost of US$14,095 *per* patient. The study by targeted NGS was complete in 3 months and by Sanger in 25 months. 111 molecular defects were identified by Sanger method and 46 by NGS considering both groups. **Conclusion:** The cost and execution time of targeted NGS panel for molecular diagnosis of patients with gonadal axis disorders are better than Sanger method and should be considered for routine diagnosis.

986T

Analysis of exome filtering techniques for the prioritization of clinically relevant genes. *K.D. Kernohan[1], T. Hartley[1], D.A. Dyment[1,2], K.M. Boycott[1,2], Care4Rare Canada Consortium.* 1) Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; 2) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.

Whole exome sequencing is a next-generation sequencing based technique which selectively sequences DNA from the ~1% of the genome that codes for proteins. This cost efficient and effective technology has revolutionized rare disease research as a tool for novel disease gene identification, and has already begun to transition into the clinical realm for diagnostics. However, one of the challenges for clinical exome sequencing is variant interpretation; each exome contains ~20,000 variants, ~300-500 of which are rare (seen in <1% of the population). Clinical labs must interrogate this data to identify the pathogenic variants in genes related to an individual's phenotype, a task which is difficult given the constantly evolving rare disease literature. Efficient exome filtering also allows for control over the identification of incidental findings. We conducted a strategic assessment of 50 exomes with known disease variants using a number of available rare disease datasets as filters: Online Mendelian Inheritance in Man (OMIM), Orphanet, Human Phenotype Ontology (HPO) annotation files and the gene panels curated at Radboud University Medical Center. Each participant was phenotyped using HPO in PhenomeCentral. Projects were then evaluated for the total number of variants which would need to be assessed with the application of each filter as well as their success rate, defined as the presence of the causal variant following filtering. We found that the clinically-curated panels from Radboud produced the fewest variants requiring assessment, however it also resulted in the lowest success rate. OMIM and Orphanet produced the largest variant lists, but with a high success rate. Personalized panels produced using the HPO terms for each participant's PhenomeCentral profile and the HPO annotation files had the same success rate as the OMIM and Orphanet panels, while providing a significantly smaller pool of variants to individually assess. We conclude that HPO term derived gene panels are an efficient and effective way to prioritize and identify clinically relevant exome variants.

987F

Candidate genes evaluated in patients with stomatocytosis and hyperbilirubinemia using whole-exome sequencing. *Y. Kim[1], Y.G. Sung[1], J. Kim[1], K-A. Lee[2].* 1) Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju, Seoul, South Korea, MD; 2) Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea, MD.

Here we reported the case of Korean female with stomatocytosis, hyperbilirubinemia and vitamin D deficiency. A 31 years old female was admitted to the gastroenterology and hepatology department in 16 July for evaluation of hyperbilirubinemia and anemia which had been detected incidentally at 2003. The peripheral blood count revealed that WBC, hemoglobin, MCV, MCHC and platelet count were 3.69 x 10^3/uL, 11.4 g/dL, 113.7 fL, 33.2g/dL and 304 x 10^9/L, respectively. And serum level of total bilirubin, direct bilirubin, calcium, LDH, vitamin D3 and PTH were 5.08 mg/dL, 0.56 mg/dL, 10.6 mg/dL, 238 U/L, <4.20 ng/mL and 181.8 pg/mL, respectively. Peripheral blood smear findings showed stomatocytosis. *UGT1A1* genotype was wild type (*1/*1). To evaluate genetic alternations related patient's phenotype, we performed whole exome sequencing with Illumina Hiseq2500 and Agilent Sureselect exome kit. The sequence reads were aligned to the reference genome (NCBI b37) using the software BWA and subsequent identification of genetic variants was done with the Genome Analysis Toolkit (GATK version 2.8.1). Annotation was performed using SnpEff (version 3.3h). Alignments are visually verified with the Integrative Genomics Viewer (IGV) v2.3. Database of dbSNP, ExAC, 1000 genomes, KRGDB, KOMIM, NCBI, ClinVar and literate review were used for mutation interpretation. In silico analysis were performed using Align-GVGD, Shift, Polyphen-2, splicing prediction programs. After filtering according to analysis steps, we obtained the lists of candidate gene and SNPs. We will perform confirmatory study using sequencing with patients and unaffected family DNA samples.

988W

Sample collection for saliva DNA may influence results from next generation sequencing for microbiome research. *M.G. Kibriya, F. Jasmine, J. Shinkle, M. Sabarinathan, H. Ahsan.* Public Health Sciences, University of Chicago, Chicago, IL.

   Background: Observations from microbiome studies in both healthy and diseased individuals are considered to be affected by factors including sample collection, handling, library preparation and next generation sequencing (NGS) platform. In this study, we evaluate the role of methods of saliva collection and processing for DNA extraction on microbiome NGS results. **Material and Methods:** This pilot study was designed to address situations that are commonly encountered in large-scale epidemiological studies: (a) Duration between saliva collection and processing, i.e., saliva collected at clinic and processed right away versus saliva collected at home and transported or shipped to the laboratory for processing; (b) Time of the day for saliva collection, i.e., first morning sample in fasting state versus randomly collected samples during the day; (c) Saliva collection with or without preservative. Saliva samples were collected from four individuals. Each individual gave saliva (fasting and random) in different tubes (Plain, Oragene, IsoHelix). Samples were processed with different gap between collection and DNA extraction. We had a total of 44 DNA samples from 4 individuals, processed for metagenomics using Illumina Nextera XT and KAPA Biosystems kits for 16S rRNA gene V3 and V4 region sequencing on MiSeq. **Result:** Each sample had a median 250165 reads passing the filter and only ~250 reads per samples were unclassified. For analysis, we focused on the bacterias (n=26) at genus level which had at least 2% reads in at least one of the 44 DNA samples. The factors that influenced the relative abundance of a specific genus included (a) person, (b) collection time within a day, (c) collection tube, as well as, (d) different days of collection. No statistically significant difference was observed arising from (a) immediate DNA extraction vs. extraction of DNA after storage at -86C; (b) extraction of DNA with or without bead beating; and (c) gap between sample collection and heat treatment - between 6 hrs and 7 days. **Conclusion:** Relative microbial abundance in saliva is different from person-to-person. However, there is day-to-day variation in relative abundance within an individual. For a particular study, uniformity in saliva sample collection protocol (in terms of collection time within a day, collection tube with preservative) is essential to compare microbial abundance between groups.

989T

Integration of copy number variation analysis with custom gene panel next-generation sequencing diagnostics. *J.M. Ellingford[1,2], C. Campbell[1], S. Barton[1], S. Bhaskar[1], S. Gupta[3], S.C. Ramsden[1], B. Panda[3], G.C.M. Black[1,2].* 1) Manchester Centre for Genomic Medicine, Central Manchester Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, St Mary's Hospital, Manchester, United Kingdom; 2) Institute of Human Development, University of Manchester, United Kingdom; 3) Ganit Laboratories, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India.

   Introduction: Copy number variants (CNVs) are increasingly identified as a cause of human genetic disease. Next-generation sequencing (NGS) techniques are a routine diagnostic test for many such disorders, but the interrogation of CNVs in targeted NGS data is limited in comparison to whole genome sequencing (WGS) techniques. Consequently, CNV detection is seldom undertaken in a clinical context. **Methods:** We benchmarked the accuracy of CNV detection in targeted NGS datasets through (i) comparison to events detected by WGS for 25 individuals, and (ii) a validated simulation methodology. Using quality assurance metrics determined through these assessments, we applied CNV detection for 133 individuals without molecular diagnoses through clinical NGS testing. High coverage targeted NGS data was generated using standard diagnostic procedures and CNV detection was performed using ExomeDepth v.1.1.6. **Results:** In comparison to WGS data, we determined a sensitivity and specificity for CNV detection of 100% and 99.8%, respectively. We found 97% of simulated events (*n*=2749), and identified that the extent of sequencing coverage, the inter- and intra-sample variability in the depth of sequencing coverage, and the composition of analysis regions are all important determinants of successful CNV surveillance. Clinical integration of CNV detection identified 17 CNVs contributing to the onset of a Mendelian disorder, and ruled out several clinical diagnoses. These findings were verified through additional molecular techniques. **Conclusions:** Our findings support the integration of CNV detection with routine targeted NGS diagnostic services for Mendelian disorders. Implementation of this strategy has transferred this from the research environment, has increased diagnostic yields and is enhancing clinical care.

990F

Correlations between predictive secondary genomic variants, family history, and clinical findings in families of children undergoing whole genome sequencing. *M.S. Meyn[1,2,3,4,5], N. Monfared[2], C.R. Marshall[6,7], D.J. Stavropoulos[7,8], R. Basran[7], D. Merico[4,6], R.H. Hayeems[2,9,10], M. Szego[6,10,11], R. Zlotnik-Shaul[5,10,14], C. Shuman[3,4], T. Nalpathamkalam[4,6], G. Pellecchia[4,6], B. Thiruvahindrapuram[4,6], M. Girdea[4,12], M. Bruno[4,12], R.D. Cohn[1,2,3,4,5], S.W. Scherer[2,3,4,6,13], P.N. Ray[2,3,4,6,7], S.C. Bowdin[1,2,5].* 1) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 4) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 9) Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada; 10) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 11) Centre for Clinical Ethics, St. Joseph's Health Centre, Toronto, ON, Canada; 12) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 13) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 14) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

The SickKids Genome Clinic is a multidisciplinary research project that conducts clinical whole genome sequencing (WGS) for children who are undergoing genetic evaluations. With parents' permission, we systematically search children's genomes for diagnostic variants that explain the patient's known phenotype and predictive secondary variants (PSVs) associated with occult or future disease. PSVs for childhood onset disorders are always returned. ~65% of parents also request return of their child's adult onset PSVs.   For our first 100 patients bioinformatics analysis of WGS data for candidate PSVs affecting 2800+ disease genes yielded 154 CNVs and 571 SVs that overlapped disease genes as well as 2957 SNVs in 1132 disease genes (~30 SNVs/genome). Two thirds of the SNVs were listed in HGMD, but manual assessment eliminated >90% of these PSVs as well as all CNVs and SVs. Using 2015 ACMG variant classification guidelines we found 24 pathogenic/likely pathogenic SNVs. An additional 12 SNVs were deemed returnable PSVs based on more limited evidence of pathogenicity. All 37 PSVs were inherited and 25% predicted adult onset disease. The majority were rare (MAF<0.0001) variants in genes that cause dominant Mendelian disorders. The rest included variants linked to attenuated Mendelian phenotypes, common disease risk variants, an X-linked variant and a protective variant.   30/100 children had at least one PSV, including 4 children with 2 PSVs and one child with 3. For children whose genomes had no reportable PSVs structured family histories obtained at the time of result disclosure were generally negative. Of note, pedigrees collected by the referring clinician rarely identified disease states predicted by a secondary PSV found in the child. This is likely due to inadequate reporting of family history, incomplete penetrance, and/or undiagnosed disease. *E.g.,* for some families reassessing a negative family history following disclosure of a secondary PSV uncovered family members who had the predicted phenotype, while in other families we found children and parents who carried a pathogenic/likely pathogenic PSV but lacked the predicted phenotype.   Comprehensive genomic searching can identify PSVs in 25-30% of children and most PSVs are rare familial mutations associated with disorders that occur in childhood. While WGS yields more false positives, it appears more sensitive than family history for identifying familial genetic disorders and can uncover occult disease states.

991W

Phenotype-based matching in mendeliangenomics.org. *J. Krier[1], A. Hamosh[2], F. Schiettecatte[3], N. Sobreira[2].* 1) Brigham and Women's Hospital, Boston, MA; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) FS Consulting, LLC; Marblehead, MA.

Introduction. Matchmaking initiatives such as GeneMatcher have demonstrated the utility of gene-based matching for identifying unrelated individuals with similar phenotypes and pathogenic variants in the same gene. Phenotype-based matching has been attempted less widely. **Methods**. As part of the Baylor-Hopkins Center for Mendelian Genomics, users submit cases to mendeliangenomics.org using phenotypic terms from the PhenoDB ontology, enabling the use of semantic-similarity based phenotype matching algorithms. We compared the following methodologies: SimUI, Distance, Wang, Jacquard, Resnick, SimGIC, PhenoDigm. For information content calculations, corpora based on phenotypic features in OMIM and mendeliangenomics.org were each tested. We used pair-based testing in which one case of a known phenotype is used as a query case, and another case is "planted" in the test set. We tested the matching algorithms on a) 44 published cases with detailed phenotypic descriptions of four known syndromes and b) 1,152 unrelated probands with 37 different OMIM phenotypes in the mendeliangenomics.org database who had been annotated with at least 5 phenotypic terms. To simulate a real-life matching query, we evaluated the ability of the algorithms to identify all cases of the same syndrome as a query case within the top 5th percentile of the complete case set based on phenotypic similarity. **Results.** We validated the algorithms by demonstrating strong matching performance using the 44 case testing set. Using data from mendeliangenomics.org, the best performing algorithm (SimGIC) successfully matched the "planted" case to the query case of the same syndrome within the top 5 phenotypic matches in 56% of matching tests. Across six selected syndromes, the SimGIC algorithm successfully identified 80% of cases sharing the same diagnosis within the top 5th percentile of the database, and above 95% for 3/6 phenotypes. **Conclusions.** Applying semantic-similarity based matching algorithms to the mendeliangenomics.org database resulted in mixed success at identifying "planted" cases of the same syndrome, but shows promise in stratifying cases of the same syndrome within the top 5th percentile large data set based on phenotypic similarity. We will next explore if the combination of phenotype-based matching and pathway-based matching offers the potential to identify novel Mendelian genes in the same pathway that cause similar phenotypes.

992T

Expanding the clinical sensitivity of shole genome sequencing – Validation of copy number variation. *R.J. Taft[1], A. Ajay[1], A. Gross[1], D. Perry[1], J. Avecilla[1], A. Chawla[1], A. Khouzam[1], E. Thorpe[1], S. Chowdury[1,2], S. Kruglyak[1], S. Tanner[1], M. Eberle[1], D. Bentley[3].* 1) Illumina Inc, San Diego, CA, USA; 2) Rady Children's Institute for Genomic Medicine, San Diego, CA USA; 3) Illumina Cambridge Ltd, Cambridge, UK.

Single nucleotide and small indel variant calling and interpretation are routine components of clinical exome and whole genome sequencing (WGS) pipelines. Traditionally, large copy number variants (CNVs) have been assessed by orthogonal technologies such as chromosomal microarrays. Here we report on the integration of large and small variant calls on a single WGS platform, and the validated deployment of copy number variants greater than 10kb as a component of the Illumina Clinical Services Laboratory's rare and undiagnosed disease WGS test offering. Using a set of 20 Coriell samples with well-characterized pathogenic CNVs, spanning from less than 50kb to trisomies, we performed a head-to-head comparison of large variants called from WGS using the Canvas algorithm and those detected by an independent clinical microarray laboratory using the Infinium CytoSNP-850k microarray. We found greater than 80% sensitivity of Canvas-called events, which exceeded that of microarrays. To assess specificity, we compared WGS Canvas calls against PacBio and BioNano data sets derived from CEPH NA12878 genome, and found a maximum false positive rate of ~20-40%. To enable rapid triaging of potential candidate CNVs, we developed a set of family-based analysis algorithms that, in the context of a proband in a family trio, typically delivers 8-10 CNVs that require manual interpretation. Up to half of these variants may be ruled out by filtering against common CNVs identified in public data sets, or an in-house database of more than 2,000 genomes, suggesting that the CNV interpretation burden from WGS data may be low. Initial CNV deployment has increased diagnostic yield - resolving both X-linked and autosomal CNVs, and complex scenarios including conditions caused by combinations of single nucleotide and CNV events. Ongoing efforts to improve our clinical CNV calling indicate that variants as small as 50bp – 10kb are tractable, suggesting that WGS can provide the complete range of variants from a single assay.

993F

The variable phenotype of heterozygous form of β-thalassemia and α-triplication. *F. Moosavi[1], A. Moosavi[1], A. Amirian[1], S. Zeinali[1,2], M.S. Fallah[2], M. Karimipoor[1].* 1) Pasteur institute of Iran, Tehran, Iran; 2) Kawsar Human Genetics Research Center, Tehran, Iran.

Objectives: Different types of thalassemia and hemoglobinopathies are common in Iran. Therefore, the combination of two or more globin chain aberrations in families is not uncommon. The clinical symptoms of β-thalassemia are modified by some genetic determinants like the number of α-globin genes. **Methods:** In this report, the prevalence of α-globin gene triplication was investigated in 550 Iranian β-thalassemia carriers. Multiplex PCR and MLPA methods were used in this study. In addition two families were referred to our center for further study of their children with intermediate forms of β-thalassemia. **Results:** Fourteen cases (2.54%) were heterozygote for α-triplication. Investigation of mentioned families revealed the combination of heterozygous form of β-thalassemia and α-triplication. **Discussion:** The results emphasizes that uncertainty still exist about the phenotypic prediction of excess of α-globin genes in combination with β- thalassemia heterozygote carriers and these issues should be concerned in genetic counseling. **Keywords:** α-triplication; α-quadruplication; β-thalassemia; thalassemia intermedia; Alpha-hemoglobin-stabilizing protein (AHSP).

994W

Genetic variations of *AKAP4* gene in infertile patients with immotile short-tail sperm defect. *E. Zare Mehrjardi[1], M. Sabbaghian[2], A. Mohseni Meybodi[3], S.M. Seifati[1], M.A. Sadighi Gilani[2], J. Hoseini[2].* 1) Islamic Azad University of Ashkezar branch, Yazd, Yazd, Iran; 2) Department of Andrology at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 3) Department of Genetics at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.

Ehsan Zare Mehrjardi[1], Marjan Sabbaghian[2], Anahita Mohseni Meybodi[3], Seyed Morteza Seifati[1], Mohamad Ali Sadighi Gilani[2], Jalil Hoseini[2] 1. Medical Biotechnology Research Center, Ashkezar Branch, Islamic Azad University, Ashkezar, Yazd, Iran 2. Department of Andrology at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran. 3. Department of Genetics at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran. **Introduction**: The immotile short tail sperm (ISTS) defect is an autosomal recessive disease that causes male infertility. A Kinase Anchoring Protein 4 (AKAP4) in the fibrous sheath (FS) of sperm flagella provides scaffold for the correct assembly of FS. The main goal of this research was to assess *AKAP4* gene variation in ISTS patients. **Material and Methods**: In this study, which was performed from 2010 to 2013, 32 infertile men with ISTS defect (>80% short tail sperm cells in at least two spermograms) and 50 normozoospermic men as control were enrolled. Semen analysis was performed for each specimen. After DNA extraction from peripheral blood samples, PCR-Sequencing was done for two regions of exon 5 in *AKAP4* gene. **Results**: Semen analysis showed significant differences in sperm concentration, normal morphology, progressive motility, and short tail sperm count between patients and controls. Sequence analysis did not show any single nucleotide polymorphisms (SNPs) in the target regions. But, a deletion related to the potential AKAP4/AKAP3 binding region was found in one patient. GAPDH was applied as internal control. Four primers were designed (with 700 base pair distance) that did not show any PCR products. Karyotype showed normal status revealed that the banding pattern of chromosomes were not affected by this deletion. **Conclusions**: Our data did not reveal any SNP in exon 5 of the gene, but further studies with more sample size and determination of levels of mRNA and protein from the *AKAP4* gene are suggested in order to elucidate final conclusion. *Key words*: Male infertility, Short tail sperm, *AKAP4* gene.

995T

The validation of pharmacogenetics in the identification of patients with Fabry disease for treatment with migalastat. *E.R. Benjamin[1], C. Della Valle[1], X. Wu[1], E. Katz[1], K.J. Valenzano[1], D.G. Bichet[2], D.P. Germain[3], R. Giugliani[4], D. Hughes[5], R. Schiffmann[6], W. Wilcox[7], J. Yu[1], J. Kirk[1], J. Barth[1], J. Castelli[1].* 1) Amicus Therapeutics,, Cranbury, NJ, USA; 2) Hôpital du Sacré-Coeur, University of Montreal, Montreal, Quebec, Canada; 3) Division of Medical Genetics, University of Versailles–St. Quentin en Yvelines (UVSQ), Montigny, France; 4) HCPA/UFRGS Porto Alegre, Porto Alegre, Brazil; 5) Royal Free Hospital, University College London, London, UK; 6) Baylor Research Institute, Dallas, TX, USA; 7) Department of Human Genetics, Emory University, Atlanta, GA, USA.

Fabry disease is an X-linked lysosomal storage disorder caused by *GLA* mutations, resulting in deficient lysosomal α-galactosidase A (α-Gal A) activity and progressive cellular accumulation of globotriaosylceramide (GL-3). Migalastat is a pharmacological chaperone that stabilizes specific mutant forms of α-Gal A (termed amenable), leading to increased physical stability, lysosomal trafficking, and intracellular activity of the enzyme. Each of 600 Fabry disease-causing mutations were expressed in HEK-293 (HEK) cells, and increases in α-Gal A activity in response to migalastat were measured using a Good Laboratory Practice (GLP)-validated *in vitro* assay (GLP HEK/migalastat amenability assay). In total, 268 migalastat-amenable mutant forms of α-Gal A were identified, defined by a relative increase of ≥1.20-fold and an absolute increase of ≥3.0% wild-type α-Gal A activity in the presence of 10 μM migalastat. Clinical validation was assessed using pharmacodynamic responses in phase 2/3 studies (73 unique amenable and nonamenable mutations in 160 male and female patients with Fabry disease). Comparison of mutant α-Gal A responses in the GLP HEK/migalastat amenability assay to responses in white blood cells from male patients in phase 2 and 3 studies to treatment with migalastat exhibited high sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively; ≥0.875). GLP HEK/migalastat amenability assay results were predictive of decreases in kidney GL-3 in males and plasma globotriaosylsphingosine in both males and females (sensitivity: 1.0 and 0.93, respectively; specificity: 1.0 and 0.69; PPV: 1.0 and 0.84; NPV: 1.0 and 0.85). The clinical study subset of amenable mutations (n=51) were representative of all 268 amenable mutations identified by the GLP HEK/migalastat amenability assay. The GLP HEK/migalastat amenability assay is a clinically validated method of identifying male and female patients with Fabry disease who are candidates for treatment with migalastat.

996F

"SMRTer Confirmation": Scalable clinical read-through variant confirmation using the Pacific Biosciences SMRT® Sequencing Platform. *S. McCalmon, K. Konvicka, N. Reddy, J. Whittaker, C. Kautzer, J. Risinger, J. Sorenson, R. Lewis, M. Kennemer, E. Olivares, A. Shinohara, I. Melnick, A. Rosendorff.* Invitae Corporation, San Francisco, CA.

Next-generation sequencing (NGS) has significantly improved the cost and turnaround time for diagnostic genetic tests. ACMG recommends variant confirmation by an orthogonal method, unless sufficiently high sensitivity and specificity can be demonstrated using NGS alone. Most NGS laboratories make extensive use of Sanger sequencing for secondary confirmation of single nucleotide variants (SNVs) and indels, representing a large fraction of the cost and time required to deliver high quality genetic testing data to clinicians and patients. Despite its established data quality, Sanger is not a high-throughput method by today's standards from either an assay or analysis standpoint as it can involve manual review of Sanger traces and is not amenable to multiplexing. Toward a scalable solution for confirmation, Invitae has developed a fully automated and LIMS-tracked assay and informatics pipeline that utilizes the Pacific Biosciences SMRT® sequencing platform. Invitae's pipeline generates PCR amplicons that encompass the variant(s) of interest, which are converted to closed DNA structures (SMRTbells) and sequenced in pools of 96 per SMRTcell. Each amplicon is appended with a 16nt barcode that encodes the patient and variant IDs. Per-sample demultiplexing, alignment, variant-calling, and confirmation resolution are handled via an automated pipeline. The confirmation process was validated by analyzing 243 clinical SNVs and indels in parallel with the gold standard Sanger sequencing method. Amplicons were sequenced and analyzed in technical replicates to demonstrate reproducibility. In this study, the PacBio-based confirmation pipeline demonstrated high reproducibility (97.5%), and outperformed Sanger in the fraction of primary NGS variants confirmed (PacBio = 93.4% and 94.7% confirmed across two replicates, Sanger = 84.8%) while having 100% concordance of confirmation status among overlapping confirmation calls.

997W

Mutation spectrum of *ACADS* gene in Korean patients with short-chain acyl-coenzyme A dehydrogenase deficiency. *C.K. Cheon[1], Y.M. Kim[1], J.M. Ko[2].* 1) Pediatrics, Pusan National University, Yangsan, Yangsan, South Korea; 2) Pediatrics, Seoul National University Children's Hospital, Seoul, Korea.

Purpose: Short-chain acyl-CoA dehydrogenase deficiency (SCADD; OMIM # 201470) is an autosomal recessive inborn error of mitochondrial fatty acid β-oxidation, presenting with a variety of clinical signs and symptoms. We sought to investigate the mutation spectrum of *ACADS* gene and associated clinical manifestations in Korean patients with confirmed SCADD. **Methods:** The study included 9 patients with SCADD from 7 unrelated families as diagnosed by biochemical profile and DNA testing. Clinical features, biochemical data, growth, and neurodevelopmental state were reviewed retrospectively. **Results:** Seven patients were found during newborn screening using tandem mass spectrometry, and 2 were diagnosed by family screening after their siblings were diagnosed. Their mean age at diagnosis was 14.1 ± 31.9 months. During follow-up ranging from 9 months to 4.5 years, no hypoglycemic event was observed, and the development and growth of the patients was normal, except in 2 siblings showing mild hypotonia and delayed gross motor development. In addition, one girl showed cyclic vomiting until the age of 2 years. We analyzed the *ACADS* genotype in all patients, identifying 7 different mutations. Of these, p.E344G was the most frequent mutation with an allele frequency of 42.8%, followed by p.P55L with 21.4%. p.G108D, and 4 novel mutations were identified: p.L93I, p.E228K, p.P377L, and p.R386H. **Conclusions:** While the *ACADS* genotype was heterogeneous, Korean patients with SCADD detected during newborn screening generally showed mild clinical manifestation without severe metabolic crisis. These data contribute to a better understanding of the distinct molecular genetic characteristics and clinical manifestations of SCADD in Korean patients.

998T

Will the real chromosome anomaly please stand up. *L. Richards, M. Chung, J. Castelli, J. Kogan, D. Rita.* ACL Laboratories, 5400 Pearl Street, Rosemont, IL 60018.

The implementation of whole genome SNP microarray copy number analysis along with fluorescence in situ hybridization (FISH) techniques in the clinical laboratory setting has had an impact on the quality and accuracy of diagnoses provided. We present three unique cases in which traditional chromosome analysis detected derivative or marker chromosomes that were later shown through the use of microarray and FISH techniques to be quite different than originally reported. We introduce the first case, where chromosome analysis was initially evaluated to be a derivative chromosome 15 involving an unbalanced translocation between chromosomes X and 15. However, microarray analysis revealed a duplication of chromosome 4, which changed the original analysis to der(15)t(4;15). In the second case chromosome analysis initially revealed an apparent derivative chromosome 2 involving an unbalanced translocation between chromosomes 2 and 14. Microarray was later performed and confirmed a derivative chromosome 2 bearing two separate interstitial deletions of the short arm of chromosome 2, but no evidence of an unbalanced translocation. The third case involved mosaicism with three different cell lines consisting of: monosomy X, monosomy X with an extra marker chromosome of unknown origin, and two apparent centric fissions of chromosome X. Microarray and multiple FISH analyses were performed that identified the marker as material from chromosome Y, and the presumed X fissions were actually one normal Y chromosome and one inverted duplicated Y chromosome. Knowing the exact origins of these chromosomal abnormalities is essential for accurate prognosis for patients and for genetic counseling. These cases illustrate that there is no simple answer in terms of which technique will most accurately provide this knowledge. G banded chromosome analysis alone is subject to interpretation error. These cases highlight the need for laboratories to employ all available techniques to attain the highest quality results for their patients. .

999F

Autozygosity mapping using SNP based chromosomal microarray: Opportunity to increase the clinical diagnostic yield in populations with high consanguinity. *S. Sampath[1], A.W. El-Hattab[2].* 1) PreventionGenetics, Marshfield, WI; 2) Pediatrics Department, Tawam Hospital, Al-Ain, United Arab Emirates.

Chromosomal microarray (CMA) test has been used as first-tier diagnostic test for the clinical diagnosis of genetic disorders in individuals with unexplained developmental delay/intellectual disability, autism spectrum disorder, and multiple congenital abnormalities, with an overall yield of 15-20%. Because of its enhanced clinical utility and impact on medical management, copy number detection using CMA test has largely replaced conventional cytogenetic tests such as chromosome and FISH analysis, although some of its limitations over these traditional tools have been exemplified in the last few years. Moreover, the use of single nucleotide polymorphism (SNP) based CMA test using aCGH + SNP microarrays conferred additional clinical utility in a) detection of allele specific copy-number, b) detection of polyploidy, and c) detection of excess homozygosity on single and multiple chromosomes. The presence of excess homozygosity in the form of uninterrupted regions/blocks of homozygous SNP alleles, may indicate, depending its total size, extent and its distribution on chromosomes, ancestral relationship, uniparental disomy (UPD) or parental consanguinity. Although the presence of excess homozygosity is by itself not diagnostic of any specific disorder, it increases the likelihood of recessive traits. Autozygosity mapping following CMA has been successfully used to identify genes underlying numerous recessive disorders in consanguineous families. Despite the success of CMA-based autozygosity mapping, its overall burden in out-bred populations significantly limits its use in a clinic setting. In this study, we describe the use of autozygosity mapping in 227 patients referred for CMA from a population with high levels of consanguinity, and demonstrate its clinical utility and overall increase in the diagnostic yield. We show that the diagnostic yield of CMA increased by at least 10% by using freely-available online tools for the detection of candidate genes and associated conditions mapping to regions of homozygosity. We also demonstrate the successful implementation of a diagnostic workflow for the routine analysis of AOH regions in a clinical diagnostic lab. Finally, using specific examples, we identify potential limitations of our approach, and make recommendations to increase the diagnostic yield by enhancing tools that automate phenotype-genotype based prioritization of gene candidates.

1000W

Validation of the OneOme RightMed™ pharmacogenomic test using the IntelliQube® to demonstrate an accurate method for pharmacogenomic genotyping. *R. Higgins, K. Bauer, B. Ramey Hartung, Ph. D., D. Bailey, Ph.D., J. Odegaard, M.D., Ph.D.* OneOme, Minneapolis, MN.

Imprecise medication can lead to delay of treatment, re-hospitalization, adverse events and increased mortality. Adverse drug reactions are the fourth leading cause of death in the United States leading to 2.2 million hospitalizations and $136 billion in costs annually.[1] Genetic factors have been shown to account for up to 95% of drug response variability.[2] Pharmacogenomics (PGx) can identify which inherited genetic variations a person has to predict how an individual will or will not metabolize a drug, ultimately guiding more effective treatment. We have developed the OneOme RightMed™ pharmacogenomic test that combines a patient's genetic profile, prescription history, and current medications with a proprietary algorithm-based platform co-developed and exclusively licensed from Mayo Clinic. The RightMed PGx test panel includes 22 genes and 90 alleles that cover more than 300 common drugs. The genotype analysis is performed using probed-based quantitative PCR assays. Here we review the results of 75 SNP genotyping assays and 4 copy number variation (CNV) assays as part of the RightMed analytical validation. The validation study included a comparison of genotype results from a collection of samples from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research in addition to 12 donor samples which were sent for Sanger sequencing confirmation. Sanger sequencing was completed at an independent laboratory, Eurofins Clinical Molecular Testing Services in Louisville, Kentucky. Results for the sequenced donor samples and reported genotypes for the Coriell samples were compared to the genotype calls generated by several intra- and inter-run precision studies performed at OneOme to assess concordance. The results of this validation study suggest that the methods for running the OneOme RightMed pharmacogenomic test panel produces both accurate and reproducible genotyping results. These findings provide confidence in the analytical validity of the RightMed test and may subsequently further the widespread adoption of the test panel to guide better prescribing decisions based on evolving pharmacogenomics evidence. 1) Johnson JA, Bootman JL. Drug-related morbidity and mortality. A cost-of-illness model. Arch Intern Med 1995;155(18):1949–1956. 2) Kalow W, et al. Pharmacogenetics 1998;8:283-289.

1001T

The Iowa Institute of Human Genetics Drug Metabolism Test: A pharmacogenomics platform to predict adverse drug response. *A.E. Kwitek[1], N.C. Smith[2], T.B. Bair[1], C.A. Campbell[1], M.S. Chimenti[1], M.J. Kimble[1], K.L. Knudtson[1], D.L. Kolbe[1], M.A. Mansilla[1], S.O. Mason[1], C.J. Nishimura[1], M. Sorenson[3], R.J. Smith[1].* 1) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 2) Institute for Clinical & Translational Sciences, University of Iowa, Iowa City, IA; 3) University of Iowa Hospitals and Clinics, University of Iowa, Iowa City, IA.

Approximately 5% of all hospital admissions are due to adverse drug responses (ADRs), resulting in over 100,000 deaths annually and $177.4 billion in related healthcare costs. Some of these ADRs are due to known inherited genetic variation. A goal in the Iowa Institute of Human Genetics (IIHG) is to identify patients with genetic risk for adverse drug responses, prior to being given the drug. Toward this goal, we developed a targeted genome enrichment and massive parallel sequencing (TGE+MPS) panel of 280 genes involving drug metabolism, transport, and action. The Drug Metabolism Test (DMT) is clinically available test through the IIHG CLIA certified, CAP accredited Clinical Diagnostics Division. Initial clinical implementation focused on reporting variants in two genes (*CYP2C19* and *CYP2D6*) affecting metabolism of clopidogrel and opioids (http://www.medicine.uiowa.edu/humangenetics/pharmacogenomic/provider). To support the clinical utility of the DMT, we performed a cost analysis of patients going to the Emergency Department (ED) at the University of Iowa Hospitals and Clinics (UIHC) for an adverse drug event. The aim of the study was to determine if preemptive testing for genes causing ADRs could have predicted patients at risk for an ADR. All patients visiting the UIHC ED for the past three years were identified and screened for ICD9 codes indicating an adverse drug response. We identified 1019 visits to the ED by 951 patients having an ADR, with related costs of nearly $52 million. Adverse response to chemotherapy drugs corresponded to over 20% of the ED visits, incurring nearly $20 million in total costs. Currently we are validating and implementing a test for inherited risk of chemotoxicity in *DPYD*, *G6PD*, *TPMT*, and *UGT1A1*. This test will be offered to the patients suffering an ADR to common chemotherapy agents (fluorouracil, capecitabine, tegafur, mercaptopurine, azathioprine, thioguanine, nilotinib, and rasburicase), to determine if their ADR could have been predicted. The DMT is a tool that can assist healthcare providers in minimizing adverse drug reactions, ideally prior to the start of therapy, and strongly complements established cancer mutation profiling.

1002F

PGx1: An assay for actionable genes in pharmacogenetics. *T. Wiltshire, O. Dong, O. Suzuki.* School of Pharmacy, University of North Carolina, Chapel Hill, NC.

Currently there are over 120 drugs that have genetic information in their package labeling and deemed as being actionable by FDA guidelines, some with "black box" warnings where an action is assigned. Of these drugs, 23 fall within the top 200 most commonly prescribed drugs. The Clinical Pharmacogenetics Implementation Consortium (CPIC) and the FDA provide guidelines that enable the translation of genetic laboratory test results into actionable prescribing decisions that promote the safe and effective use of specific drugs. It is estimated that during a person's lifetime 96% of consumers will be prescribed a drug that has actionable PGx information, **yet few of these tests are currently used in any clinical settings**. There have been a number of major barriers to implementation of pharmacogenetic diagnostics, including lack of knowledge by clinicians, but **a major factor is the availability and cost of tests, reimbursement, and delay in return of results.** Here, we address how to capture PGx data efficiently, in a cost-effective and broadly applicable manner. We have designed an assay set that covers the pharmacogenetic "actionable genome", that is genes that have guidance (CPIC) for the use of the genetic information. We have used a molecular inversion probe (MIP) platform to develop a custom pull-down sequencing assay (PGx1). Designs were performed for twenty-one genes using MIPgen resulting in 1800 probes, with each probe optimized to specifically capture 112 bp of sequence. MIP sets were tested and re-designed or re-balanced for optimal performance. The resultant set covers all coding regions, exon/intron boundaries, and 1 kb of upstream sequence (amounting to 120 kb of sequence). PGx1 tests in a 48 DNA sample multiplexed MiSeq sequence run show that the assay provides greater than 35X average coverage. As a clinical diagnostic tool however, there are only specific polymorphisms that have direct guidance; sequencing depth was optimized for these polymorphisms. 42 polymorphisms have been validated in a CLIA/CAP environment such that the genotypes can be directly reported as a clinical result. These genes/polymorphisms cover pharmacogenetic impact across a number of different disease/drug indications, eg. cardiovascular, pain, psychiatric. We have used this assay as an educational tool with PharmD students, and in high-risk populations (cath lab patients) to explore pharmacoeconomic analysis.

1003W

Retrospective comparison of traditional single gene testing and massively parallel sequencing-based gene panels in the diagnosis of pulmonary arterial hypertension. *K. Sumner[1], D.H. Best[1,2,3].* 1) Inst. for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, UT; 2) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 3) Dept. of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

Background: Pulmonary arterial hypertension (PAH) is a rare, but lethal disorder most commonly caused by autosomal dominant (with reduced penetrance) mutations in the *BMPR2* gene. PAH can also result from mutations in the *ACVRL1*, *ENG*, *CAV1*, *KCNK3*, and *EIF2AK4* genes. **Method:** A retrospective analysis was performed to determine the mutation detection rate in (n=68) PAH patients submitted for diagnostic testing of the *BMPR2* gene using traditional testing modalities (i.e. Sanger sequencing, MLPA). Similarly, the mutation detection rate was determined for (n=41) PAH patients referred for a massively parallel sequencing (MPS)-based PAH gene panel containing multiple PAH related genes combined with deletion/duplication analysis by comparative genomic hybridization array (aCGH). **Results:** Sanger sequencing of the *BMPR2* gene detected a clinically significant sequence variant in 10 out of 68 patients (14.7%).Testing by the MPS-based PAH gene panel identified a clinically significant variant in 14 out of 41 patents (34.1%). The PAH panel detected eight mutations in the *BMPR2* gene and six in the other PAH genes tested *ENG* (n=1), *KCNK3* (n=1), *ACVRL1* (n=2), *EIF2AK4* (n=2). **Conclusion:** Diagnostic testing in PAH patients using traditional testing modalities is ultimately more time-consuming and costly than those utilizing MPS.

1004T

Transcriptomic profiles in peripheral blood between women with unexplained recurrent implantation failure and recurrent miscarriage and the correlation with endometrium. *J. Huang[1], H. Qin[2], X. Shi[1], W. Cheung[1], C. Wang[1,3], T. Chan[2], T. Li[1].* 1) Department of Obstetrics and Gynaecology, the Chinese University of Hong Kong, Hong Kong SAR, China; 2) School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China; 3) Li Ka Shing Institute of Health Sciences, The Chineses University of Hong Kong, Hong Kong SAR, China.

The gene expression profiles are different in the endometrium of RIF, RM and fertile women during the window of implantation. Women with RIF and RM have both been reported to have altered expression of endometrial receptivity markers. But it is not known whether any difference in transcriptome in peripheral blood during the window of implantation between women with recurrent implantation failures (RIF) and unexplained recurrent miscarriages (RM). This is an observational prospective study. In total 9 subjects were recruited, 3 RIF, 3 RM, and 3 controls. Paired samples (endometrium and peripheral blood) simultaneously collected from the same subjects precisely timed on the 7th days after luteal hormone surge (LH+7). This study carried out in a reproductive medicine center of a teaching Hospital. The paired samples were extracted for total RNA. The transcriptome was determined by RNA-Seq. Differentially expressed genes were validated by quantitative real time PCR (qPCR). Among nearly 19,000 genes had sequencing readings, there were 1595 genes correlation coefficients more than 0.5. The correlation coefficients of the 9 paired samples were all over 0.64, and it seems that the correlation between blood and endometrium were higher in RIF than those in RM, with $p < 0.05$. However, among the three studied groups, RM and RIF showed distinct different gene expression patterns in endometrium, but in blood the two groups could not be separated. So far, it is still infeasible to use blood transcriptome to investigate the endometrial gene expressions.

1005F

Clinical versus research analyses of clinical genomic data: Two products of one test. *D.R. Adams[1,2], B. Pusey[2], C. Lau[2], G. Yu[2], T.C. Markello[2], C. Tifft[1,2], C. Toro[2], M. Malicdan[1,2], W.A. Gahl[1,2].* 1) NHGRI, NIH, Bethesda, MD; 2) Undiagnosed Diseases Program and Undiagnosed Diseases Network, Common Fund, NIH, Bethesda, MD.

Clinical exome sequencing has been established as a powerful and productive tool in genomic diagnostics. However, a substantial number of exome studies do not produce a definitive diagnosis. In many bioinformatics and referral centers, selected non-diagnostic cases are referred for a subsequent set of analyses, sometimes referred to as research analyses. Research analyses have not been well defined by the field and often rely on local expertise, disparate computational approaches and variant filtration strategies. The variety of approaches may be beneficial for the exploration of novel analytic techniques, but simultaneously confusing to the patient whose data is being analyzed. Furthermore, the goals of research exome analysis are diverging from those of a standard clinical analysis. Clinical analyses continue to be effectively optimized for identifying DNA sequence variants of interpretable significance. Research analyses are, in consequence, concerned with finding variants of unknown significance that nonetheless have potential value for follow up functional validation. The presence of differing goals for clinical and research analyses suggest that different strategies may be appropriate for variant reporting, sharing and annotation in each respective setting. We present: 1. The research exome analysis framework currently employed by the NIH Undiagnosed Diseases Program (UDP) as an example and starting point for discussion of possible future research analysis standards. 2. An analysis of the differences and similarities between clinical and research exome analyses as they have manifested over the last 8 years of the UDP's existence. 3. A description of the variant reporting, sharing and annotation mechanisms we have developed for research-level variants. The rapid implementation of genome-scale sequencing in the clinical environment has blurred the line between clinical and research molecular diagnostics. A clear delineation of the processes used for diagnostic and discovery uses of the technology has the potential to allow for the optimization of both processes. Furthermore, consent and return of results practices will benefit from a better understanding of the differences and similarities between these concurrently generated products of genomic sequencing.

1006W

PseudoSlayer: Clinical-grade variant identification within segmental duplications using NGS data. *D.J. Kvitek, Y. Li, S. Lin, E. Gafni.* Invitae, San Francisco, CA.

Segmental duplications (SegDups) are long DNA stretches of nearly identical sequence that occur two or more times in the genome, arising from duplication events that appear fixed in the population. SegDups complicate variant calling from short-read next-generation sequencing (NGS) data because sequence reads cannot be unambiguously aligned due to the shared sequence identity between SegDup regions. This makes clinical-grade variant calls in these regions difficult or impossible using NGS data and current variant identification approaches.   We have developed a general method for identifying SNPs, indels and CNVs within SegDups. The method consists of two steps. The first step is a bioinformatics screen using NGS data in which sequence reads derived from both the clinical gene of interest and the SegDup are analyzed for the presence of variants using a reference sequence with the gene of interest present but the SegDup(s) masked. If no variants of interest are identified in the screen, no further work is required, and the sample is negative for variants across that gene. If variants of interest are identified, the variant location is disambiguated with additional laboratory assays such as long-range PCR or MLPA, if necessary.   The PseudoSlayer method eliminates the need to perform additional laboratory assays on screen-negative samples, significantly reducing the time and cost for diagnostic testing using gene panels. We have applied this method successfully to several genes, including the Lynch Syndrome gene *PMS2* and its pseudogene *PMS2CL*, an internal 8 exon triplication of the myopathy gene *NEB* and the spinal muscular atrophy genes *SMN1/2*. We have performed an extensive clinical validation of the method for *PMS2*, achieving 100% sensitivity and specificity for SNPs, indels and deletions/duplications across a total of 295 known positive and 34,926 known reference-matching sites.   We are in the process of applying the PseudoSlayer method to other genes of clinical importance that have SegDups elsewhere in the genome. Among them are the autosomal dominant polycystic kidney disease gene *PKD1*, the Shwachman-Diamond syndrome gene *SBDS*, the incontinentia pigmenti gene *IKBKG* (*NEMO*), the Koolen-de Vries Syndrome gene *KANSL1*, and the Gaucher disease gene *GBA*.

1007T

Heterozygous deletions of *FREM1* are not associated with trigonocephaly. *A.J. Dawson[1,2,3], J. Liu[1,2,3], S. Marles[2,3], C.R. Greenberg[2,3], A. Mhanni[2,3], K. Hovanes[4].* 1) Genomics Laboratory Diagnostic Services Manitoba, Winnipeg, Manitoba, Canada; 2) Program of Genetics and Metabolism Winnipeg Regional Health Authority Winnipeg, Manitoba , Canada; 3) Department of Biochemistry and Medical Genetics University of Manitoba Winnipeg, Manitoba, Canada; 4) Combimatrix Irvine, California, USA.

The *FREM1* (*FRAS1*-related extracellular matrix 1) gene (MIM 608944) encodes an extracellular matrix protein involved in the formation and organization of basement membranes in multiple organ systems throughout development. Homozygous loss-of-function mutations in *FREM1*, mapped to 9p22.3, have been shown to cause bifid nose with or without anorectal and renal anomalies (BNAR [MIM 608980]) and Manitoba oculo-tricho-anal (MOTA [MIM 248450]) syndromes. Craniosynostosis is a common craniofacial abnormality resulting in premature fusion of calvarial sutures (coronal, sagittal, metopic and lambdoid). Causative mutations have been identified in more than 10 genes and affect each calvarial suture differently, with variant gene expression patterns in each suture. Trigonocephaly is a rare form of craniosynostosis resulting from premature closure of the metopic suture and usually occurs sporadically. Trigonocephaly type 1 (TRIGNO1 [MIM 190440]) is caused by heterozygous mutations in the *FGFR1* gene (MIM 136350). Heterozygous large deletions of 9p, spanning multiple genes including *FREM1*, have been associated with a syndromic form of trigonocephaly; however, there is variable expressivity and limited genotype-phenotype information. Heterozygous missense mutations of *FREM1* have been reported to be associated with isolated trigonocephaly, with incomplete penetrance, and thus have been designated as TRIGNO2 (MIM 614485) with dominant inheritance. However, in this study using chromosomal microarray analysis, we have identified seven patients with a 60 Kb deletion of *FREM1* spanning exons 8-23 and two patients with a 200 Kb deletion spanning exons 1-6 and including gene *LOC389705* (unknown function). None of these patients show signs of an abnormal metopic suture despite heterozygosity for *FREM1* deletions. All patients are First Nations of Cree or Ojibwa-Cree descent. Three of the seven patients have inherited the identical deletion from their phenotypically normal father. In addition, one patient was homozygous for the 60 Kb *FREM1* deletion and was diagnosed with MOTA syndrome. Our findings suggest that although homozygous deletions of *FREM1* are pathogenic, the heterozygous deletion can be considered a common variant in the Manitoba First Nations population. Therefore, the OMIM designation of dominant *FREM1* deletion mutations resulting in TRIGNO2 is premature.

1008F

Improvements in molecular inversion probe (MIP) assays to increase sensitivity for detection of copy number variants (CNVs). *R. Sapolsky, M. Chadha, O. Mizrahi Man, J. Li, A. Shukla, T. Chen, M. Shapero, J. Schmidt, E.T. Fung.* Affymetrix, part of Thermo Fisher Scientific, Santa Clara, CA.

There is great interest in improving the sensitivity of methods for the detection of CNVs, particularly for complex and/or limiting samples. Relevant situations include liquid biopsies in oncology, pre-implantation genetic diagnosis and non-invasive prenatal diagnosis, and mosaicism. When combined with microarray technology, MIP technology has been used successfully to detect clones present as low as 20% from 80 ng input DNA. We sought to determine whether aberration fraction or input mass could be improved through modifications of the assay. Using the OncoScan® assay as a starting point, we focused on the following steps for assay modification: the ratio of MIPs to genomic DNA concentration, exonuclease activity used to enrich for covalently closed, circular MIP molecules, PCR amplification, and two-color array detection methods. Genomic DNA isolated from cell lines and from clinical samples with varying levels of a spiked-in minor clone was processed through modified MIP assays and analyzed on 96-format array plates designed with high replicates of universal tag probes. Signal profile plots of A signal versus B signal were assessed to determine the quality of A/B allelic separation. Log2 ratio in regions of interest was computed and compared to corresponding log ratio in samples without copy number (CN) aberrations. Sensitivity was determined using spike-in mixtures of 95% normal and 5% aberrated regions, while specificity was determined by measuring calls in normal regions. Titration of the ratio of a 48K MIP panel to genomic DNA from 2000:1 to 51,000:1 revealed a modest increase in MIP circle formation as assessed by real time PCR. While the temperature of the anneal reaction had a demonstrable effect, anneal reaction duration did not greatly impact MIP circle formation. The combined activity of Exo I, III, and VII was optimal for increased CNV sensitivity. The introduction of a hot-start DNA polymerase showed a marked improvement in the separation of genotype clusters on profile plots as compared to conventional Taq DNA polymerase. By combining these modifications with changes to the multi-color array detection such as chemical cross-linking, we improved CN detection sensitivity from 80 ng to ~20 ng and to ~5% minor aberrant clone, with specificity and sensitivity above 90%. These assay modifications may be useful in applications such as analysis of low tumor burden FFPE samples, liquid biopsies, or other low DNA input applications.

1009W

Chromosomal Assessment by Rapid LAMP Analysis (CARLA) – A new test for chromosomal aneuploidy. *H. Shani[1], K. Barazanik[2], V. Wei[2], M. Mauro[2], L. Lodico[2], S. Klugman[1], Z. Williams[2].* 1) Reproductive and Medical Genetics. Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, NY, USA; 2) Program for Early and Recurrent Pregnancy Loss. Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, NY, USA.

Purpose: Rapid and point-of-care testing for genetic abnormalities has been a longstanding goal for molecular diagnostics in pre-implantation, prenatal and postnatal diagnostics. Existing methods for chromosomal assessment are time-consuming, costly and require a centralized high-complexity laboratory. Consequently, diagnosis and treatment is delayed and testing is often not an option in resource-poor settings. We aimed to develop a rapid, inexpensive test that can be performed at point of care for the detection of chromosomal aneuploidy. Methods: The new diagnostic tool, Loop-mediated isothermal amplification (LAMP), previously used for qualitative analysis of target DNA sequences was adapted to detect chromosomal aneuploidy. To our knowledge this is the first time, LAMP has been adapted for quantitative detection of human specific genomic sequences and we termed the test Chromosomal Assessment by Rapid LAMP Analysis (CARLA). CARLA was applied to genomic DNA isolated from peripheral blood samples, cell cultures, and amniotic fluid. Samples were tested using 6 primer sets for two unique single-copy target genes: the *SRY* and *RPA4* genes on chromosomes Y and X, respectively. We optimized the report assay, and used target-specific fluorescent probes that enable real-time monitoring of fluorescence increase during product amplification. Under these conditions, we were able to readily distinguish between male and female sample within 20 minutes and using as little as 1 ng of genomic DNA. By performing CARLA with serial dilutions of target DNA, we were able to quantify target DNA to as low as 3400 target copy number. Normalizing to a target sequence DNA on an autosomal chromosome (*CFTR* gene on chromosome 7) enabled diagnosis of 46XX, 46XY, 47XYY and 45X karyotypes. Summary: CARLA can provide quantitative assessment of relative copy numbers of genomic DNA targets in under 20 minutes without the need for complex laboratory equipment or training. When optimized and adapted, this assay may enable accurate detection of aneuploidy for all human chromosomes. This would enable testing for aneuploidy and CNVs in a wide range of clinical situations that would be orders of magnitude faster, easier, and less expensive to perform than our current modalities.

1010T

Conventional cytogenetics: Gold standard for mapping of large genomic imbalances on chromosome. *M. Rahman[1], N. Rao[1,2], L. Spiteri[3], T. Vo[3].* 1) Clinical Gen Ctr, Specialty Diagnostics Inc, La Mirada, CA; 2) David Geffen UCLA School of Medicine, Dept of Pathology, Los Angeles, CA 90024; 3) Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656.

Conventional cytogenetics represents G-banded routine and or high resolution karyotyping of human chromosomes. Recently, FISH and chromosomal microarray analysis (aCGH) techniques aided in the identification of submicroscopic aberrations such as microdeletions and duplications, often undetectable by karyotype studies. We report here three cases with large genomic imbalances identified by chromosomal microarray analysis (CGH + SNP array) but the precise location of the rearrangement sites or the genomic imbalance had to be confirmed on specific chromosome bands by conventional cytogenetics. **Methodology**: DNA from blood was hybridized to a 400K oligonucleotide array, with probes at an average of 10.5 kb throughout the genome and at 5 kb on the X chromosome. Copy losses or gains are detected in comparison with a control DNA sample. Routine karyotype on G-banded cells is performed and reported. **Results:** *Case 1:* aCGH showed pathogenic GAIN of 11.201 Mb at 18p11.32p11.21, which contains 66 genes and a pathogenic LOSS of 7.152 Mb at 18q22.3q23 containing 39 genes. *Karyotype analysis revealed additional material at chromosome 18q22.3 [46,XX,add(18)(q22.3)] resulting in the LOSS of terminal 18q and confirming GAIN of whole 18p (18p11.32p11.21) at the distal end of 18q which may be a result of a meiotic recombination of a pericentric inversion in a carrier parent. Case 2:* aCGH showed pathogenic LOSS of 3.8 Mb at 14q32.32q32.33 that includes 67 genes, LOSS of 10.5 Mb at 15q11.1q13.2 with 223 genes including SNRPN and UBE3A and LOSS of 0.67 Mb at 15q13.3 with two genes, respectively. *Karyotype analysis revealed an unbalanced translocation between chromosome 14q32.3 and chromosome 15q12, resulting in the loss of distal 14q and the proximal15[45,XX,der(14) t(14;15)(q32.3;q12)], confirming the structural location of genomic imbalance on the derivative chromosome. Case 3:* aCGH showed pathogenic GAIN of 26.11 Mb that contains 190 genes at Yp11.31q11.23, most likely representing an additional whole Y chromosome. *Karyotype analysis revealed an additional Y chromosome with a pericentric inversion involving Yp and Yq [47,XY, inv(Y) (p11.3q11.23], confirming the chromosomal location of genomic imbalance.* **Conclusion:** These are few examples of the many reported in the literature that clearly support the notion that karyotyping studies are essential and complement the microarray results when large genomic imbalances are seen.

1011F

Comprehensive, high through-put workflow for automated gDNA isolation from iSWAB oral samples. *B. El Fahmawi[1], K. Durvasula[2], J. Roeder[2], T. Butts[2].* 1) Mawi DNA Technologies, Hayward, CA; 2) Omega Biotek, Norcross, GA.

High-quality DNA is a critical factor in determining success of many molecular biology applications such as next generation sequencing, qPCR, microarrays etc. Other potential applications include clinical molecular diagnostics, genetic analyses, pharmacogenomics, animal speciation studies, and forensics. Buccal swabs or saliva are the most commonly used DNA sources for these types of studies because of their non-invasiveness and ease of collection. However, these existing non-invasive technologies suffer from low genomic DNA recovery along with high bacterial contamination. Current swab-based products require manual swab removal and are not automation friendly. Mawi DNA Technologies' iSWAB collection devices addresses these issues with its innovative tube insert design that maximizes the release of cells collected on swabs into an existing proprietary selective mammalian cell lysis buffer and allows for long term, room temperature storage of the sample while preventing the release of bacterial DNA. The design is particularly attractive because it eliminates the need for swab retention and is compatible with a variety of commercially available swabs. LIMS compatible barcodes are also included on each iSWAB device for efficient sample traceability and storage purposes. Omega Bio-tek offers the Mag-Bind Blood & Tissue DNA HDQ 96 kit for rapid and reliable isolation of high-quality DNA and its usage has been automated on an open ended liquid handling platform, the Hamilton Microlab® STAR. Here we provide a comprehensive, automated work flow capable of gDNA extraction from 96 iSWAB oral samples in less than 2 hours. The purified DNA was quantified and assessed for its suitability for downstream applications such as real-time PCR. The bacterial contamination of the iSWAB samples was also estimated. Our results indicate that the DNA obtained from the iSWAB samples using Omega Bio-tek's extraction chemistry was of sufficiently high quality and suitable for a variety of downstream applications. The Omega/Hamilton protocol provides an essential framework for the qualification of DNA preparations to ensure cost-effectiveness and reproducibility at different laboratories.

1012W

Accrediting Australasia's first clinical Whole Genome Sequencing (WGS)

service. *L. Burnett[1,2,3,4], A. Borcz[2,3], M. Buckley[1,2,3,4], P. Budd[2,3], G. Chadbond[1,2,3], L. Constantinescu[2,3], J. Copty[1,2,3], M. Cowley[2,3], M. Dinger[1,2,3], L. Ewans[2,3,5], M. Gonzalez[1,2,3], T. Groza[2,3], W. Kaplan[2,3], M. Lubka-Pathak[1,2,3], B. Lundie[1,2,3], J. Mattick[3], S. Mead[1,2,3,4,5], D. Miller[6], P. Morris[2,3], M. Pinese[2,3], T. Roscioli[1,2,3,5], D. Saroufim[1,2,3], A. Statham[1,2,3], J. Stockmyer[1,2,3], K. Ying[1,2,3].* 1) Genome.One at the Garvan, Darlinghurst, NSW, Australia; 2) Kinghorn Centre for Clinical Genomics (KCCG) at the Garvan, Darlinghurst, NSW, Australia; 3) Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; 4) SEALS Genetics, NSW Health Pathology Service, Randwick, NSW, Australia; 5) St Vincent's Clinical School, UNSW Australia, Darlinghurst, NSW, Australia; 6) Illumina Inc. La Jolla, San Francisco, CA.

Background: Garvan's KCCG operates the Illumina HiSeq X-Ten whole genome sequencing (WGS) system. This enables a minimum 30x coverage of the human genome. Application of WGS to clinical practice requires accreditation. Worldwide, very few WGS facilities have gained accreditation and, to our knowledge, none have achieved the international standard of ISO 15189. Aims: To describe the process used to obtain NATA/RCPA (National Association of Testing Authorities/Royal College of Pathologists of Australasia) accreditation for human clinical WGS, including certification to ISO 15189. Methods: Following WGS facility commissioning, a 12-month plan progressively upgraded infrastructure and processes to accreditation requirements. A new company ("Genome.One", a subsidiary of the Garvan) was established to ensure management control over processes involved in clinical report production. Key accreditation developments included a Quality Management System to control procedures and documents, audit and nonconforming event management, and a clinical Code of Conduct for all staff. Innovations included wikis and job ticketing systems facilitating rapid communication between expert teams. WGS-specific issues included design and execution of validation plans to meet regulatory requirements for in-house in-vitro devices (IVDs; laboratory-developed tests; LDTs) and for massively parallel sequencing; online web requesting; information systems meeting regulatory requirements and capable of integrating phenotype ontologies with genotypes; optimisation of filtering pipelines; and structured clinical reports compliant with ACMG Guidelines. Results: NATA/RCPA accreditation to ISO15189 was achieved at first application. Discussion: The initial WGS IVD detects SNVs and indels <20nt in bioinformatically-defined panels. While it can diagnose individuals, it is optimised for familial trios. The second version of our WGS clinical product will also include CNV detection at resolutions comparable to current microarray technologies. Although our IVD already outperformed Sanger sequencing during validation, we will routinely orthogonally confirm medically significant findings in the short term.

1013T

Proposal to clarify population frequency filter criterion for classifying

sequence variants as benign. *S. Harrison[1], B. Funke[1], R. Karchin[2], H. Rehm[1], M. Greenblatt[3], L. Biesecker[4], ClinGen Sequence Variant Interpretation Working Group.* 1) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 2) Johns Hopkins University School of Medicine, Baltimore, MD; 3) University of Vermont College of Medicine, Burlington, VT; 4) National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD.

In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published a guideline for variant interpretation that provides an evidence-based framework to classify variants. The guideline defined 28 criteria, each with an assigned code that addressed heterogeneous types of variant evidence. The committee recognized that the guideline was a starting framework that would evolve over time. ClinGen's Sequence Variant Interpretation (SVI) Working Group aims to optimize and update the ACMG/AMP guidelines for use in ongoing variant curation efforts by ClinGen-supported Expert Panels. Here we address changes to the criterion of "population allele frequency" that allows variants to be classified as benign without the presence of other supporting data. This criterion is designated as "BA1" (for Benign, stand Alone) and was intended to be a genome-wide filter such that any variant above this frequency can be dropped from further consideration of pathogenicity for a Mendelian disorder, with few clearly indicated, well-recognized exceptions. We have changed several attributes of this criterion. The new criterion is "Allele frequency is >0.05 in any general continental population dataset of at least 2,000 alleles for a gene without a gene-specific recommendation." First, we have clarified that the tested individual does not need to be matched in ethnic origin to the population dataset. Second, we have specified that the comparisons can be made to individual, continental population datasets and not necessarily to aggregate data. We have specified that these are population datasets and not controls, since in most cases phenotypes are unknown. The definition also clarifies that the comparison should be performed only on datasets of 2,000 or more alleles, instead of individuals, which is important for X and Y. We have also recognized that there may be alleles that could be pathogenic for a Mendelian disorder at a frequency >.05 and have proposed that ClinGen establish and maintain such a gene-specific exception list. Finally, this specification of the ACMG/AMP rules allows for gene-specific BA1 frequency filters that may be numerically lower than .05, based on criteria established by ClinGen Expert Panels and reviewed by the SVI. These updates to the BA1 criterion will substantially change the practices for interpreting variants and wide-ranging dissemination, discussion, and feedback is warranted.

1014F

Evaluation of the NextGENe CNV caller for use in the clinical laboratory.
T. Lewis[1], J. Durtschi[2], B. Brulotte[2], S. Dames[1], G. Pont-Kingdon[1], R. Mao[3], P. Bayrak-Toydemir[3]. 1) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah; 2) ARUP Laboratories, Salt Lake City, Utah; 3) ARUP Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, Utah.

Statement of Purpose Using Next Generation Sequencing (NGS) data to detect exonic level deletions/duplications is very cost effective and time-saving for diagnostic laboratories, which will be replacing the traditional MLPA or array CGH. Analysis of allele frequency and read depth are the two most common methods. We used read depth data to detect copy number variation and examined the robustness of CNV Caller in the NextGENe software, version 2.4.11. Methods Used NGS read depth data was generated using a 0.5 Mb Agilent SureSelect custom target enrichment panel that captured 1193 exons from 83 genes followed by sequencing on the Illumina NextSeq platform. Samples were sequenced on Illumina NextSeq instruments. Fastq data was processed by BWA aln alignment, and Picard MarkDuplicates for duplicate removal. The average read depth of coverage was 3342. The NextGENe software utilizes the ratio of read depth of the sample to a control set then uses a Hidden Markov Model to account for dispersion (noise) and calculate copy number variations at each region. Ten samples with normal aCGH results were established as the control set. To control for gender, the 10 samples were computationally altered to represent males (single chrX) or females (2 chrX). 114 samples were compared against the gender matched controls as were seven samples known by aCGH to harbor a copy number variation. Summary of Results Using the initial data set of 114 samples, a quality control cutoff was established using the average dispersion for each sample. Of the 93 samples (81%) passing QC, thirty-nine of the samples had no calls. In the remaining 54 samples, there were 95 calls of copy number variation. 85 of the calls represented 12 regions that were called in 5 or more of the samples. These 12 regions were marked as potential false positives. A total of ten unexpected CNV calls were seen in the remaining 8 samples and need to be further evaluated. Seven known positive controls were analyzed using the NextGENe CNV caller and the appropriate gender control set. In each instance, the copy number variation was accurately identified by the software. The preliminary results indicate that the CNV Caller in the NextGENe software is robust enough to consider for use in the clinical lab. Positive results need to be confirmed by a second method but the elimination of aCGH for the majority of samples (85 of 114 samples, 75%) would be both a time and cost saving measure.

1015W

An automated approach for identification and pre-classification for clinical reporting of CNVs and SNVs.
A. O'Hara, Z. Che, V. Wasnikar, S. Shams. BioDiscovery, Inc., El Segundo, CA.

Clinical case review is focused on identification and classification of pathogenic copy number variations (CNVs) and single nucleotide sequence variants (SNVs), while excluding common benign alterations and frequently observed artifacts. Published ACMG recommendations and guidelines call for inclusion and classification of copy number variations (CNV) and sequence variants (single nucleotide variants or SNVs) into different well-defined classes (Pathogenic, Benign, VOUS, etc.) based on a series of set criteria. Variant Interpretation Assistance (VIA) is a system that uses a decision tree machine learning approach to set up rules for automatic pre-classification of CNVs and SNVs. This sophisticated tool includes pre-classification based on user specified designations (defaults include benign, likely benign, VUS, likely pathogenic, pathogenic and artifact); the VIA system then uses region specific functions, such as SIMILARITY, OVERLAP, PREVIOUS_SIMILAR_CASES, and POSITION to allow the user to set up lab specific rules. This includes the presence of genomic contents within the affected interval (intergenic v. intragenic, intronic v. exonic). For CNV, the type (e.g. gain, loss, etc.) and length of the alteration must be considered. For SNV, the type of alteration (nonsense, frameshift, insertion, deletion, etc.) must be considered. Finally, systematic evaluation requires comparison of CNV and SNV to both internal and external databases, including in-house legacy sample results and well-established databases including OMIM, DECIPHER, and ClinGen, respectively. VIA tracks results in real time from previously annotated samples to allow for interpretation of newly uploaded samples based on the evolving in-house results archive. Microarray results on mixed array platforms from normal and affected samples were tested with VIA. A set of previously reviewed samples with a mix of benign and pathogenic classified calls was imported into the system. VIA then pre-classified these same samples and the results were compared. All pre-classification results were annotated with the logic used by VIA to keep a complete audit trail of calling. This system was able to correctly identify pathogenic results for affected samples while finding benign results for normal samples. Final results were then auto-reported in ISCN nomenclature. VIA is capable of quickly and reliably classifying CNVs from microarray and includes an extensive audit trail for potential regulatory demand.

1016T

panelcn.MOPS: CNV detection in targeted panel sequencing data for diagnostic use. *G. Povysil[1], A. Tzika[2], J. Vogt[2], V. Haunschmid[1], L. Messiaen[3], K. Wimmer[2], G. Klambauer[1], S. Hochreiter[1].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria; 3) Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Birmingham, AL.

While various copy number variation (CNV) detection methods exist for whole-genome and whole-exome sequencing data, highly accurate methods for targeted panel sequencing data that are suitable for a diagnostic setting are still missing. The challenges with analyzing this kind of data include the small size and number of target regions as well as their uneven coverage. For clinical applications a method should furthermore be able to detect both short CNVs affecting only single exons or just parts thereof as well as longer CNVs that affect multiple exons or even an entire gene. Another important issue is the risk of incidental findings. Our new method panelcn.MOPS for copy number detection extends cn.MOPS to targeted panel sequencing data. We optimized the design of the count windows, the read counting procedure, the parameters of the model and the segmentation algorithm for targeted panel sequencing. Additionally, several quality control criteria both for samples and targeted exons have been implemented to increase the confidence in called CNVs. In contrast to other CNV detection methods all targeted regions are exploited for the detection of CNVs, but only results for user-selected genes are reported to avoid the risk of incidental findings. We have tested panelcn.MOPS on simulated and real sequencing data. The real sequencing data was enriched with the TruSight cancer panel that targets 94 cancer predisposition genes including *NF*1/2, *BRCA*1/2 and *APC*. The performance of panelcn.MOPS was compared on a data set of 150 samples against several CNV detection tools including NextGENe, ExomeDepth, CoNVaDing, and VisCap. The size of the CNVs ranges from a 20bp deletion affecting only part of an exon over duplications of several exons to a 350kb deletion of an entire gene. In contrast to the other methods, panelcn.MOPS not only achieved a sensitivity of 100%, but also the highest specificity. Furthermore, we do not only provide panelcn.MOPS as an R package, but also as a standalone program with a practical graphical user interface. Therefore, panelcn.MOPS can be conveniently used by users without any programming experience. Our results show that panelcn.MOPS accurately predicts CNVs in targeted panel sequencing data. Consequently complementary biotechnologies to detect CNVs, such as MLPA, can be omitted in order to reduce time and costs.

1017F

Whole exome NGS at the Geneva Genome Clinic: A successful translation to diagnostics through a multidisciplinary team. *E. Hammar[1], S. Fokstuen[1], P. Makrythanasis[1,2], M. Guipponi[1,2], E. Ranza[1], K. Varvagiannis[1], F.A. Santoni[1,2], A. Mauron[4], S.A. Hurst[4], C. Moret[4], C. Gehrig[1], J. Bevillard[2], A. Vannier[1], S. Gimelli[1], E. Stahaki[1], A. Bottani[1], A. Paoloni-Giacobino[1], M. Lidgren[1], F. Sloan-Béna[1], L. D'Amato-Sizonenko[1], T. Nouspikel[1], J.-L. Blouin[1], S.E. Antonarakis[1,2,3].* 1) Genetic Medicine and laboratory, Univerity Hospital of Geneva (HUG), Geneva, Geneva 4, Switzerland; 2) Department of Genetic Medicine and Development, University Medical Center, Geneva 4, Switzerland,; 3) Institute of Genetics and Genomics of Geneva (iGE3), 1 rue Michel-Servet, 1211 Geneva, Switzerland; 4) Institut Ethique Histoire Humanités IEH2, University Medical Center, Geneva 4, Switzerland.

The advances of next generation sequencing (NGS) technologies enable their application in clinical care. Despite this most promising progress, there are several significant challenges for their implementation in clinical diagnostics. In Geneva, we use whole exome sequencing (WES) followed by targeted bioinformatics analysis of individual gene panels for the diagnosis of Mendelian disorders and we have created a multidisciplinary working group, the Genome Clinic Task Force which meets once a week. During these meetings, clinical cases and results are presented, the interpretation of identified variants is debated and the final laboratory reports are critically discussed. Reimbursement and ethical issues such as informed consent, disclosure of incidental findings and/or variants of unknown clinical significance (VUS) are also addressed. In Switzerland, reimbursement of diagnostic NGS tests is integrated in the public health insurance since January 2015. During the pilot year 2015 and the beginning of 2016, a total of 198 cases were analyzed. We found pathogenic variants (class 4 or 5) in 35% of patients with developmental delay, 43% of patients with cardiac diseases, 38% in patients with various neurological diseases and in 41% with various other mendelian phenotypes. The average detection rate of (likely) causative variants was 38%. The decision to report VUS was made on an individual case basis. In order to render our diagnostic use of NGS even more efficient and patient-friendly, our current aims are to 1) improve different steps of the workflow in order to accelerate the analysis process 2) continuously update the variant interpretation methods 3) continuously learn from the challenges during the genetic counseling sessions, such as the informed consent process and 4) develop and implement specific strategies for prenatal exome diagnosis and diagnostic exome analyses of trios.

1018W

Application of exome sequencing for diagnosis and research of rare Mendelian disorders - A preliminary Indian experience. *M. Hebbar[1], A. Shukla[1], J. Shah[1], S.S. Lewis[1], S.L. Bielas[2], K.M. Girisha[1].* 1) Department of Medical Genetics, Kasturba Medical College, Manipal University, Manipal, India; 2) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA.

   Congenital disorders with severe morbidity and mortality have an estimated global birth prevalence of 5% to 7%. The Indian population is unique in being extremely divergent and heterogeneous genetically. The burden of genetic disorders in India, though enormous, remains unknown till date. The practice of consanguineous marriages, as high as 30% in some regions, further precipitates several rare disorders in this population. The high burden of infectious and nutritional diseases and scarcity of appropriate medical care overshadows the diagnosis and care of rare Mendelian disorders. Whole exome sequencing (WES) is emerging as a powerful, easy and cost-effective tool to investigate these disorders in a middle income country like India. Here, we share our experience of utility of WES in a tertiary care centre in India. Forty six families with rare Mendelian genetic disorders underwent WES. The cohort consisted of a spectrum of pediatric neurodevelopmental disorders (34%), skeletal dysplasias (36%) and other rare disorders (30%) mainly from southern part of India. Massively parallel sequencing of whole exome was outsourced and done using the NextSeq500 Sequencer (Illumina Inc., San Diego, CA, USA) in combination with the NextSeq™ 500 High Output Kit (2x150bp). Bioinformatic and clinical data analysis was done locally. A definitive molecular aetiology could be established in 21 (46%) families. Apart from these, we also ascertained two novel genes.. Twenty one families had autosomal recessive conditions, 19 of which had homozygous variants consequent to the high degree of consanguinity in this part of the country. Pathogenic variants delineated included 12 missense, five nonsense, three splice site, three frameshift deletions and a large deletion spanning 2 exons. Molecular diagnosis could not be established in 23% families. Here we report the utility of exome sequencing in a tertiary health care hospital for the first time in India. .

1019T

Reduction in pipette tip consumable cost and waste in the molecular biology. *A. Safavi.* Grenova LLC, 483 Southlake Blvd. Richmond, VA 23236.

   As the number of samples processed in labs using high-throughput processes has increased, there has been a corresponding rise in the level of waste, such that in 2010, four million pounds of plastic pipette tips were disposed of after just one use. The TipNovus high-throughput automated pipette tip cleaning system from Grenova enables laboratories to safely reuse sanitized tips and thereby reduce cost and waste output. Purpose: To validate the Grenova TipNovus tip washing and drying system to ensure that the tips used to add DNA and/or TAQ can be used multiple times following a wash and dry cycle on the Grenova TipNovus System. Precedures - PCR plates were placed on the Hamilton STARlet and ran with water and new tips. Taq was then added by the Hamilton STARlet with new tips. This was done to give a baseline (blank) data for the tips. The plates where then loaded on the thermocyclers and once completed the plates were read. The next step was to test that the Grenova TipNovus could remove any residual DNA after washing and drying the tips. The plates with MMix were placed on the Hamilton STARlet, along with the deep well plates. The PCR set up program was selected on the Hamilton STARlet and DNA was added to the PCR plates with new tips. Taq was added by the Hamilton STARlet using new tips. Plates were sealed, shaken and placed on the thermocyclers. The tips that were re-racked back on the Grenova tip racking tray were taken off the Hamilton STARlet and placed in the Grenova TipNovus and the wash and dry cycles were ran. To test for residual carryover of DNA, the cleaned tips were used to add water in place of DNA on the Hamilton STARlet to the PCR plates containing MMix. New tips were used to add Taq to the plates on the Hamilton STARlet. Plates were sealed, shaken, and placed on the thermocyclers. This was repeated for 25 runs washing the tips each time after patient DNA was added to the plates and always using new tips to add Taq to the water and patient DNA plates. All plates were read on Synergy Plate Reader. The ability of TipNovus to sanitize tips was tested by comparison and validation studies of new and washed tips. A correlation (linear regression) was obtained to monitor reuse of the disposable tips for DNA and TAQ transfer after multiple wash cycles and the rusults concluded that the Grenova TipNovus operated in a way that will allow the use of washed and dried tips to be used multiple times in the Molecular Biology Department.

1020F

Introducing genetics training and competency assessment into day-to-day working practice. *S. Deans[1], F. Khawaja[1], J. Matindale[2], I. Bamsey[3], B. Lings[3], V. Davison[4], A. Seller[4], S. Abbs[5].* 1) UK NEQAS for Molecular Genetics, Edinburgh, United Kingdom; 2) Sheffield Childrens NHS Foundation Trust, Sheffield, United Kingdom; 3) Certus Technology, Hems Mews, Exeter, United Kingdom; 4) Health Education England, Birmingham, United Kingdom; 5) Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom.

The requirement of medical laboratories to work to International Standard ISO15189 has introduced the need to evidence the training and ongoing competency of staff at all levels for all tasks performed. The demonstration of staff to accurately follow testing protocols has been widely adopted, however the measurement of training and competency for routine genetic laboratory duties is problematic. UK National External Quality Assessment Service (NEQAS) for Molecular Genetics has developed the online Genetics Training Assessment and Competency Tool (G-TACT) in conjunction with Certus Technology, with the aim to provide continuing assessment of the ability of individuals working in laboratories to perform both routine and non-routine tasks. These currently include tasks that are undertaken by Sample Receptionists, Duty Scientists, Data Analysts, and Report Authorisers, thereby covering a range of roles within the laboratory. Examples of routine tasks being assessed include logging a sample into a LIMS, answering requests from service users, analysing and interpreting genomic data, and authorising clinical reports. Participation involves entering a virtual genetic testing laboratory and navigating between different workstations to complete tasks. For each participation, the system generates a random scenario, so no individual will complete the exact same task. The system can introduce errors or problems to expose the participant to more challenging scenarios they may encounter, e.g. handling of samples which are not fit for purpose, and analysing and reporting sub-optimal results. The system automatically assesses the participant's handling of the scenario. The modular format enables new workstations, tasks and roles to be introduced as the laboratory environment changes. A manager is assigned to each individual so a laboratory based performance review can be integrated into the appraisal system and training needs can be identified locally and in real time. Multiple assessments can be assigned throughout the year. Currently the assessments cover germline and somatic molecular testing services, and additional assessments are being developed for future, including cytogenetic services. G-TACT offers, for the first time, a consistent approach to ensure individuals at all levels across all UK genetics laboratories, and potentially worldwide, are appropriately trained and competent to perform their assigned tasks to a high quality standard.

1021W

The Exome Clinic and the role of medical genetics expertise in interpretation of whole exome sequencing results. *D. Baldridge[1], J. Heeley[1,2], M. Vineyard[1], L. Manwaring[1], T. Toler[1], E. Fassi[1], E. Fiala[1], S. Brown[3], M. Wiling[1], D.K. Grange[1], B.A. Kozel[1,4], M. Shinawi[1].* 1) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO; 2) Now at Mercy Clinic - Kids Genetics, Mercy Children's Hospital St. Louis, St. Louis, MO; 3) Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO; 4) Now at National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

Purpose: Evaluation of the clinician's role in optimal interpretation of clinical whole exome sequencing (WES) results and the utility of post-exome auxiliary studies and follow up clinical tests in determining the pathogenicity of genetic variants. Methods: Retrospective chart review of the first 155 patients who underwent clinical WES in our Exome Clinic and direct interaction with the ordering geneticist to evaluate the process of interpretation of results. The clinical assessment was categorized as concordant or discordant with the laboratory's classification of the sequence data to analyze how the geneticist's interpretation influenced the final diagnosis. Results: The most common primary indication was neurological problems (~66%), followed by multiple congenital anomalies (~10%). The diagnostic laboratory reported 238 genetic variants, with an average of 1.5 variants per patient and a range from 0 to 6. The overall diagnostic yield was 35% based on sequencing data. After assessment by the medical geneticist, incorporation of detailed phenotypic and molecular data, and utilization of additional diagnostic modalities, the classifications of 97 total variants were changed, altering the final diagnosis of the patient in 21 cases and increasing the final diagnostic yield to 43%. Thirty-two (48%) out of the 66 definitive cases had mutations in genes described in 2011 or later. Seven patients of our cohort were included in initial case series that described novel genetic syndromes and 23% of patients were involved in subsequent research studies directly related to their results or involved in efforts to move beyond clinical WES for diagnosis. We found a higher diagnostic yield for females, patients with a craniofacial anomaly, and patients with an abnormal head circumference. Gene Ontology analysis showed enrichment for genes involved in face morphogenesis/development and brain development. Incidental findings in 10% of our cohort patients often resulted in additional interventions in both the probands and their carrier relatives. The clinical management was directly altered due to the WES findings in 12% of definitively diagnosed cases. Conclusions: Our results emphasize the clinical utility of WES, demonstrate the significant role of the medical geneticist in the diagnostic process of patients undergoing WES, and illustrate the benefits of post-analytical diagnostic work-up in solving the "diagnostic odyssey.".

1022T

Dominant *de novo* variants in consanguineous families: Increased diagnostic yield with WES trio analysis. *A. Begtrup, L. Folk, K. Retterer, B. Friedman, J. Juusola.* GeneDx, Inc., Gaithersburg, MD.

Whole exome sequencing (WES) is a powerful tool for the diagnosis of patients with complex and poorly defined genetic disorders however this technology is more expensive with a longer turn-around-time compared to traditional single gene and gene-panel tests. Individuals with parental consanguinity are more likely to have homozygous variants, therefore one approach to minimize cost and testing time in these individuals is to limit analysis to genes within regions of homozygosity (ROH). Utilizing ROH previously identified via SNP microarray, we developed a targeted exome sequencing test that focuses WES analysis to ROH (ROH Slice). Because this approach does not require samples of parents for analysis the cost is reduced. In this study, we compared the diagnostic yield of ROH Slice to trio WES in cases with a parental kinship coefficient (KC) representing second cousins or closer (KC >0.01). Our clinical diagnostic laboratory performed ROH Slice on 165 samples with a positive rate of 19.4% (32/165), and WES on 502 consanguineous trio cases with a positive rate of 29.5% (148/502). The diagnostic yield of consanguineous WES trio cases was significantly higher than the diagnostic yield of ROH Slice (P=0.0113). Although consanguineous WES trio cases were enriched for positive homozygous variants (31.9%) compared to WES trio cases without detectable parental consanguinity (4.8%; P<0.0001), de novo variants and variants associated with autosomal dominant or X-linked inheritance also contributed heavily to the diagnostic yield in consanguineous WES trio cases. Cases with a parental kinship coefficient representing second cousins (KC ≤0.025 and >0.01) were much less likely to have homozygous positive results. In these cases, more than half of the reported positive variants were in genes with autosomal dominant inheritance (53.6%) and de novo in origin (56.6%), whereas almost one-third (31%) of positive variants were associated with autosomal recessive inheritance. Therefore, while ROH Slice for consanguineous families is a viable option when parental samples are unavailable or full WES is cost prohibitive, WES trio analysis provides a higher diagnostic yield, particularly due to the ability to detect de novo variants associated with autosomal dominant inheritance.

1023F

Comparison of the diagnostic yield and clinical indications for whole exome sequencing among adult and pediatric patients. *A. Crunk, R. Willaert, S. DeWard, K. Retterer, J. Juusola.* GeneDx, Gaithersburg, MD, USA.

The diagnostic yield for whole exome sequencing (WES) ranges from 22-28%, and is affected by clinical indication, age of the proband, and analysis of proband-only versus trio cases. The adult population presents unique diagnostic challenges both in the clinic and in the laboratory. Adults accounted for 19.8% (1874/9464) of the total WES cases completed in our laboratory by March of 2016. The most common diagnoses in adults 30 and over were fatigue (33.4%), muscle weakness (29.5%), and migraine (20.4%). In contrast, the most common diagnoses in pediatric cases, less than 18 years of age, were generalized hypotonia (34.8%), developmental delay (33.3%), and seizures (25.3%). Phenotypic features of adults 18-29 years were a blend of those seen in the other age groups. The definitive diagnostic rate of WES was 17.8% (335/1874) when the proband was over 18, significantly lower than 27.6% (2096/7590) positive rate for pediatric cases (p<0.0001). Dividing the adults further, the positive rate in adults 18-29 years was 24.3% (198/814), but decreased to 12.9% (137/1060) (p<0.0001) in probands 30 years or older. The majority (72%) of pediatric cases were submitted and analyzed as traditional trios, defined as a case including a proband and both biological parents, while only 41% of adult WES cases were trios. In contrast to the higher positive rate of WES trios (29%) compared to proband-only (21%) in the pediatric population, there was no significant difference in adults. Proband-only testing in young adults 18-29 years provided a definitive result in 24.8% compared to 25.4% (p=0.9147) with trio testing. Similarly, among adults 30 years and older, there was no significant difference in diagnostic rate when comparing proband-only cases (13.7%) to traditional trios (14.5%, p=0.7494). Our data demonstrate that adults, specifically those over 30 years of age, are undergoing WES for different clinical indications with lower detection rates than pediatric cases and may benefit less from the trio-based approach.

1024W

Integration of Exomiser into a clinical whole exome sequencing analysis bioinformatics pipeline to aid variant interpretation. *J. Ji, L. Shen, J. Buckley, M. Bootwalla, D. Ruble, J. Galler, D. Maglinte, C. Quindipan, G. Raca, S. Saitta, J. Biegel, X. Gai.* Children's Hospital Los Angeles, 4650 Sunset Blvd., Los Angeles, CA 90027.

Whole exome sequencing (WES) has shown great efficacy in diagnosing patients with suspected genetic disorders. However, the need for multiple manual data interpretation steps and limited adaptation of robust informatics tools for variant prioritization limit rapid turn-around time in a clinical setting. We implemented a semi-automated workflow that integrates a number of publicly available tools and resources including Phenotips for clinical phenotype capture using Human Phenotype Ontology (HPO) terms and Exomiser for variant prioritization (http://www.sanger.ac.uk/science/tools/exomiser). A total of 28 cases (24 singleton-WES and 4 trio-WES) were sequenced using the Agilent SureSelect Human All Exon V6 kit and the Illumina NextSeq 500 sequencing system. Twenty-six cases had known pathogenic or likely pathogenic variants; one case had a variant of unknown clinical significance; and one case was negative. We manually reviewed all rare variants in a blinded manner, facilitated by priority gene lists generated from HPO terms, and achieved 100% concordance rate with prior diagnoses. We then applied Exomiser, a tool that prioritizes genes and variants in a VCF format based upon patients' phenotypes defined in HPO terms, to identify potential disease-causing variants in the 28 cases. In 25/26 (96%) cases, Exomiser ranked the known pathogenic or likely pathogenic variants as the top candidate with high scores (>0.85). In one case, Exomiser initially ranked the known pathogenic variant in 7th place. When a more specific HPO term was added however, the ranking improved to 1st place. For the case with a variant of unknown clinical significance, Exomiser ranked it in 8th place. Exomiser and manual review results were consistent for the negative case, where Exomiser showed low scores for all candidate variants. In conclusion, we have successfully tested and implemented a phenotype-driven automated WES variant analysis pipeline in a clinical diagnostic setting. With Exomiser incorporation, automated analysis is quick, accurate, and unbiased. Our study also suggests that complete and specific clinical phenotypes are critical for phenotype-driven exome analysis.

1025T

A retrospective analysis of exome sequencing cases using the GenePool™ Genomics Platform. *A. Vladimirova[2], T. Klingler[2], R. Goold[2], E. Puffenberger[1].* 1) The Clinic for Special Children, Strasburg, PA 17579; 2) Station X, 185 Berry Street, Suite 2001, San Francisco, CA 94040.

The Clinic for Special Children (CSC) is a rural pediatric non-profit medical practice serving uninsured Amish and Mennonite (Plain) children with genetic disorders. The clinic strives to identify genetic causes of childhood disability and disease and uses modern genetic technologies to diagnose and treat patients. Whole exome sequencing (WES) and data analysis in conjunction with deep phenotyping has enabled the scientific community to achieve great success in identifying the molecular bases of disease. The CSC has used these technologies successfully as well over the past several years. The CSC employs a diagnostic pipeline for new patients that involves detailed phenotyping, targeted mutation detection, chromosomal microarray analysis, and exome sequencing in order to generate a molecular diagnosis for the patient. Due to a deep knowledge of segregating mutations in the Plain populations, nearly 50% of all new patients receive a diagnosis through targeted mutation detection while roughly 3% have diagnostic copy number changes. Of the remaining patients, our diagnostic yield for clinical exomes is approximately 49%. We present a validation study of solved WES cases from the CSC where we demonstrate the ability to efficiently identify putative causative variants in GenePool, a cloud-based genomics platform for analysis of genomics data. We utilized the built-in analytical workflows for trio analysis and the pipelines designed for population-size cohort analyses. The latter analyses compared groups of affected and unaffected individuals. We used GenePool's interactive visualization filters with the comprehensive library of annotations to quickly prioritize the list of potential causative variants to a small highly-relevant set and validated our results. GenePool allowed us to efficiently screen for pathogenic variants associated with autosomal recessive and *de novo* dominant phenotypes, as well as with more complex genetic diseases. Rapid diagnosis is crucial to optimal patient outcomes, and GenePool solves a critical part of this process by enabling the analysis and identification of a small set of putative pathogenic variants in a short time frame. In this study, we found high concordance between GenePool variant prioritization and the prior *ad hoc* manual prioritization. The study we present was conducted in specific regional founder populations, but it provides important lessons for WES studies in non-founder populations.

1026F

Overcoming the challenges of regions of high homology in the clinical genomics laboratory using short-read based next generation sequencing platforms. *D.W. Close, N.S. Reading, G. Pont-Kingdon, C.P. Vaughn, K.V. Voelkerding, A. Kumanovics.* ARUP Institute of Clinical and Experimental Pathology, Salt Lake City, UT.

Regions of homology in the genome, including pseudogenes and other segmental duplications, represent a significant percentage of the genome and lead to unique challenges for targeted genomics in molecular diagnostics. This is particularly true when using short-read based next generation sequencing (NGS) platforms. Certain genes, including *IKBKG* and *PMS2*, are frequently targeted in disease specific panels but nearly identical pseudogene DNA fragments are co-captured and cannot be differentiated from functional gene DNA using standard NGS alignment workflows. Often what is required is supplemental long-range PCR followed by gene specific Sanger sequencing; a process that leads to significant increases in costs and longer turn around times. In this work, we analyze regions of homology relevant to targeted NGS panels and present methodology for processing and analyzing NGS data in the context of segmental duplications containing pseudogenes. In one approach, we demonstrate a method where functional gene specific amplicons are introduced to extracted genomic DNA followed by standard capture and sequencing. Data is then analyzed using a modified NGS pipeline for mapping just the functional gene specific reads, thus eliminating the need to perform supplemental Sanger Sequencing. We highlight an example of this method that centers on the *IKBKG* (NEMO) gene located on the X chromosome. Variants in *IKBKG* have been reported in various forms of Primary Immunodeficiency and Ectodermal Dysplasia disorders, with a number of these variants falling within exons 5-12, a region with 99.9% identity to an inverted segmental duplication located ~36,000 bp upstream of the functional gene. We show that our amplicon spiking method leads to highly accurate sequencing specific to exons 5-12 of the functional gene, without the need for supplemental Sanger sequencing. The method can be applied to other genes with confounding homologous regions such as *PMS2*, and when combined with other methods and bioinformatics tools allow interrogation of gene regions previously inaccessible by standard short-read NGS workflows.

1027W

Development of standardized pathogen detection assays using CleanAmp™ Master Mixes. *A.P. McCaffrey, S. Shore, J.M. Henderson, G. Zon, R.I. Hogrefe.* TriLink BioTechnologies, San Diego, CA.

Acute gastroenteritis (AGE) affects over 179 million people and causes over 600,000 hospitalizations and 5000 deaths in the U.S. annually (CDC). Infectious AGE can be caused by a wide variety of different pathogens including bacteria, viruses and parasites. Rapid, definitive and economical identification of the causative AGE agent could inform medical decisions while reducing the inappropriate use of antibiotics. Standardized molecular methods and multiplexing will make pathogen identification more efficient and economical, with a huge potential global impact. The molecular diagnostics field demands assays with more speed, specificity, and sensitivity. PCR is an established and reliable technology that meets these criteria. Although there are a myriad of published protocols using PCR for detection of specific diseases there is still a lack of established and standardized assays that cover the detection of certain infections. CleanAmp™ dNTPs are a universal Hot Start technology that can be applied to many different types of PCR assays including Real-Time PCR, Multiplex PCR and RT-PCR. CleanAmp™ dNTPs are 2' deoxynucleotide triphosphates with a 3' thermolabile protecting group that prevents incorporation of the dNTPs at room temperature. As temperature is increased the protecting group begins to fall off creating a standard dNTP that can be naturally incorporated into the growing cDNA or PCR strand. CleanAmp™ dNTPs do not have a specific activation temperature but their kinetics are determined by a combination of heat, time and pH that can be optimized for each assay type. CleanAmp™ dNTPs have proved to increase amplification specificity and yield in a variety of assays. We will demonstrate the use of CleanAmp™ dNTPs in robust Multiplex Real-Time PCR and RT-PCR assays that are designed for screening and detection of pathogenic bacteria and viruses in patient stool samples. Primer sets were designed to consensus regions of several strains of each pathogen to maximize detection. Assays were optimized using positive control bacterial isolates and were then tested on DNA and RNA isolated from patient stool samples that may contain inhibitors. Using commercially available CleanAmp™ Master Mixes (CleanAmp™ 2x Multiplex Master Mix and CleanAmp™ 2x RT-PCR Master Mix) we aim to establish and standardize several assay panels for detection of the pathogens causing infectious diarrhea that will make diagnosis and treatment faster and more reliable.

1028T

Use of saliva and salivary DNA for comprehensive genotyping based on RealFast and StripAssays. *C. Oberkanins, A. Berndt, B. Hauser, H. Puehringer.* ViennaLab Diagnostics, Vienna, Austria.

Nucleic acids extracted from blood are the most common sample types used in molecular genetic diagnostics. Blood is an abundant and reliable specimen, yielding high amounts of DNA and RNA with homogenous quality. Since drawing blood is an invasive procedure, and DNA extraction is time consuming and requires special equipment, salivary DNA and saliva direct-to-PCR represent promising alternative approaches in clinical diagnostics. Saliva is considered to represent an ultra-filtrate of blood, and thus contains in abundance a huge variety of diagnostically valuable molecules. Apart from being known to contain polymerase chain reaction (PCR) inhibitors, saliva is mucous and inhomogenous and thus often difficult to handle. Moreover, its human DNA content is by far lower compared to blood. Here we examined blood-derived DNA, salivary DNA and saliva direct-to-PCR to carry out reliable genotyping based on single-plex RealFast and multi-plex StripAssays for up to 22 mutations. Our data suggest that saliva collected and stored in dedicated devices is suitable to be used for isolating ample amounts of high-quality DNA, but also for direct-to-PCR approaches towards comprehensive molecular diagnostics.

1029F

Taking another look: Results of high-resolution microarray analysis in patients with previous microarray testing. *M.R. Sdano, M.M. Martin, R.J. Vanzo, M.A. Serrano, P.A. Mowery-Rushton, E.R. Wassman.* Lineagen, Inc., Salt Lake City, UT.

Chromosomal microarray (CMA) has become the first tier test for individuals with intellectual disability (ID), developmental delay (DD), multiple congenital anomalies (MCA), and/or autism spectrum disorder (ASD). While it is well established that CMA has a higher diagnostic yield than other cytogenetic tests such as karyotype and fluorescence *in situ* hybridization (FISH), few studies have looked at the diagnostic yield between microarray platforms. The purpose of this study was to compare the results of a patient's previous microarray with the results of a newer ultra-high resolution CMA. Patient data was retrospectively collected from 7,412 patients consecutively referred for clinical microarray testing by their healthcare provider between August 2012 and May 15, 2016. CMA was performed using a commercial microarray optimized for diagnosis of ID/DD/MCA/ASD through the addition of 88,435 custom probes of relevance to these disorders to an Affymetrix array platform resulting in a 2.8M probe microarray (FirstStep$_{Dx}$ PLUS). Referral information or case medical records revealed that fourteen patients (0.02%) previously had CMA with one of a variety of lower resolution microarrays. For seven of the 14 patients with a previous microarray, new information was obtained by FirstStep$_{Dx}$ PLUS, with four of these patients receiving a new or updated genetic diagnosis. The four new diagnoses were cerebral cavernous malformation (CCM) syndrome (patient 1), KBG syndrome (patient 2), 49,XXXXY (patient 3; previously diagnosed as 47,XXY), and a mosaic 13q deletion (patient 4). Additionally, FirstStep$_{Dx}$ PLUS was able to further refine the breakpoints and gene content of previously identified deletions and duplications in three of the seven patients. This case series highlights the benefits of ultra-high resolution CMA over lower-resolution platforms since clinically relevant information was obtained in 7/14 (50%) patients. Identifying previously unreported copy number variations led to a new genetic diagnosis in four of 14 patients (28.6%), suggesting that re-analysis of patients who previously had low resolution CMAs should be considered. These new diagnoses resulted in more personalized medical management as well as new recurrence risk information for the family. When choosing a CMA platform for a patient, it is important to choose the highest resolution platform available in order to have the greatest chance to identify a genetic condition, thus impacting clinical management.

1030W

Automated evaluation of *de novo* **copy number variants from SNP array trios using pCNV.** *A.P. Wiita[1], C.A. Mather[1], J.Y. An[2], S.J. Sanders[2].* 1) Laboratory Medicine, University of California, San Francisco, San Francisco, CA; 2) Psychiatry, University of California, San Francisco, San Francisco, CA.

   Introduction: The identification of clinically relevant copy number variants (CNVs) in single nucleotide polymorphism (SNP) arrays can be complicated. While some pathogenic variants may be inherited from a carrier parent, many deleterious variants will be *de novo.* Our ability to confidently assign *de novo* status to small CNVs, however, is hampered by noise inherent to the assay and minimum size thresholds for CNV reporting. We therefore used a new computational method, pCNV, to identify and assign confidence values to *de novo* CNVs identified in parent-child trios. **Methods:** SNP arrays were performed using the Illumina CytoSNP-850K assay and initially evaluated using Illumina BlueFuse software for standard clinical reporting. The pCNV method was initially described by Sanders *et al.* (*Neuron* (2015) 87:1215). Briefly, all CNVs were predicted using existing algorithms and the specificity of de novo CNV predictions was estimated using a per CNV p-value (pCNV), based on the per SNP variability in Log R Ratio (LRR) and the number of SNPs consistent with a deletion/duplication based on B Allele Frequency (BAF). *De novo* CNV calls were compared between the pCNV and the BlueFuse outputs. **Results:** We evaluated 16 trios and one quintet. Twenty-one CNVs that met our reporting criteria were identified in the probands. Manual review of the probands and relatives revealed that five of these CNVs were *de novo*. These were confirmed by pCNV, which also reported one other, 0.94 Mb, *de novo* duplication. This duplication was confirmed on review of the BlueFuse output. **Conclusions:** The pCNV program accurately identified all reported *de novo* CNVs in our study set. This algorithm can automatically identify *de novo* CNVs from trio samples without manual review of the data, which could streamline the process of identifying and evaluating CNVs of uncertain significance. The finding of an additional, *de novo* CNV in our sample set raises consideration that pCNV could be used to detect causative CNVs that might otherwise be dismissed because they are too small to meet standard reporting criteria in the clinical laboratory. Furthermore, our study demonstrates the potential utility of performing trio SNP array analysis, analogous to routinely-performed trio whole exome sequencing, to enhance detection of *de novo* variants which may be of clinical relevance but would otherwise be ignored.

1031T

Modification of Eureka Genotyping for detection of RNA without conversion to cDNA. *J. Curry[1], P. Dickinson[2], A. Schweitzer[2].* 1) Eureka Genotyping, Affymetrix part of ThermoFisher Scientific, Hercules, CA; 2) Affymetrix part of ThermoFisher Scientific, Santa Clara, CA.

   Analysis of RNA often suffers from a bias due to its conversion to cDNA prior to analysis. We have modified the Eureka Genotyping Solution, which utilizes next generation sequencing and enables the simultaneous genotyping of hundreds of thousands of DNA samples for tens to thousands of loci, for the direct interrogation of RNA. This solution is based on ligation dependent PCR and uses interrogation site probes as well as sample index barcodes that are added during amplification. The utility is demonstrated by performing a highly multiplexed reaction that uses a commercially available RNA ligase to ligate DNA probes hybridized to RNA templates. Next generation sequenced data is created from the resulting PCR products. Each read is assigned to a sample (based on index) and to a locus. The sequence produced from left and right probe ligation is examined to determine if/what type of RNA splicing occurred by reading from the left portion into the adjoined/ligated right portion. Here we show the results of a 778-plex panel of probes designed to interrogate the RNA produced from housekeeping genes and from human gene exons selected for cancer fusion gene detection. The expected exon junctions were found in the house-keeping genes. Additional uses of Eureka Genotyping for RNA include the targeted study of expression, allele-specific expression, alternative splicing, and fusion gene detection. This direct from RNA detection is a simplified assay that also removes the RNA to cDNA conversion bias.

1032F

Phenotypes of Distinction": When and how to integrate unique phenotypic information into variant interpretation. *B. Johnson, S. Topper, S. Aradhya, A. Entezam, Y. Kobayash, K. Mitchel, C. Morgan, R. Nussbaum, J. Tahiliani, K. Nykamp.* Invitae, San Francisco, CA.

For well-described hereditary diseases with a specific phenotype, there is a relatively high prior probability that pathogenic variant(s) will be detected, if the appropriate genes have been sequenced. The presence of a unique phenotype in a patient can provide a powerful line of evidence for variant classification. We set out to establish a systematic approach for integrating unique phenotypic data into variant interpretation. In the course of sequence variant classification, the potential exists to conclude incorrectly that a novel VUS found in a patient with a disease must be causative. This incorrect conclusion arises from an overestimate of the prior probability that any variants in a patient found in a gene consistent with the patient's disorder must be causative. The likelihood that any variants detected in a gene cause disease, depends on how distinctive the phenotype of the individual is, the degree of locus heterogeneity, the fraction of locus heterogeneity that is accounted for by the genes being tested, and the prevalence of phenocopies in the population. It is also modulated by the observed genotype in the patient. The ACMG Interpretation of Sequence Variants guidelines address this subject with a single criterion: PP4, "Patient's phenotype or family history is highly specific for a disease with a single genetic etiology". We sought to clarify the usage of this rule with more detailed, gene-level guidelines. We use a point system for variant interpretation that is based on the ACMG guidelines. We defined a new set of evidence-based criteria that can be applied during variant interpretation when the following criteria are met: (1) our diagnostic yield is >75% for the gene(s) tested, (2) the clinical features described in a given patient must be so specific that they are essentially pathognomonic for the disorder, and (3) the patient's genotype must match the expected inheritance of the disease. The relative weight of the criteria is determined by the likelihood that a patient's observed genotype explains disease. As a result of these new criteria, we have been able to classify multiple rare variants as likely pathogenic or pathogenic that would have otherwise been classified as VUS by an approach that ignores phenotype. Thirteen patients have received a positive diagnosis for primary ciliary dyskinesia, L1 syndrome, and pyridoxine responsive epilepsy who would have received uncertain findings without the use of these criteria.

1033W

Signature of positive selection at the *CDK6* locus in Pima Indians of Southwest Arizona and association with type 2 diabetes. *P. Kumar, P. Piaggi, A. Nair, P. Chen, S. Kobes, R. Hanson, C. Bogardus, L. Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ.

Polygenic diseases often disproportionately affect certain ethnic groups. In some cases, positive selection of genomic regions may have led to population-specific enrichment of risk alleles. Here, we analyze the signatures of positive selection in the Pima Indians of Arizona who suffer from a high prevalence of obesity and type 2 diabetes (T2D). SNPs detected in whole genome sequence (WGS) data in 296 Pima Indians were integrated with 1000 Genomes data. Population differentiation analysis (F_{ST}) between Pima and Caucasian (CEU) identified rs2237572 (F_{ST}; 0.8) as one of top 1% differentiated signals. The derived allele at rs2237572(C) had a higher frequency in Pima (86%) and Chinese (CHB; 80%) as compared to CEU (4%). This variant maps to the *CDK6* gene which regulates the cell cycle. Overexpression of CDK6 and cyclin D3 in β-cells results in efficient replication and proliferation; therefore, the *CDK6* locus was studied further. Summary statistics of neutrality, including nucleotide diversity and Tajima's *D* across *CDK6* showed reduced diversity and a negative *D* in Pima as compared to CEU and CHB based on coalescence analysis for different recombination events. An integrated haplotype score (iHS) analysis for extended haplotype homozygosity of *CDK6* identified a long range haplotype around rs2237571 (iHS score; -3.4471 and F_{ST}; 0.3), which shares the same linkage disequilibrium block as the high F_{ST} variant rs2237572. This extended haplotype (600 kb) contained the rs2237571 derived allele (G), suggesting positive selection in the Pima population. Genotyping of the variant rs2237571 in a population-based sample of Pima Indians showed that the derived allele (G) modestly associated with T2D (N=7,710 *P*=0.001 after genomic control, odds ratio=1.23 per copy [95% confidence interval, 1.10-1.41]; adjusted for age, sex, birth year, and first 5 principal components derived from GWAS). Among 300 normal glucose tolerant Pima Indians assessed for acute insulin response to glucose bolus, carriers of the derived allele had an impaired insulin response (P=0.0023, beta as multiplier=-0.090) even after adjustment for differences in fatness. Our results suggest an important genetic role and selection of *CDK6* gene in Pima Indians and their association with T2D and insulin secretion. Future studies to functionally characterize *CDK6* variation are planned.

1034T

The high prevalence of *UGT1A1* TA8 allele in the Tunisian population as a possible African ancestral protection factor in congenital unconjugated hyperbilirubinemia diseases. *N. Abdelmoula[1], N. Abdellaoui[1], R. Abdelhedi[1,2], N. Kharrat[2], H. Ameur[3], R. Louati[1], D. Dlansi[1], W. Aloulou[3], S. Aloulou[3], P. Lohse[4].* 1) Lab Histology, Univ Medicine, Sfax, Tunisia; 2) Lab Molecular and cellular screening process , CBS, Sfax, Tunisia; 3) Medical university, Sfax, Tunisia; 4) Dep Clinical Chemistry, Munich, Germany.

UGT1A1 which is the major UGT1 gene product is located on chromosome 2q37. Expression of UGT1A1 is, in part, controlled by a polymorphic dinucleotide repeat within the UGT1A1 promoter TATA box consisting of between 5 and 8 copies of a TA repeat, with A(TA)6TAA considered as the wild type. In the regulatory region of UGT1A1, many polymorphic variations such as the phenobarbital responsive enhancer module (PBREM) are described. The A(TA)7TAA allele has been identified as the most frequent allele in Caucasian populations whereas A(TA)8TAA allele, described for the first time by Beutler E et al. (1998) in Afro-Americans with different genotype combinations and later on detected in Sub-Saharian Africans, remains the most rare allele worldwide including in North African and Arab populations. Here we report results of a Tunisian study in which our goal was to determine, using bidirectional sequencing, the UGT1A1 gene mutations and TATA box polymorphisms in Crigler Najjar 1 and 2 patients as well as their parents and family members. Among 82 Tunisian alleles (42 subjects), A(TA)8TAA was detected 13 times (16 percent-PC-) versus A(TA)6TAA 45 times (54 PC) and A(TA)7TAA 24 times (30 PC). Genotypes were distributed as the following: TA6/6 in 5 subjects (12,2 PC), TA6/7 in 14 subjects (34,1 PC), TA7/7 in 14 subjects, TA7/8 in 3 subjects (7,3 PC) and TA8/8 in 5 patients (12,2 PC). It is the first study in which such high frequency of TA8 allele was reported particularly with but also without the association to the Tunisian CN1 mutation (UGT1A1 exon 3 mutation: c.1070A-G). TA8 subjects included 5 double homozygous CN1 cases, 2 double heterozygous parents and one TA7/8 heterozygous subject. Moreover, TA7/8 status of UGT1A1 was not associated with obvious manifestations of Gilbert trait. The same observation was done for the only Tunisian reported case of TA7/8 among the healthy cohort (n= 137) of Chaouch L et al. 2011. Additionally, CN patients with TA8/8 status seem to have better outcome with a survival rate of 10 years for 1 out of our 5 cases. We suggest that the TA8 UGT1A1 promoter polymorphism may be a prevalent allele in our Tunisian population and may be a protection factor against high bilirubin levels. This protection may be secondary to the DNA features of the ancestral African haplotype especially in the regulatory region of UGT1A1 (such as the absence of -3275T-G at PBREM which is frequently associated to TA7 allele).

1035F

Pinpointing *IVD* regulatory variants that impact leucine metabolism in a region under selection in East Asians. *E.A. Brown[1,2], R. Tewhey[1,2], J. Vitti[1,2], S. Schaffner[2], P. Sabeti[1,2].* 1) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge, MA.

Only a few genetic loci have been firmly linked to metabolic adaptations despite genome-wide selection scans indicating that infectious disease, climate, and diet have exerted strong recent (50kyr) selective pressure on human metabolism. While we previously used the Composite of Multiple Signals (CMS) test to identify and narrow in on a strong signal of positive selection at the locus of the metabolic gene isovaleryl dehydrogenase (*IVD*) in East Asian populations, neither the underlying adaptive sequences nor their phenotypic effects have been characterized. To address this, we comprehensively characterized genetic variants falling in this selection region for their ability to regulate gene expression. We chose to focus on variants overlapping the selection peak that associated with *IVD* expression and metabolites of *IVD* in the blood for two reasons: First, our broad goal was to extend selection scans to find candidates for adaptation by intersecting our CMS loci with loci that associate with gene expression because the majority of selected loci likely function through non-coding regulatory alterations. Secondly, *IVD* catalyzes digestion of leucine, an essential amino acid in the human diet with roles in muscle anabolism and insulin signaling early in development, which makes it an intriguing candidate for adaptation. Leucine is found at high concentrations in eggs and cheeses, as well as soy beans, which originally were domesticated in East Asia. Using the luciferase reporter gene assay, we identified two variants whose derived alleles increased expression of IVD and are at 80% frequency in East Asians. We confirmed these alleles endogenously regulate *IVD* using CRISPR/Cas9-based homologous replacement of alleles in the same cell lines. Additionally we dissected the sequence determinants of regulation using a multiple parallel reporter assay (MPRA) to bash the locus to identify transcription factor motifs responsible for regulatory function. Lastly, we further refined the selection signal by dating the functional alleles and analyzing selection measures across a diverse set of human populations. Detecting and exploring regulatory variants of *IVD* lays the groundwork for understanding natural variation in important human traits affected by leucine in diverse human populations. .

1036W

Natural selection at circadian clock–fasting glucose-related gene loci associated with the cryptochrome 2 and melatonin receptor 1B genes. *I. Yoshiuchi.* Dept Med, Yoshiuchi Med Diabetes Inst, Kamakura, Kanagawa, Japan.

Type 2 diabetes mellitus (T2DM) is highly heritable and shows significant variation in prevalence among populations. The circadian clock has been linked with diabetes and obesity, and short sleep is associated with risk of diabetes and obesity. According to the thrifty gene hypothesis, natural selection for diabetes and glucose-related genes is important during feast and famine because such genes control glucose levels. Human adaptations to environmental changes in food supply, lifestyle, and geography may have influenced the selection of genes associated with the metabolism of glucose, lipids, and energy. The cryptochrome 2 (*CRY2*) and melatonin receptor 1B (*MTNR1B*) genes are associated with fasting glucose, and are also related to the circadian clock . Single nucleotide polymorphisms (SNPs) in *CRY2* and *MTNR1B* have been linked with fasting glucose levels. Here, we investigated selection at circadian clock –fasting glucose-related gene loci associated with *CRY2* and *MTNR1B* . We performed a three-step method to detect selection at the *CRY2* and *MTNR1B* loci using HapMap population data. We used Wright's F-statistics as a measure of population differentiation, the long-range haplotype test to identify extended haplotypes, and the integrated haplotype score (iHS) to detect selection at *CRY2* and *MTNR1B* . We found high population differentiation between European and African populations at one *CRY2* and one *MTNR1B* fasting glucose-associated SNP using Wright's F-statistics. The iHS also showed evidence for natural selection at the *MTNR1B* locus. These findings provide evidence for natural selection at the *CRY2* and *MTNR1B* loci . Further investigation is warranted into adaptive evolution at circadian clock –fasting glucose-related gene loci.

1037T

Association between variants of melanocortin 1 receptor *(MC1R)* gene and the systemic lupus erythematosus development. *N.F. Cagnin[1], F.B. Barbosa[1], L. Marano[1], C.T. Mendes-Junior[2], E.A. Donadi[3], A.L. Simoes[1].* 1) Genetics, University of São Paulo, Ribeirão Preto, São Paulo, Brazil; 2) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil; 3) Division of Clinical Immunology, Ribeirão Preto Medical School, University of São Paulo, Brazil.

Systemic lupus erythematosus (SLE [MIM: 152700]) is an autoimmune disease, which has higher incidence and prevalence in African descent people of non-tropical regions compared to European descendants and Africa inhabitants. It has been proposed that vitamin D deficiency contributes to the development of SLE, because the active vitamin form has biological functions in the immune system. The capacity to synthesis of this vitamin is lower in darkskinned individuals. They have much of melanin pigment that acts as a physical barrier to the UVB rays incidence, essential for the vitamin synthesis. The melanin synthesis is triggered by activation of melanocortin 1 receptor (MC1R) and polymorphisms in coding region of the receptor gene contributes to the determination of human skin pigmentation phenotype. Facing the possible association between *MC1R* (MIM: 155555/HGNC:6929) gene and SLE, this study aimed to investigate the association between SNPs of *MC1R* gene and the occurrence of SLE and, compare ethnic profiles of patients and control by Ancestry Informative Markers (AIMs). Using genetic sequencing, a total of 18 SNPs distributed throughout the whole coding region of *MC1R* gene was genotyped in DNA samples from 138 patients and 123 controls. Departures from Hardy-Weinberg equilibrium weren't found. Allele, genotype and haplotype frequencies were obtained and compared between patients and controls using logistic regression test, in which sex and age variants were controlled. The haplotype B1 of *MC1R* (p=0.016; OR=0.443; CI=0.219-0.855), the allele T of the 1558GT locus (p=0.012; OR=0.422; CI=0.216-0.827), that features this haplotype, and its heterozygous GT genotype (p=0.026; OR=0.434; CI=0.208-0.906) were associated with protection to SLE. Additionally, we genotyped 12 AIMs and their allelic frequencies were used to estimate ancestral composition of sample groups. The ancestral composition of patient group and control was similar: 64% and 70% European contribution, 22% and 17% of African and 14% and 13% of Amerindian, respectively, and there was no genetic difference between the groups. These results revealed that there is an association between *MC1R* gene and the SLE development.

1038F

Genetic adaptation and Neandertal admixture shaped the immune system of modern human populations. *H. Quach[1,2,3], M. Rotival[1,2,3], J. Pothlichet[1,2,3], Y.-H.E. Loh[1,2,3], M. Dannemann[4], N. Zidane[1,2,3], G. Laval[1,2,3], E. Patin[1,2,3], C. Harmant[1,2,3], M. Lopez[1,2,3,5], M. Deschamps[1,2,3,5], N. Naffakh[6], D. Duffy[7], A. Coen[8], G. Leroux-Roels[8], F. Clément[8], A. Boland[9], J.-F. Deleuze[9], J. Kelso[4], M.L. Albert[7,10], L. Quintana-Murci[1,2,3].* 1) Human Evolutionary Genetics, Institut Pasteur, Paris, Ile de France, France; 2) CNRS, URA3012, Paris 75015, France; 3) Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris 75015, France; 4) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany; 5) Université Pierre et Marie Curie, Cellule Pasteur, Paris 75015, France; 6) Institut Pasteur, Molecular Genetics of RNA Viruses Unit, Paris 75015, France; 7) Institut Pasteur, Dendritic Cell Immunobiology Unit, Paris 75015, France; 8) Center for Vaccinology, Ghent University and University Hospital, Ghent 9000, Belgium; 9) Centre National de Génotypage, CEA, Evry 91000, France; 10) Department of Cancer Immunology, Genentech, California 94080, USA.

The immune response to stress is a highly complex phenotype. Inappropriate immune activity can increase susceptibility to infectious, inflammatory, and autoimmune diseases, the clinical manifestations of which vary considerably in humans, at both individual and population levels. Yet, the extent of population differences in immune responses, and their genetic and evolutionary determinants, remain undefined. Here, we characterized, using RNA-sequencing, the transcriptional response of primary monocytes, from 100 Africans and 100 Europeans, exposed to various bacterial and viral stimuli — ligands activating Toll-like receptors pathways (TLR1/2, TLR4 and TLR7/8) and influenza virus — and mapped expression quantitative trait loci (eQTL). We find that marked differences in immune responses exist within and between human populations, due to the contribution of multiple *cis*- and *trans*-acting regulatory variants. These include, for example, a *TLR1 trans*-regulatory hotspot associated with the expression of a large gene network of 101 genes, which decreases the expression of multiple pro-inflammatory genes in Europeans only. Furthermore, we show that natural selection has preferentially targeted loci affecting monocyte gene expression, and identify a list of immune-responsive regulatory variants that have conferred different selective advantages in African and European populations. Specifically, our analysis uncover attenuated inflammation as an evolutionarily advantageous mechanism, as attested by the cases of *TLR1*, *CCR1* and *DUSP14*. Finally, we demonstrate that admixture with Neandertals has shaped monocyte gene expression by introducing regulatory variants into European genomes that affect preferentially responses to viral challenges, and report novel cases of adaptive introgression, as highlighted by the *cis*-eQTL regulating *PNMA1* in response to influenza infection. Overall, our analyses provide a comprehensive view of the impact of population genetic differences on transcriptional responses to innate immunity activation, and highlight evolutionarily important determinants of host immune responsiveness. The regulatory variants identified in our study constitute a useful resource for evaluating the role of such variants in the molecular and cellular mechanisms underlying host immunity to infection and susceptibility to disease, at the individual and population levels.

1039W

Insights into the geographical distribution of genetic admixture of unrelated volunteer donors and recipients of stem-cell transplants. *A. Madbouly[1], K. Besse[1], Y. Wang[2], J. Byrnes[2], C. Ball[2], N. Myres[2], M. Maiers[1].* 1) Bioinformatic Research, National Marrow Donor Program, Minneapolis, MN; 2) Ancestry.com, San Francisco, CA, USA.

Genetic ancestry of self-described groups may vary across geographic locations in the US, a phenomenon documented anecdotally but not thoroughly explored in the literature. We studied the genetic ancestry of 995 HLA matched donor/recipient (DR) pairs from the Be The Match® registry with a focus on regional ancestry differences among ethnic groups. We hypothesized that, along with historical events, donor/transplant center distribution and socioeconomic factors might influence the geographical spread of some genetic admixtures. We genotyped 995 DR pairs on the Illumina OmniExpress chip with approximately 730,000 SNPs. Self-reported race and ethnicity was collected for donors at the time of registry recruitment. Recipients' race and ethnicity was recorded at the transplant hospital once at the time of diagnosis and again after transplant. The majority of the study cohort (94%) self-identified as European Caucasian (CAU). The rest identified as Hispanic (HIS) (3.5%), African-American (1%) and Asian or Pacific Islander (1.5%). Address zip code information was available for 99% of recipients but only 59% of donors. Genetic ancestry was estimated by applying the AncestryDNAÔ ethnicity estimator pipeline, which provides a vector of 26 admixtures. Some admixtures were combined for the analysis due to small counts and minimal impact such as detailed African (AFR) admixtures. We then mapped the geographical distribution of European (EUR) and non-EUR genetic admixtures for self-reported CAU and non-CAU individuals, optimizing geographical regions for subject privacy. The main self-reported race groups showed average proportions of AFR and EUR admixtures compatible with Bryc and colleagues (2015). However, our results revealed larger Amerindian admixture in self-reported HIS, especially among recipients. When stratifying regionally, systematic differences emerged in admixture distribution among similar race groups mostly interpretable by historic events. Separating donors and recipients suggested possible additional influences, such as donor and transplant center geographical spread. Importantly, we observed differences in the distribution of non-majority admixtures such as increased AFR admixture in self-reported CAU donors (but not recipients) in some southern states suggesting a possible socioeconomic link. This work has the potential of guiding stem-cell donor registry strategies on volunteer donor recruitment and donor and transplant center planning.

1040T

The MHC Diversity in Africa Project (MDAP) Pilot - 125 African high resolution HLA types from 5 populations. *M.O. Pollard[1,2], N. Park[3], C. Pomilla[1], S. Peacock[5], D. Dilernia[4], D. Gurdasani[1], M. Quail[3] on behalf of MDAP Investigators.* 1) Genetic Epidemiology, Department of Medicine, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) Human Genetics Informatics, Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 3) Sequencing Research And Development, Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 4) Emory Vaccine Center, Emory University, Atlanta, GA; 5) Histocompatibility and Immunogenetics Laboratory, Cambridge University Teaching Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Hills Road, Cambridge, UK.

The major histocompatibility complex (MHC), or human leukocyte antigen (HLA) in humans, is a highly diverse gene family with a key role in immune response to disease; and has been implicated in auto-immune disease, cancer, infectious disease susceptibility, and vaccine response. It has clinical importance in the field of solid organ and bone marrow transplantation, where donors and recipient matching of HLA types is key to transplanted organ outcomes. The Sanger based typing (SBT) methods currently used in clinical practice do not capture the full diversity across this region, and require specific reference sequences to deconvolute ambiguity in HLA types. However, reference databases are based largely on European populations, and the full extent of diversity in Africa remains poorly understood. Here, we present the first systematic characterisation of HLA diversity within Africa in the pilot phase of the MHC Diversity in Africa Project, together with an evaluation of methods to carry out scalable cost-effective, as well as reliable, typing of this region in African populations. To sample a geographically representative panel of African populations we obtained 125 samples, 25 each from the Zulu (South Africa), Igbo (Nigeria), Kalenjin (Kenya), Moroccan and Ashanti (Ghana) groups. For methods validation we included two controls from the International Histocompatibility Working Group (IHWG) collection with known typing information. Sanger typing and Illumina HiSeq X sequencing of these samples indicated potentially novel Class I and Class II alleles; however, we found poor correlation between HiSeq X sequencing and SBT for both classes. Long Range PCR and high resolution PacBio RS-II typing of 4 of these samples identified 7 novel Class II alleles, highlighting the high levels of diversity in these populations, and the need for long read sequencing approaches to characterise this comprehensively. We have now expanded this approach to the entire pilot set of 125 samples. We present these confirmed types and discuss a workflow for scaling this to 5000 individuals across Africa. The large number of new alleles identified in our pilot suggests the high level of African HLA diversity and the utility of high resolution methods. The MDAP project will provide a framework for accurate HLA typing, in addition to providing an invaluable resource for imputation in GWAS, boosting power to identify and resolve HLA disease associations.

1041F

Population diversity in transcriptional responses of macrophages to TLR7/8 signaling. *R. Song[1], I. Dozmorov[1], R. Levitz[2], C. Liang[1], B. Zhang[1], B. Wakeland[1], C. Arana[1], K. Viswanathan[1], J. Kahn[2], E.K. Wakeland[1].* 1) Department of immunology, UT southwestern medical center, Dallas, TX; 2) Department of Pediatrics, Division of Infectious Diseases, University of Texas Southwestern Medical Center, Dallas, TX,.

The Toll-like receptors (TLR) of the innate immune system play a key role in the recognition of pathogens and the initiation of a robust innate immune response. Although an innate immune response is essential for resistance to pathogen infection, the magnitude and qualities of innate immune responses are quite variable in the human population. Here we performed transcriptome sequencing of monocyte-derived macrophages generated from 185 healthy individuals, including 10 replicates from repeat donors, to profile gene expression patterns in responses to TLR7/8 signaling. Extensive qualitative and quantitative diversity was apparent in the response of individuals in this panel. Cluster analysis discriminated seven distinct clusters of genes expressed in response to the TLR7/8 agonist (R848) with correlated expression variations within this panel. Interestingly, the induction of cytokine and chemokine transcription varied among donors and majority of these genes have very reproducible variability of their expression in 10 replicates. In addition, it was shown that donors with high expression of cytokine and chemokine genes after R848 stimulation were also high responders to Respiratory syncytial virus infection. Furthermore, it was shown that distinct/similar time dynamics of gene expression in high and low responder in a time course of R848-induced gene expression. Especially, expression of phagocytosis genes was increased in only low responder after 12 hours of R848 stimulation. These results suggest that variations in gene expression among individuals following TLR7/8 stimulation involve coordinate variations in clusters of genes which mediate specific effector functions, suggesting that genetic polymorphisms in "master" regulators may diversify macrophage activation and function within human populations.

1042W

Prevalence of immunity-dependent disease-associated SNPs in human populations. *A. Cherednichenko[1], V. Stepanov[1,2], E. Trifonova[1,2], K. Vagaitseva[1,2], A. Bocharova[1].* 1) Research Institute of Medical Genetics, Tomsk, Russian Federation; 2) National Research Tomsk State University, Tomsk, Russian Federation.

The immune system provides protection from a wide range of pathogens. Pathogens have represented an significant selective force in the process the adaptation of human populations to changing environment. Therefore, immune system is one of the functions enriched with genes under natural selection. Here, allele frequency distributions for gene polymorphisms associated with autoimmune and allergic diseases, as well as with the regulation of cytokines levels and immunoglobulin E, were studied in 26 populations of Northern Eurasia in comparison with data in 31 populations from 1000 Genomes, HapMap, HGDP projects. The frequencies of genotypes demonstrated wide variability. The level of genetic differentiation of worldwide populations was 12.5%. It was shown that allele frequency for most SNP demonstrated significant correlation with one or more climatic and geographic parameters as well as with the incidence of parasitic diseases. Significant correlations between average expected heterozygosis in 27 loci in 57 populations and average annual level of precipitations, range of average temperatures, absolute latitude, temperature of the coldest month and filariasis were found. Population groups exhibited clustering according to their geographic location. Signals of natural selection were revealed for some genomic regions of studied polymorphic markers. The results demonstrate that genetic diversity in immune-related genes in human populations was partially shaped by adaptation to climate and/or parasitic load driven by natural selection. This work was supported by Russian Foundation for Basic Research (grant # 15-04-02442).

1043T

Evaluation of local ancestry differences in multiple sclerosis among US minority populations. *A. Beecham[1,2], L. Amezcua[3], A. Chinea[4], N. Isobe[5], P. Manrique[1], B.T. Lund[3], A. Levy[3], D. Conti[6], G. Beecham[1,2], P. De Jager[7], S. Delgado[8], J. Oksenberg[5], J. McCauley[1,2].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 3) Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 4) San Juan MS Center Guaynabo, Puerto Rico, USA; 5) Department of Neurology, University of California at San Francisco, San Francisco, CA, USA; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 7) Program in Translational NeuroPsychiatric Genomics, Institute for Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, USA; 8) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple Sclerosis (MS) exhibits variable prevalence across populations, with European populations having a higher prevalence than either Hispanic or African. In addition to environmental influences, this observation could be partly due to a greater genetic risk in European populations. Genetic association studies in individuals of European descent have identified 110 established MS risk variants in 103 discrete loci outside of the Major Histocompatibility Complex. Our goal was both to characterize the local ancestry across these loci in genetically admixed samples from Hispanic and African American populations as well as to use local ancestry to identify novel loci which may be related to MS. Data from the Illumina ImmunoChip were available on Hispanic (190 cases and 46 controls) and African American (973 cases and 257 controls) samples. Local ancestry was computed across 125,830 variants (inclusive of 186 fine-mapping regions for both MS and non-MS loci) with RFMix, after first phasing the haplotypes using Beagle. Reference data from 1000 Genomes (400 from each of the African, Asian, and European populations) were used for both phasing and local ancestry calculation. Separately in Hispanic and African American samples, logistic regression was used to test for the association of both the number of European haplotypes and the number of Asian haplotypes observed at each position, after controlling for global European and Asian ancestry. Global ancestry was computed by averaging local ancestry estimates across ~40,000 independent variants outside of known MS loci. In both Hispanic and African American populations we find evidence ($p \le 0.01$) of European admixture difference across known MS loci between cases and controls. In African Americans, at a threshold of $p \le 0.001$, in addition to the established *CD58* locus, which was previously identified in an admixture scan of African Americans; we saw evidence for a decreased number of European haplotypes in cases at the *NOD2* locus (a well-known Crohn's disease locus; novel to MS). In Hispanics, at a threshold of $p \le 0.001$, we saw an increase of both European and Asian haplotypes at the *NXPE1* locus (an established ulcerative colitis locus; novel to MS). The increase in Asian haplotypes among cases was also observed in African Americans ($p \le 0.05$). These results highlight the value of local ancestry evaluation in admixed populations to better characterize known risk loci and to potentially identify novel MS loci.

1044F

Malaria endemicity and genetic risk of hypertension in Africans. *S.M. Raj[1], M. Hansen[2], F. Schlamp[1], J. Hirbo[2], D. Hu[2], S. Soi[2], L. Scheinfeldt[2,3], A. Ranciaro[2], S. Thompson[2], W. Beggs[2,4], C. Lambert[5], M. Ibrahim[6], T. Nyambo[7], S. Omar[8], C. Wambebe[9], D. Meskel[10], G. Belay[10], A.G. Clark[1], S. Tishkoff[2].* 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, 14853; 2) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA; 3) Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, 19122, USA; 4) Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, 19104, USA; 5) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA; 6) Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan; 7) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 8) Kenya Medical Research Institute, Center for Biotechnology Research and Development, Nairobi, Kenya; 9) International Biomedical Research in Africa, Abuja, Nigeria; 10) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia.

Africans show high incidence and prevalence of the diseases malaria and hypertension. Medical and epidemiological evidence suggests that there may be etiological links between the two diseases. However, these links have not been directly tested at a genetic level. We used population genetics-based approaches to identify correlations between the two diseases in a dataset including 1 million Illumina SNP genotypes in 634 individuals belonging to 39 African populations. First, we used two methods to identify signatures of natural selection, bayenv and hapflk, flagging regions that show high population differentiation as well as correlation with environmental factors. We identified four regions, including one on chromosome 11p15 containing the genes *PPFIBP2* and *CYB5R2*, that showed both high population differentiation as well as a correlation only with malaria endemicity out of a set of twelve environmental variables. SNPs in this region were also associated with elevated blood pressure in published genome-wide association studies. Therefore, we carried out a candidate-gene based association study for SNPs in the region that show malaria-based selection signatures to identify associations with both blood pressure and malaria endemicity. We found a significantly associated SNP ($q<0.02$), with the derived allele corresponding to increased blood pressure. Correlations between SNPs associated with blood pressure and malaria endemicity suggest that genetic variants that play a role in adaptation to malaria may also play a role in risk of hypertension in individuals of recent African origin.

1045W

Genetic contributions to the MHC and LRC regions of a South African population and its effect on TB vaccine efficacy. *M. Salie, H. Schurz, P.D. van Helden, M. Moller, E.G. Hoal.* Biomedical Science, Stellenbosch University, Tygerberg, Western Cape, South Africa.

   Genes in the MHC and LRC regions play important roles in the host immune response to disease and in reproduction, two biological processes essential for the maintenance of the human population. Due to rapid evolution and balancing selection more than 7000 class-I alleles and 400 different KIR profiles have been identified. As populations from geographical regions are under different selective pressures, the frequencies of these loci are known to vary. Differences in disease susceptibility, drug-, and vaccine efficacy between individuals of different ethnicities, highlights the importance of elucidating the genetic profiles of distinct populations. The South African Coloured (SAC) population represents a population with a unique genetic composition with influences from African, Asian and European populations. The global ancestral contributions over chromosomes 6 (MHC) and 19 (LRC) were inferred in the SAC population using three software packages: ADMIXTURE, LAMP-LD and RFMix. LAMP-LD and RFMix was used to infer locus-specific ancestry of the HLA class-I and LRC region. Three reference populations were used: CEU, LWK and JU (Khoisan, Ju/'hoan from Namibia). SAC study participants were genotyped on the Affymetrix 500K Array.We noted differences in the estimated global ancestral contribution to chromosomes 6 and 19 between the different software packages. LAMP-LD inferred a greater proportion of LWK for both chromosomes (44% and 47%), whereas RFMix inferred a greater CEU contribution (34% and 36%). ADMIXTURE showed a greater contribution from the LWK for chromosome 6 (39%) but the CEU for chromosome 19 (37%). Locus-specific analysis noted a larger LWK contribution to the HLA class-I genes, LAMP-LD (>50%) and RFMix (>30%). Conversely, the CEU was noted as having a larger contribution to the LRC region, LAMP-LD (49%) and RFMix (44%). For TB vaccine efficacy we noted that the class-I alleles that bind TB vaccine epitopes Ag85B-ESAT-6, Ag85B-TB10.4 and Mtb72f occurred at low to moderate levels in the South African Black, San and mixed populations. While the SAC population may have a greater proportion of San ancestry based on genome-wide data, the MHC and LRC regions are predominantly of black African or European ancestry, dependent on the software used. Based on class-I allele distribution across SA populations, none of the vaccines currently undergoing clinical trials would be equally efficacious across all SA populations.

1046T

An introgressed Neanderthal haplotype at the alpha-defensin CNV. *X. Xu, J. Armour.* School of Life Science, University of Nottingham, Nottingham, UK.

   The human alpha defensin genes encode an important class of innate immunity effectors. Alpha defensin 1 (*DEFA1*) and 3 (*DEFA3*) are interchangeable gene variants located in a mixed multiallelic copy number variable region (CNV) on 8p23.1. Modern humans have between 3 and 16 copies of *DEFA1/DEFA3*. Recent GWAS studies of IgA nephropathy patients identified two SNPs associated with the disease in the LD block of the *DEFA1A3* CNV. To better understand the origin of modern human diversity at the DEFA1A3 CNV, a phylogenetic analysis of the telomeric and centromeric sequence flanking the CNV was carried out. We identified a putative Neanderthal introgressed haplotype in present-day European populations. Using a molecular clock approach, we found that the upper limit of divergence time between the putative introgressed haplotype and Altai Neanderthal haplotype (220,000 ybp) was more recent than the lower limit of the divergence between other human haplotypes and Neanderthal (670,000 ybp). S* statistics also showed a cluster of derived SNPs in high LD, a characteristic signature of recent introgression. The full length of the introgressed haplotype (~135 kb) was also consistent with Neanderthal introgression. We also demonstrated evidence of introgression inside the *DEFA1A3* CNV repeat. A global distribution of the introgressed haplotype showed that it was mainly distributed in Europe and Indo-European related populations, but absent from African and East Asian populations. We therefore believe this pattern to be a result of a European-specific introgression, after the divergence of Europeans and Asians. The frequency of the introgressed haplotype in European populations is 17.7%, with a higher frequency towards Western Europe. We further evaluated natural selection in this region, and found evidence of positive selection from the integrated Haplotype Score (iHS). The *DEFA1A3* and/or *DEFA5* genes are the likely target of selection. This is the first evidence of archaic introgression reported in a multiallelic copy number variable region, and the potential role of these variants in IgA nephropathy needs to be further evaluated.

1047F

Characterization of *OCA2* gene diversity: Genetic variation and association with skin, hair and eye pigmentation in an admixed Brazilian population. *G. Debortoli[1], L. Marcorin[2], M.L.G. Oliveira[1], A.L.E. Pereira[2], N.C.A. Fracasso[1], J.D. Massaro[3], A.L. Simoes[1], E.A. Donadi[3], E.C. Castelli[4], C.T. Mendes-Junior[2].* 1) Departamento de Genética, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 2) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 3) Departamento de Clínica Médica, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 4) Departamento de Patologia, Universidade Estadual Paulista (FMB-UNESP), Botucatu, SP, Brazil.

The *Oculocutaneous Albinism II* (*OCA2*) gene is located in chromosome region 15q11 and has been shown to be involved in normal variation of skin, eye and hair pigmentation. Notwithstanding, the diversity of *OCA2* gene region has not been studied so far and with the extent proposed by this study in an admixed population, as the Brazilian. Thus, to verify if other variation sites across the *OCA2* gene may be associated with human pigmentation, libraries of all the coding exons, intronic boundaries and regulatory regions were analyzed by next-generation sequencing in a Brazilian admixed population sample. Sample was composed of 339 unrelated individuals from São Paulo State, which were stratified according to eye, hair and skin pigmentation. DNA libraries were prepared using Haloplex Target Enrichment System (Agilent Technologies) and sequenced in MiSeq platform (Illumina). CutAdapt, BWA and GATK packages were used for trimming adapter sequences, alignment and genotype calling, respectively. CADD and SNVrap applications were used to annotate the variations effects on the gene sequences. A total of 143 variation sites were identified. Eighteen of these (with 3 of them being novel) presented strong association with at least one pigmentation feature. When subdividing the variation sites associated with pigmentation traits in upstream and promoter region, 5'UTR, coding region (plus intronic boundaries), and 3'UTR, there were identified 5 variations located near or in the promoter region, 2 in the 5'UTR region, 11 in coding regions (2 in exons and 9 in intronic boundaries), and none in the 3'UTR. Associations found relate mostly to dark skin, both light and dark hair and blue eyes. Some findings corroborates with other literature studies such as rs7497270*T allele being associated with dark phenotypes (black hair in our study, p = 0.0045; OR = 19.78), while some associations linked to different traits in other studies were also found (rs7495174*G associated with red-hair - p = 0.0063; OR = 64.47). Among the 3 novel variation sites reported here, one located in chr15:28,347,113 was associated with dark skin (p = 0.0279; OR = 25.45). In addition, an association that was not previously reported was found between rs200081580*G and blue eyes (p = 0.0116; OR = 42.60). These findings emphasize *OCA2* has an important role in human hair, eye and skin pigmentation. FINANCIAL SUPPORT: FAPESP (Grant 2013/15447-0) and CNPq/Brazil (Grant 448242/2014-1 and Fellowship 309572/2014-2).

1048W

Prevalent homozygous *TGM1* mutation in lamellar ichthyosis patients from consanguineous marriages originating from a clustered population in Northern Malaysia. *Z.A.D. Pramono[1], G.K Oh[1], U. Surana[2], W.C. Tan[3], K.F. Leong[4].* 1) Department of Research, National Skin Centre, Singapore, Singapore; 2) A*STAR Intitute of Cellular and Molecular Biology, Singapore; 3) Department of Dermatology, Hospital Sultanah Bahiyah, Alor Setar, Kedah, Malaysia; 4) Institute of Pediatry, General Hospital, Kuala Lumpur, Malaysia.

Autosomal recessive congenital ichthyosis (ARCI) is heterogeneous group of ichthyosis with 9 causative genes have to date been identified. Large part of ARCI is formed by lamellar ichthyosis that (LI) that mainly is caused by mutations in *TGM1*. We performed genetic analyses for 4 patients with LI from 3 families whereby all are consanguineous marriage originating from the same district in North Malaysia. Mutation analyses were performed by PCR and direct sequencing. Possible large deletions were examined by qPCR. We detected homozygous mutation g1166a (Arg389His) in the TGM1 gene in all 4 patients. None of the patients showed large deletions in the *TGM1* gene by qPCR. The Arg389His has previously been reported in LI patients from the US and Japan. The mutation in the highly conserved residue located in the center of the core domain of TGase 1 peptide, close to the active sites, has been predicted to lead to serious loss of TGase activity. This is the first report on the detection of a homozygous *TGM1* mutation among LI patients from consanguineous marriages in a clustered population in Malaysia. The result of this study highlights the need of public education on the risk of consanguineous marriages. Further study may reveal the prevalence of this mutation in the same ethnicity nationwide as well as the region.

1049T

MITF and ASIP genetic variation and associations with skin, hair and eye pigmentation in an admixed Brazilian population. *C. Mendes-Junior[1], A.L.E. Pereira[1], G. Debortoli[2], M.L.G. Oliveira[2], N.C.A. Fracasso[2], J.D. Massaro[3], A.L. Simoes[2], E.A. Donadi[3], E.C. Castelli[4], L. Marcorin[1].* 1) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 2) Departamento de Genética, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 3) Departamento de Clínica Médica, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 4) Departamento de Patologia, Universidade Estadual Paulista (FMB-UNESP), Botucatu, SP, Brazil.

Microphthalmia associated transcription factor, MITF, and *Agouti signaling protein, ASIP*, are genes involved in the melanin synthesis pathway. *MITF* is known for its regulatory role in many processes like expression of *TYR, TYRP1* and *DCT*, which are directly involved in melanin synthesis in melanosomes. *MITF* expression can be regulated by *MC1R* interaction with α-MSH, stimulating *MITF* expression, or with *ASIP,* as an α-MSH antagonist, preventing *MITF* expression. Although *ASIP*'s rs6058017 have been associated with darker skin and eye phenotypes in Caucasian populations, there aren't known association studies linking *MITF* diversity and pigmentation traits. This study aimed to find genetic variation in *MITF* and *ASIP* associated with pigmentation traits in a Brazilian admixed population sample ($n = 340$) from the state of São Paulo. Samples were stratified regarding eye, hair and skin color and presence of freckles and premature hair greying. Libraries were prepared using Haloplex Target Enrichment System (Agilent Technologies) for MiSeq Illumina sequencing platform. Adapter trimming, alignment and genotype calling were done using cutadapt, BWA and GATK, respectively. Haplotypes and missing alleles were inferred computationally using PHASE, however, the known phase between variable sites (obtained with GATK) was considered. There were found 177 variation sites in the promoter, intronic, regulatory UTRs and coding regions in both genes (55 in *ASIP* and 122 in *MITF*). For *ASIP*, 17 variations (3 novel) are strongly correlated mostly with dark skin, light hair and blue eyes, and corroborates the previous findings regarding *ASIP*'s rs6058017 association with darker skin ($p=0.0001$; OR=10.89), eye ($p=6.02 \times 10^{-6}$; OR=13.81) and hair ($p=0.0011$; OR=20.63). The strongest correlations found were *ASIP* intronic rs7710617 with type VI skin color (Fitzpatrick) ($p=1.19 \times 10^{-9}$ and OR=330.52). For *MITF*, 19 variations (4 novel) are strongly correlated with dark brown eyes, either the most dark or the most light skin and hair colors and the presence of freckles. Besides singletons, the strongest correlations were found for rs74845300 and rs80232752, which are in perfect linkage disequilibrium within *MITF* upstream regulatory region and associated with black hair ($p=0.0002$ and OR=41.10). These results emphasize the role played by *ASIP* and *MITF* variations in human pigmentation. FINANCIAL SUPPORT: FAPESP (Grant 2013/154470) and CNPq/Brazil (Grant 448242/2014-1 and Fellowship 309572/2014 -2).

1050F

Evolution of the 'fused' gene family repeated region across primates. *V. Romero[1], I. Inoue[1], H. Nakaoka[1], K. Hosomichi[2].* 1) Human Genetics, National Institute of Genetics-Japan , Mishima, Shizuoka, Japan; 2) Department of Bioinformatics and Genomics, Kanazawa University, Kanazawa, Japan.

The Epidermal Differentiation Complex (EDC) on chromosome 1 is a dense cluster of genes expressed during the late differentiation of the epidermis and is responsible for maintaining hydration and structure of the skin. The cluster organization of this complex might suggest adaptation during evolution and is strongly associated with a variety of skin disorders. EDC includes the CE precursor family, S100A family and the 'fused' gene family. The 'fused' genes members are: cornulin, filaggrin, filaggrin-2, hornerin, repetin and trichohyalin. These genes have a similar organization showing an initial S100 domain followed by a tandem repeat region and a C-terminal domain. In this study, we focus on the repeat region of each gene and the evolution pathway among primates with short divergence time to fully understand any recent repeat variation. We demonstrated that each member of the 'fused' genes undergo different selective pressure, which suggests a complex model for the adaptive evolution of the EDC. Our results showed that the filaggrin repeat regions within these specific genes have the highest nucleotide diversity between them. The phylogenetic analyses across primates suggest that recent independent duplication events created the repeat sequences of filaggrin in macaque and orangutan, and trichohyalin in marmoset, which are the only ones with phenotype variation related to copy number variation. The comparisons of the repeat sequences detected the signatures of positive selections on filaggrin, filaggrin2, hornerin and repetin within a species and specific codons and branches across species. We suggest that repeat variation across species may be a consequence of non-random events accompanied by species-specific adaptive traits.

1051W

Mutational analysis of causative genes for autosomal recessive spinoc-erebellar degeneration (AR-SCD) to delineate molecular epidemiology of early-onset SCD. *Y. Hama, M. Kanai, Y. Aoshima, Y. Mukai, S. Watanabe, M. Murata, Y. Takahashi.* National Center Hospital, National Center of Neurology and Psychiatry, Kodaira, Japan.

Background: Spinocerebellar degeneration (SCD) is a group of disorders characterized by progressive ataxia caused by dysfunction and atrophy of cerebellum or its afferent or efferent projections. Hereditary SCD is genetically heterogeneous, composed of autosomal dominant (AD), autosomal recessive (AR) or X-linked SCD with more than 50 causative genes. To establish efficient diagnostic algorithms for early-onset SCD excluding AD-SCD, it is imperative to elucidate the molecular epidemiology in these disorders. Purpose: To elucidate the molecular epidemiology of early onset SCD excluding AD-SCD by the mutational analysis of causative genes for AR-SCD. Methods: Subjects with the age of onset under age 40 who were diagnosed as SCD excluding SCA1, SCA2, SCA3/MJD, SCA6, SCA8, SCA12, SCA17, SCA31, DRPLA, ataxia oculomotor apraxia type 2 (AOA2) or ataxia with vitamin E deficiency (AVED) were enrolled in this study. DNA samples obtained with informed consent were subjected to the mutational analysis of the whole exons of *APTX*, *SACS* and *SPG7* employing direct nucleotide sequencing method. This study was approved by the Institutional Review Board of National Center of Neurology and Psychiatry. Results: Among 191 patients with the diagnosis of SCD registered in our cohort, 36 patients consisting of 24 males and 12 females fulfilled the enrollment criteria. The mean age of onset was 10 years old (ranges: 0 - 39). In this cohort, 36 variants were identified. Among them, we identified novel compound heterozygous frameshift variants in *SACS*, which were considered as pathogenic mutations. The patient was thirty five years old female with slowly progressive gait disturbance and dysarthria complicated with cognitive impairment and deafness. Her brain T2-MRI showed low intensity lesions in pons and high intensity lesions in the outer edge of thalamus. All her clinical presentation, neurological findings, and MRI were characteristic of ARSACS. No pathogenic mutations were identified in *APTX* or *SPG7*. Conclusion: This study revealed that *AOA1*, *AR-SACS*, or *SPG7* is not a frequent cause in the early-onset SCD in the Japanese population. Considering the substantial genetic heterogeneity of AR-SCD, exome-first approach would be efficient for genetic diagnosis of early-onset SCD.

1052T

Admixture mapping of established HLA and non-HLA multiple sclerosis risk variants in 24,786 individuals shows evidence for European origin in Hispanics. *C. Chi[1,2], X. Shao[1], B. Rhead[1,2], E. Gonzalez[3], J.B. Smith[3], C. Schaefer[4], L.F. Barcellos[1,2], A.M. Langer-Gould[3,5].* 1) Genetic Epidemiology and Genomics Laboratory, UC Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, UC Berkeley, Berkeley, CA; 3) Kaiser Permanente Southern California, Department of Research & Evaluation; 4) Kaiser Permanente Division of Research, Oakland, CA; 5) Southern California Permanente Medical Group, Los Angeles Medical Center, Neurology Department.

Multiple Sclerosis (MS [MIM 126200]) is a demyelinating autoimmune disease with both genetic and environmental risk factors. MS genetic risk variants are hypothesized to originate in Europeans, but it is unclear whether risk variants exist in other ancestral populations, as previous studies have focused primarily on White populations. In this study, we performed admixture mapping in a diverse group of African American, Hispanic, and non-Hispanic White individuals (2,153 MS cases, 22,633 controls) from Northern and Southern California for whom we have whole genome data and race-ethnicity data through self-report. Preliminary estimates from fastStructure show evidence for case-control differences in global European ancestry proportions in Hispanics. However, ancestry estimates were similar between cases and controls in non-Hispanic White and African-American populations. In Hispanics, an average of 54% and 52% European ancestry was observed from fastStructure in cases and controls, respectively. Local ancestry estimation using LAMP show this global difference was reflected by ancestry estimates for 110 non-HLA MS risk variants established through recent GWAS. Local ancestry for each risk variant was considered together and individually. The local European ancestry proportion between Hispanic cases and controls for the top 20 risk variants had an average difference of 3.09%, most of which revealed greater European ancestry in cases. In contrast, the average difference in all non-Hispanic Whites for the top 20 MS risk variants was only -0.045%. Greater European ancestry was not observed in African-American cases nor controls for the majority of MS risk variants. Local ancestry estimates were also derived for the major histocompatibility complex (MHC) region overall and separately in class I, II and III regions. A similar trend in Hispanics was most notably observed in the MHC class I region, with an average of 47% and 44% European ancestry in cases and controls, respectively. The result that Hispanics exhibit greater European ancestry in cases versus controls supports the hypothesis that MS risk variants are primarily of European origin. However, the absence of this difference in African Americans suggests the possibility of undiscovered MS risk variants in other ancestral populations. Further work is underway to fully characterize the ancestral contributions of each candidate region, and to identify new risk variants through whole genome admixture mapping.

1053F

Language impairment in ASD and schizophrenia resulting from opposite failed domestication itineraries of the human brain. *A. Benítez-Burraco[1], E. Murphy[2], W. Lattanzi[3].* 1) Philology, University of Huelva, Huelva, Spain; 2) Division of Psychology and Language Sciences, University College London, London, United Kingdom; 3) Institute of Anatomy and Cell Biology, Università Cattolica del Sacro Cuore, Rome, Italy.

Schizophrenia (SZ) and autism spectrum disorders (ASD) are pervasive, highly prevalent neurodevelopmental disorders entailing social and cognitive deficits, including marked language deficits. Numerous genes have been associated with both conditions, but it is not clear how language deficits arise from gene mutation or gene dysregulation. In this study we have adopted an evo-devo approach to address this issue. On the devo side, we will build on the hypothesis that SZ and ASD represent opposite poles within a developmental continuum of human cognition, also encompassing the normal faculty of language. On the evo side, we have built on the hypothesis that the emergence of a modern faculty of language was brought about by changes in the human brain/skull, but also by the self-domestication of the human species. We show that people with ASD exhibit less marked domesticated traits at the morphological, physiological, and behavioural levels, whereas people with SZ show more marked domesticated traits. We also show that many candidates for both conditions are represented among the genes known to be involved in the "domestication syndrome" (the constellation of traits exhibited by domesticated mammals, which seemingly results from the hypofunction of the neural crest), and among the set of genes involved in language function that interact with them. We have performed comparative *in silico* analysis of genome-wide brain expression datasets, from humans and chimps. This allowed us to demonstrate that several of these genes exhibit altered expression in the brains of ASD and SZ patients. Interestingly, some candidates for domestication and language-readiness exhibited similar expression trends between ASD patients and chimps, while opposite trends between SZ patients and chimps, in brain areas involved in language processing. We conclude that ASD and SZ (and specifically, language deficits in each condition) may represent abnormal ontogenetic itineraries for the human faculty of language, resulting in part from opposite changes in genes important for the "domestication syndrome" and, ultimately, from the normal functioning of the neural crest.

1054W

Polygenic adaptation and how it may explain the 'evolutionary paradox'. *P.F. O'Reilly, E. Vassos, J. Maxwell, C.M. Lewis.* MRC SGDP Centre, King's College London, London, United Kingdom.

Tests for detecting natural selection in the human genome have typically focused on identifying signatures of selection at individual loci. This implicitly assumes alleles of sufficiently large phenotypic effect to leave a discernible trace in contemporary genetic data. However, GWAS have demonstrated that most human phenotypes are governed by many hundreds or even thousands of variants of tiny effect. Therefore, the mark left on the genome by selection will often be polygenic and too subtle to detect at any one locus. Using the latest publicly available GWAS we demonstrate that the risk allele frequencies for numerous diseases and psychiatric disorders are systematically lower than expected by chance, while other traits associated with increased fecundity, such as height, have systematically higher than expected frequency. Next we investigate how risk alleles for psychiatric disorders associated with particularly low fecundity may be maintained in the population (the 'evolutionary paradox'). Using polygenic risk scores and UK Biobank data, we show that having an increasing genetic load for disorders such as schizophrenia and autism is in fact beneficial for a number of behavioural traits, such as cognitive ability, risk taking and well-being - until an extreme genetic load is reached, which produces negative outcomes. Thus whether an allele is beneficial or deleterious is dependent on the overall genetic load relating to a phenotype in the individual, rather than being a fixed feature of the allele.

1055T

Association of sleep disorders with neuropeptide receptor genes polymorphism in male population in Russia/Siberia. *V. Gafarov, E. Gromova, M. Voevoda, I. Gagulin, V. Maximov, A. Gafarova, D. Panov.* FSBI Institute of Internal and Preventive Medicine, NOVOSIBIRSK, NOVOSIBIRSK, Russian Federation.

Objective: To study the association gene of candidate NPSR1 rs324981 with sleep disorders in the open population of men 45-64 years of Novosibirsk. Methods: The study of the association candidate gene polymorphisms with sleep disorders was carried out during the examination of a random representative sample of men 45-69 years (n = 1770). The response rate was 61%. The median age is 56.5 year. Every 12 subject was selected for genotyping (n = 147). To assess the level of sleep was used a questionnaire which was filled with self-test. Statistical analysis was performed using SPSS-11.5. Results: The level of sleep disorders in the male population of 45-64 years was 79.9%. The frequency of homozygous C / C genotype of neuropeptide S (gene NPSR1 rs324981) was 19.4%, T / T genotype occurs in 27.8%, C / T genotype - 52.8%. Men dominated the T allele of -54.2%, and the C allele - 45.8% growth trend Fnd dissatisfaction with the quality of their sleep among men. Men T- allele carriers, most evaluated their sleep as "satisfactory" in 69% of cases, ($\chi2$ = 15,713 df = 8, p <0.05). Conclusion Association found men carrier T - allele of neuropeptide S (gene NPSR1 rs324981), a sleep disorder. Supported by Grant of Russian Foundation for Humanities No14-06-00227/a.

1056F

Assessing population genetic evidence of selection across neuropsychiatric phenotypes. *E.R. Beiter[1], E.A. Khramtsova[2], C. Simonti[3], S. Dalvie[4], C. Van Der Merwe[4], D. Stein[4], J.A. Capra[3], J. Knowles[5], L. Davis[6], B. Stranger[2].* 1) Washington University, Saint Louis, MO; 2) Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL; 3) Vanderbilt Genetics Institute; Department of Biological Sciences, Vanderbilt University, Nashville, TN; 4) Department of Psychiatry, Faculty of Health Sciences, University of Cape Town; 5) Department of Psychiatry and The Behavioral Sciences, Univesity of Southern California; 6) Vanderbilt Genetics Institute; Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN.

Variation in neuropsychiatric traits is present across diverse human populations, has persisted across recorded history, and has a genetic basis primarily accounted for by common (minor allele frequency > 5%) single nucleotide polymorphisms (SNPs). Given recent observations that polygenic risk for neuropsychiatric phenotypes is correlated to many non-neuropsychiatric complex traits (Bulik-Sullivan et al., 2015), and that SNPs with high minor allele frequency (MAF) contribute disproportionately to some neuropsychiatric phenotypes (Davis et al., 2013), we were interested in systematically characterizing signatures of positive selection across nine neuropsychiatric traits. We performed a multi-level analysis using genome-wide association study summary statistics provided by the Psychiatric Genomics Consortium to detect evidence of 1) extreme integrated haplotype scores (iHS > 2.0) demonstrating recent strong positive selection (i.e., hard sweeps), 2) population differentiation (F_{ST} > 0.3), 3) polygenic selection, and 4) enrichment of trait-associated alleles in regions intolerant to Neanderthal introgression (S-score < 5%), among SNPs nominally associated with each trait (p < 5.0 x 10[-3]). For each set of trait-associated SNPs, we used SNP-SNAP tool to generate 500 sets of null SNPs matched on minor allele frequency (MAF), distance to nearest gene, and gene density, then compared the proportion of trait-associated SNPs with extreme iHS, F_{ST} and S-scores to their respective matched null distributions. Consistent with expectations for polygenic phenotypes, no trait displayed significant enrichment of population differentiated SNPs or strong recent positive selection. However, we did identify enrichment of risk alleles for schizophrenia (p < 0.002) within regions of the genome intolerant to Neanderthal introgression. Finally, a recently published method (Berg and Coop, 2014) for detecting polygenic selection was also applied to each phenotype. The polygenic adaptation approach detects coordinated and systematic shifts in MAF of trait-associated SNPs while accounting for genetic drift measured across 52 populations. Taken together, our results suggest that individual common risk alleles for neuropsychiatric phenotypes are unlikely to have been the targets of strong recent positive selection (i.e., hard sweeps).

1057W

An investigation of the evolution and worldwide distribution of the READ1 element in *DCDC2* **and its contribution to language development.** *M.M.C. DeMille[1], J.C Frijters[2], K. Tang[4], C. Geissler[4], N.R Powers[1], B.M. Bowen[3], A.K. Adams[3], D.T. Truong[1], J.R. Gruen[1,3].* 1) Department of Pediatrics, Yale University School of Medicine, New Haven, CT; 2) Child and Youth Studies, Brock University, St. Catharines, ON, Canada; 3) Department of Genetics, Yale University School of Medicine, New Haven, CT; 4) Department of Linguistics, Yale University, New Haven, CT.

The DYX2 locus (6p21.3), containing dyslexia candidate genes *DCDC2* and *KIAA0319*, is the most replicated locus for reading disability (RD) and language impairment (LI) worldwide. READ1, located in intron 2 of *DCDC2*, is a multi-unit transcriptional control element with approximately 40 known alleles. Previous research has shown that repeat unit (RU) length groupings of alleles of READ1 (RU1-1 and RU2-long) are associated with RD. We have also demonstrated epistasis between READ1 alleles and a RD risk haplotype spanning *KIAA0319*. Furthermore, an *in vitro* reporter gene assay demonstrated that READ1 alleles alter expression from the *DCDC2* promoter and chromatin conformation capture indicates that READ1 interacts physically with a region upstream to the *KIAA0319* promoter *in vivo*, suggesting regulation of *KIAA0319* expression as the mechanism of epistasis between READ1 and *KIAA0319*. Taken together, READ1 in *DCDC2* and *KIAA0319* are strongly implicated in RD, with functional studies elucidating a potential molecular mechanism in its etiology. We have investigated the evolution and worldwide distribution of the READ1 element, with an eye toward language development. While *DCDC2* was identified as a RD gene, it may also influence normal variation in language. First, we estimated the mutational history of the READ1 element by examining primate and archaic hominin sequences. We found that the duplication of RU1 occurred after the split between chimps and humans. Additionally, we found that RU2 did not have greater than 4 repeats until after that split. We also examined the worldwide distribution of READ1 alleles in 58 populations among five different continental groups and discovered correlations between the allele frequencies of READ1 and several attributes of the languages used in a given culture: phoneme number (RU1-1, $p = .006$), consonants (RU1-1, $p = .0151$ and RU2-short, $p = .0179$), tones (RU1-1, $p =.0084$), and ratio of consonants to vowels (RU1-1, $p = .0091$; RU2-long, $p = .0056$; rs9461045T, $p = .0104$). We also showed a significant interaction between the SNP rs9461045 (a SNP in the *KIAA0319* risk haplotype) and READ1 alleles in this correlation. The higher the proportion of RU2-long in the population, the stronger the negative relationship between the derived allele T of SNP rs9461045 and the number of consonants in a population. We hypothesize that alleles of READ1 may have influenced the type of written and spoken language developed by a given culture.

1058T

Doubts and wavering in mitochondrial DNA forensics: While interpreting length heteroplasmy conundrum of various Sindhi and Baluchi ethnic groups of Pakistan. *S. Bhatti[1], M. Aslamkhan[2], S. Abbas[3], M. Attimonelli[4], H.K. Aydin[5], E.M.S. de Souza[6].* 1) Department of Human genetic and Molecular biology, University of Health Sciences, Lahore, Punjab, Pakistan; 2) Department of Human Genetics and Molecular biology, University of Health Sciences Lahore-54600, Pakistan; 3) Institute of Molecular Biology and Biotechnology, University of Lahore-54600, Pakistan; 4) Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, Bari, Italy; 5) Department of Medical Biochemistry, Ege University School of Medicine, Bornova Izmir, Turkey; 6) Instituto Nacional de Pesquisa, Manaus Programa de Pós Graduação em Genética, Conservação e Biologia Evolutiva. Instituto Nacional de Pesquisas da Amazônia Av. André Araújo, 2936, Aleixo, Manaus, AM 69060-001, Brazil.

The insight heterodox genetics of mtDNA provides new perspectives at the level of human mitochondrial control region heteroplasmy, which is significant in evolutionary as well as forensic interpretation. The main goal of this study is to interrogate the recurrence and resolve the ambiguity of blurry spectrum of heteroplasmy in the human mtDNA control region of 50 Baluchi and 116 Sindhi unrelated individuals. Sanger sequencing was used classically, which was further investigated by minisequencing. Only 20% Baluchi and 25.8% Sindhi were homoplasmic, whereas rest of 80 % Baluchi and 74.1% Sindhi showed at least one heteroplasmy within the specimen. In total 166 individuals have length heteroplasmy (LH) found to be at position 16189, 303-315, 568-573, and 514-524, whilst point mutation heteroplasmy (PMH) was detected at position 73, 16093, 16189 and 16234 respectively. Overall LH was observed albeit high frequency in Sindhi ethnic group (82%) rather than Baluchi's (37%), whereas PMH accumulation relatively more (24%) in Baluchi ethnic groups followed by low percentage (11.2%) observed in Sindhi ethnic groups. The obtained results underline the necessity that growing knowledge of heteroplasmy, helps to develop a consciences in the forensic community that heteroplasmy plays a pivotal role in the forensic interpretation on a regular basis and knowledge of its biological underpinnings has a vital niche in the forensic sciences. Limited studies has focused on heteroplasmy, yet scientific attention should be given, in order to determine its magnitude in different ethnic boundaries. **Key words:** Mitochondrial DNA heteroplasmy, Control region, Sanger sequencing, Baluchistan, Sindh, Pakistan.

1059F

p.P347Q Rhodopsin mutation is a recurrent founder allele in Brazilian individuals with autosomal dominant retinitis pigmentosa. *E.M. Jones[1,2], F.B.O. Porto[3,4], Z. Soens[1,2], R.T. Simoes[5], Y. Li[1,2], R. Chen[1,2,6,7,8].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) INRET - Clínica e Centro de Pesquisa, Belo Horizonte, Minas Gerais, Brazil; 4) Departamento de Retina e Vítreo, Centro Oftalmológico de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; 5) Instituto de Ensino e Pesquisa da Santa Casa de Belo Horizonte, Belo Horizonte, Minas Gerais, Brazil; 6) Structural and Computational Biology and Molecular Biophysics Graduate Program, Baylor College of Medicine, Houston, TX; 7) The Verna and Marrs Mclean Department of Molecular Biology, Baylor College of Medicine, Houston, TX; 8) Program in Developmental Biology, Baylor College of Medicine, Houston, TX.

   PURPOSE: The p.P347Q *RHO* mutation is responsible for almost 30% of the autosomal dominant Retinitis Pigmentosa (adRP) in Minas Gerais (MG) – Brazil. The purpose of this study was to describe: the segregation of the p.P347Q *RHO* mutation in 6 unrelated families from the small isolated city of Tabuleiro-MG; the haplotype reconstruction; and the clinical characteristics of the p.P347Q allele. **METHODS:** Visual acuity, retinography, autofluorescense and OCT were performed on 23 patients, and full-field ERG on 6 patients. Blood samples were collected on 23 patients (17 affected and 6 non-affected, ranging from 2 to 61 years old). Three probands were sequenced using a targeted capture panel including 224 known retinal disease genes. Next-generation sequencing (NGS) data was then aligned, processed and filtered against control datasets using an in-house custom pipeline. Identification of p.P347Q from NGS data in each of the unrelated probands was validated through Sanger sequencing. DNA samples of the probands' parents and additional unrelated individuals linked to Tabuleiro were then collected and Sanger sequenced for the presence of p.P347Q. Informative captured SNPs were used to infer haplotypes across each of these individuals. **RESULTS:** A C to A transversion at position 1040 (p.P347Q) of the cDNA has been identified in all 12 RP individuals currently sequenced for the variant. All individuals with this mutation had early-onset night blindness and substantial visual field losses as teenagers. All patients who underwent full field ERG testing demonstrated non-recordable ERGs. Autofluorescence retinography showed maculopathy and a hipoautofluorescent ring around the optic disc in all patients. Lastly, on the basis of conserved haplotypes linked to this pathogenic variant, we propose that p.P347Q spread through a cohort of geographically isolated families by a preceding founder effect. **CONCLUSIONS:** The p.P347Q allele represents a founder mutation causing a rapidly progressive form of adRP in our distinct population. The identification of this founder effect and subsequent mutation may greatly simplify the molecular diagnosis of relevant disease in this population with important implications for the speed and cost-efficiency of diagnostic screening and prognostic counseling.

1060W

***RUNX2* and skull evolution: Insights from ancient DNA and gene expression in calvarial sutures of nonsyndromic cranisynostosis patients.** *W. Lattanzi[1], L. Di Pietro[1], M. Barba[1], P. Frassanito[2], M. Baranzini[1], L. Massimi[2], G. Tamburrini[2], A. Benítez-Burraco[3].* 1) Institute of Anatomy and Cell Biology, Università Cattolica del Sacro Cuore, Rome, Italy; 2) Institute of Neurosurgery, Università Cattolica del Sacro Cuore, Rome, Italy; 3) Department of Spanish Philology and its Didactics, Universidad de Huelva, Spain.

   The concerted skull-brain evolution occurred through a modification in the rate and timing of neurocranium ossification, specifically in the frontal region, with delayed ossification of the metopic suture being a key feature. The *RUNX2* gene, encoding a master transcription factor and downstream effector of most signaling pathways orchestrating craniofacial development, played a key role in this process. Signs of positive selection in the modern human variant of the gene have been demonstrated, by deciphering ancient hominin genomes*;* also, human and primate RUNX2 proteins show significant aminoacid changes. In humans, *RUNX2* mutations are responsible for craniofacial phenotypes including metopic synostosis. The aim of this study is to assess whether *RUNX2* functional status changed during evolution, as being involved in site-specific suture patterning that modified skull shape over time. Bioinformatic tools were used to align and compare the *RUNX2* gene in Neanderthal, Denisovan and modern humans (AMH), and map all the sequence changes that have occurred during recent evolution. The functional effect of the detected changes, with regard to alternative isoforms' transcription and gene expression regulation, was predicted *in silico*. Thereafter, *RUNX2* isoform-specific expression was analyzed in suture specimens collected from nonsyndromic craniosynostosis patients. We identified > 570 fixed changes within noncoding regions of the *RUNX2* gene, including nucleotide substitutions, indels, and annotated human SNPs; these were mostly overlapping in AMH-versus Neanderthal and Denisovan genes. Selected changes mapped within the two promoters of the gene, although they apparently do not affect isoforms' expression, based on *in silico* prediction. In addition, changes in the 3'UTR were detected that were putatively able to affect miRNA binding. Additionally, we found that the expression of *RUNX2* isoforms increased in fused coronal, lambdoid and metopic, but not sagittal, sutures compared with site-matched unfused sutures. Interestingly, by comparing suture sites, we detected an isoform-specific decrease of *RUNX2* expression in metopic versus other suture sites. These results allowed confirming a specific role of *RUNX2* in determining the changes in skull morphogenesis during AMH evolution. This should reasonably affect also brain and cognitive abilities, given that *RUNX2* is functionally related to genes important for brain development and cognitive function, including language.

1061T

Genes that are routinely subject to inactivation and genes that routinely escape inactivation are highly expressed, and are candidates in Turner syndrome phenotype. *S. Brotman[1], M. Wilson Sayres[1,2].* 1) School of Life Sciences, Arizona State University, Tempe, AZ; 2) Center for Evolution and Medicine, The Biodesign Institute, Tempe, AZ.

Individuals with Turner syndrome (45,X) exhibit wide variation in phenotypes. Differential dosage of genes on the X chromosome likely contributes to this phenotypic variation due to the partial or complete loss of one X chromosome in individuals with Turner syndrome. The process of X-inactivation silences expression of genes on all X chromosomes in excess of one. However, approximately 15% of the genes located on the silenced X chromosome(s) escape inactivation. This allows individuals with a typical (46,XX) karyotype to have expression of two copies of a gene, whereas individuals with Turner syndrome will only have expression from one copy, making the subset of escaping genes likely candidates for affecting the phenotype of people with Turner syndrome. In this study, we use an evolutionary approach to identify candidate genes that may contribute to the variation in phenotype among individuals with Turner syndrome. We confirm previous results that methylation pattern differs between genes with different inactivation statuses, in individuals with two X chromosomes, but not in individuals with Turner syndrome. We then jointly analyze estimates of variable X-inactivation status across multiple 46,XX individuals alongside patterns of gene expression on the X chromosome across five tissues in ten species. Our results suggest that genes that are subject to inactivation in all individuals and genes that escape inactivation in all individuals show high levels of expression (whereas genes that are heterogeneously inactivated across individuals are typically lowly expressed). This is consistent with selection acting on these two classes of dosage sensitive genes on the X chromosome. We expect variation in these genes to affect phenotype when the dosage is altered from normal levels.

1062F

Medical Genome Reference Bank – A comprehensive and integrated genome/phenome resource generated from a disease deplete, elderly population. *M.E Dinger[1], A. Stone[1], W. Kaplan[1], P. Lacaze[2], R.L. Woods[2], J.J. McNeil[2], S. Redman[3], B. Armstrong[3], D.M. Thomas[1].* 1) Garvan Institute, Sydney, NSW, Australia; 2) Monash University, Melbourne, VIC, Australia; 3) Sax Institute, Sydney, NSW, Australia.

The Medical Genome Reference Bank (MGRB) will house whole genome sequencing (WGS) data and rich phenotypic information from 4,000 healthy Australians over 70 years of age. The reference bank will act as a powerful filter to distinguish between causal and passenger genetic variation, and will be a resource to maximise the efficiency of disease-specific genomic analyses in both the research and clinical setting. MGRB participants, consented through contributing studies, 45 and Up (Sax Institute, Sydney), and the ASPirin in Reducing Events in the Elderly (ASPREE) clinical trial (Monash University, Melbourne), are free from cardiovascular disease, degenerative neurological disorders and of a history of a cancer. WGS is performed on the Illumina HiSeq X-Ten platform at the Garvan Institute of Medical Research (Sydney, Australia) under clinically accredited conditions (ISO 15189). Data is aligned and variant files are generated using best practice (GATK, BWA) pipelines to make results interoperable with other cohorts. Joint-called variants are loaded into an analytical framework (developed in partnership with Genome England) to enable the interrogation of genomic variants together with advanced filters based on clinical traits and genomic annotations. Curated data will be openly accessible to the international research community through an MGRB website. Preliminary features will include a data Beacon, extensive variant annotation, complex population-based clustering queries, visualisation of variant data (e.g. genome viewer/ gene networks) and analysis tools for assessing the genetic burden of individual variants and variant subsets. While basic demographic and phenotypic information will be incorporated into the MGRB data portal, researchers are invited to apply for access to comprehensive genotypic and clinical information to support high-level integrative analysis. The MGRB data portal is scheduled for launch in late 2016, providing researchers access to data from 1,500 individuals in the first instance; increasing to 4,000 individuals throughout 2017.

1063W

CGHD: A database with a user-friendly interface for genome-wide association study by whole-genome sequencing and genome-wide genotyping microarray of the Taiwanese population. *W. Fan[1], S. Su[1,2], C. Wang[2,3], W. Chung[1,2,3,4].* 1) Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan; 2) Chang Gung Immunology Consortium, Chang Gung Memorial Hospital and Chang Gung University, Taoyuan, Taiwan; 3) Department of Dermatology, Drug Hypersensitivity Clinical and Research Center, Chang Gung Memorial Hospital, Taipei, Linko and Keelung, Taiwan; 4) School of Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan.

The Chang Gung Memorial Hospital reported the construction of a Taiwanese population reference database (CGHD; Chang Gung Human Database) including 250 whole-genome sequencing and 756 genome-wide genotyping microarray of healthy Taiwanese individuals. Here we identify through this high-coverage sequencing (31.7-fold on average), 15.7 million single-nucleotide variants (SNVs) and 2.6 small insertions and deletions (small Indels, size ≤ 50 base pairs). We also catalogued structural variants (SVs), including 1.4 million insertions and deletions, and 6,962 copy-number variants (CNVs). The CGHD was effective for imputing genotypes of the Taiwanese population genome wide. These data demonstrate the value of high-coverage sequencing for constructing population-specific variant panels and for identifying causal rare variants of complex human disease phenotypes in genetic association studies, such as phenytoin-related severe cutaneous adverse reactions. In addition, we build a webserver with a user-friendly interface for searching and analyzing variant with the information of gene structures, CGHD, 1000g, Ex6500, ploy-phen2 and so on. The availability of such in-house database will form a solid foundation for optimizing the quality of genomic research and test conducted in Taiwan.

1064T

Testing differential genetic flow across Sinai and Bab-el-Mandeb. *D. Platt[1], L. Parida[1], S. Basu[2], P. Zalloua[3,4].* 1) IBM T. J. Watson Research Center, Yorktown Hgts, NY; 2) Department of Mathematics, Purdue University, West Lafayette, IN; 3) Lebanese American University, School of Medicine, Beirut 1102 2801, Lebanon; 4) Harvard School of Public Health, Boston, MA, USA.

It has been established that SNP evidence suggests that the dominant out-of-Africa movement occurred across the Sinai rather than Bab-el-Mandeb, even though recombination event data suggests stronger admixture across Bab-el-Mandeb. Subsequent gene flow could have occurred at either or both, especially as trade routes opened up carrying spices, silk, and incense, along with water-borne commerce on the Red Sea, which supported rich trading empires in Djibouti and Yemen. These activities make the region an excellent testing ground to identify broader-based genetic lineages between populations. PCA extracts correlated structure in data allowing for identification of a small number of dimensions that capture most of the detail of Euclidean distances. If those are genetic distances, marking populations differentiated by accumulated mutations in isolation, then a large portion of evolutionary history can be identified in a small number of components. However, PCA is insensitive to structure on the thin axes, e.g. smaller numbers of alleles that emerged recently that distinguish genetically similar populations. Fisher's linear discriminant with Rao's extension provide a way to identify differentiation along those thin axes, as well as p-values via random permutations in a general ANOVA decomposition framework. Such frameworks include nested hierarchic structures, such as AMOVA. We applied multiple modes of analysis using 1087 samples from SW Asia and Africa. For simple one-way ANOVA decomposition, we show that Linear Discriminant Analysis (LDA) yields stronger differentiation compared to chance via random permutations of assignments) for regional coarser grained groups rather than finer grained national groups. Further we show much more homogeneity across the Sinai (more homogenous than chance, p-value = 0.0302) rather than Bab-el-Mandeb (less homogenous than by chance, p-value=$6.29 \cdot 10^{-24}$), contrasting the Saudi Peninsula with the connecting African regions. We report and contrast these results with other ANOVA decompositions on LDA, on computational homology, and with Patterson et al's f3 and f4 statistic results. Our data supports the conclusion that the dominant genetic flow occurred over Sinai, with subsequent admixture primarily by that route.

1065F

ACTN3 R577X and ACE I/D gene variants influence performance in elite sprinters: A multi-cohort study. *I.D. Papadimitriou[1], A. Lucia[2], Y.P. Pitsiladis[3], V.P. Pushkarev[4], D.A. Dyatlov[5], E.F. Orekhov[5], G.G. Artioli[6], J.P.L.F. Guilherme[6], A.H. Lancha Jr[6], V. Ginevičienė[7], P. Cieszczyk[8], A.M. Karlowska[8], M. Sawczuk[8], C.A. Muniesa[9], A. Kouvatsi[10], M. Massidda[11], C.M. Calò[11], F. Garton[15], P.J. Houweling[15], G. Wang[3], K. Austin[3], A.M. Druzhevskaya[12], I.V. Astratenkova[13], I.I Ahmetov[14], D.J. Bishop[1], K.N. North[15], N. Eynon[1].* 1) Institute of Sport, Exercise and Active Living (ISEAL), Victoria University, Victoria 8001, Australia; 2) Universidad Europea and Research Institute Hospital 12 de Octubre, Madrid, Spain; 3) FIMS Reference Collaborating Centre of Sports Medicine for Anti-Doping Research, University of Brighton, Brighton, UK; 4) Clinical Physiology Laboratory, Urals Research Centre for Radiation Medicine, Chelyabinsk, Russia; 5) Ural State University of Physical Culture, Chelyabinsk, Russia; 6) School of Physical Education and Sport, University of Sao Paulo, Sao Paulo, Brazil; 7) Department of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania; 8) Department of Physical Culture and Health Promotion, University of Szczecin, Szczecin, Poland; 9) Faculty of Physical Activity, Universidad Europea de Madrid, Alcobendas, Spain; 10) Department of Genetics, Development and Molecular Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece; 11) Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy; 12) Sports Genetics Laboratory, St Petersburg Research Institute of Physical Culture, St. Petersburg, Russia; 13) Department of Physiology, St Petersburg State University, St. Petersburg, Russia; 14) Sport Technology Research Centre, Volga Region State Academy of Physical Culture, Sport and Tourism, Kazan, Russia; 15) Murdoch Childrens Research Institute, Melbourne, Australia.

To date, studies investigating the association between ACTN3 R577X and ACE I/D gene variants and elite sprint/power performance have been limited by small cohorts from mixed sport disciplines, without quantitative measures of performance. To examine the association between these variants and sprint time in elite athletes we collected a total of 555 best personal 100-, 200-, and 400-m times of 346 elite sprinters in a large cohort of elite Caucasian or African origin sprinters from 10 different countries. Sprinters were genotyped for ACTN3 R577X and ACE ID variants. On average, male Caucasian sprinters with the ACTN3 577RR or the ACE DD genotype had faster best 200-m sprint time than their 577XX (21.19 ± 0.53 s vs. 21.86 ± 0.54 s, p=0.016) and ACE II (21.33 ± 0.56 vs. 21.93 ± 0.67 sec, p=0.004) counterparts and only one case of ACE II, and no cases of ACTN3 577XX, had a faster 200-m time than the 2012 London Olympics qualifying (vs. 12 qualified sprinters with 577RR or 577RX genotype). Caucasian sprinters with the ACE DD genotype had faster best 400-m sprint time than their ACE II counterparts (46.94 ± 1.19 s vs. 48.50 ± 1.07 s, p=0.003). Using genetic models we found that the ACTN3 577R allele and ACE D allele dominant model account for 0.92 % and 1.48 % of sprint time variance, respectively. Despite sprint performance relying on many gene variants and environment, the % sprint time variance explained by ACE and ACTN3 is substantial at the elite level and might be the difference between a world record and only making the final.

1066W

Blood nitric oxide regulation and genetic adaptation of Tibetans to high altitude hypoxia. *Y. He[1,4], C. Cui[2], X. Qi[1], W. Ouzhuluobu[2], Y. Peng[1], H. Zhang[1], Y. He[1,4], W. Yangla[2], W. Dejiquzong[2], W. Baimakangzhuo[2], W. Duojizhuoma[2], Y. Pang[2], W. Cirenyangji[2], W. Baimayangji[2], W. Gonggalanzi[2], C. Bai[2], W. Bianba[2], T. Wu[3], B. Su[1,2].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 2) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa 850000, China; 3) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining 810012, China; 4) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100101, China.

Tibetans are indigenous highlanders who are well adapted to high-altitude hypoxic conditions. As an important vasodilator, nitric oxide(;NO);has been highlighted in previous studies with a potential role in hypoxic adaptation of Tibetans. It was shown that highlander Tibetans exhibited greatly elevated levels of circulating NO products when compared to lowlanders. However, the observed higher blood NO concentration of high-altitude Tibetans could be caused merely by altitude difference because the previous report did not test NO levels in lowlanders living at high altitude. In this study, using ozone-based chemiluminescence, we measured blood NO concentration of 1,581 individuals, including 335 Han and 520 Tibetans living in Lhasa (elevation: 3,680m), 507 Tibetans living at 4,700m and 219 Tibetans living at 5,370m. We found that consistent with previous report, high-altitude Tibetans have a higher NO level compared to lowlanders (OR=1.87, p=0.0025). But surprisingly, Han residents in Lhasa exhibited a significantly higher NO level than Lhasa Tibetans (OR=2.35, p<0.0001). At higher altitudes (4,700m and 5,370m), the NO levels of Tibetans are lower than the level in Lhasa (p<0.0001). These data suggest that the elevation of blood NO level at high altitude is not specific to Tibetans. Rather, Tibetans showed a blunted response to NO elevation under hypoxia. We propose that similar with the regulation of hemoglobin, the regulation of blood NO plays a role in both acclimatization of lowlanders and genetic adaptation of Tibetans to high altitude hypoxia.

1067T

Geographic and historic changes in runs of homozygosity among more than 1,000,000 individuals sheds light into the recent demographic history of US population. *A. Kermany, C. Ball, J. Byrnes, P. Carbonetto, K. Chahine, R. Curtis, E. Elyashiv, J. Granka, H. Guturu, E. Han, E. Hong, N. Myres, K. Noto, K. Rand, Y. Wang.* Ancestry.com DNA, LLC, San Francisco, CA.

Runs of Homozygosity (ROH) are indicators of segments of chromosomes identical by descent between parental haplotypes. Distribution of such runs along the chromosome contains information regarding the demographic history of the population under study, in particular it reveals trends in consanguinity. In this study, we analyze the distribution of runs of homozygosity – chromosomal locations, number of runs and lengths of runs - as well as estimated inbreeding coefficient (F) among more than 1,000,000 consented AncestryDNA customers. We report on observed variations in distribution of ROH based on geographic origins - inferred from the available pedigree data - admixture proportions as well as birth year cohort. In particular, we present our results on variations in the distribution of ROH within 19 communities within the US population - identified based on analysing a network of genetic matches in the database - and investigate differences in patterns of ROH between each group and comment on the inferred demographic history within each group.

1068F

Nanopore sequencing reveals evolution of poxvirus genome structure driven by conflict with human defense mechanisms. *T. Sasani, K. Cone, R. Layer, N. Elde, A. Quinlan.* Department of Human Genetics, University of Utah, Salt Lake City, UT.

Viruses are locked in conflict with host organisms and rapidly adapt to combat antiviral host responses. To prevent viral propagation, human immune defense pathways involve a number of host proteins that work in concert. However, the Vaccinia (VACV) genome encodes two host-range factors, E3L and K3L, that each disrupt key host defenses. In the absence of E3L, VACV has been shown to rapidly adapt by duplicating K3L in tandem arrays, which confers a significant increase in fitness (Elde et al. 2012). Additionally, the appearance of an H47R mutation in a single copy of the K3L gene confers a fitness benefit comparable to duplication of wild-type K3L.The precise nature and composition of these duplications has proven difficult to characterize, primarily due to the fact that short-read sequencing technologies are unable to sequence through arrays of K3L duplications (measuring ~500 bp per copy). In this study, we utilized Oxford Nanopore (ONT) long-read sequencing technology to characterize VACV populations (lacking the E3L gene) that survived selective pressure during successive passages in HeLa cells. Using this approach, we detected individual viral genomes harboring up to 24 tandem copies of K3L, and find that the distribution of K3L copy number in VACV populations is highly heterogeneous. Also, by analyzing individual reads that contain K3L duplications, we determined that the H47R mutation is not uniformly present or absent in tandem arrays of K3L. Rather, we find that the mutation is lost and gained in subsequent duplications, perhaps due to recombination between viral genomes. For the first time, using ONT sequencing, we have uncovered the interplay between allelic diversity and structural variation during viral evolution.Accurate representation of variation in structurally unstable genomic regions is critical to understanding the genetic basis of human disease. For example, structurally labile chromosomal regions (e.g., 17p12) yield complex rearrangements driven by non-allelic homologous recombination and underlie developmental disorders. Highly variable repeat expansions within *C9ORF72* are also implicated in amyotrophic lateral sclerosis (ALS), but are effectively impossible to characterize with short read sequencing technologies. Therefore, inspired by our success in VACV, we will present new approaches for applying long-read ONT sequencing to the characterization of complex structural variation (SV) in the human genome.

1069W

Whole genome sequencing in a large pediatric African American sample. *P. Sleiman[1,2], L. Tian[1], D. Li[1], X. Chiang[1], H. Hakonarson[1,2].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

African Americans have been underrepresented in genetic research, a trend that has continued with the advent of NGS. Samples classed as African/African American make up 8.5% of the ExAC database despite harboring more variation than other populations. We have generated deep whole genome sequencing data on 900 pediatric samples including 532 children of African American ancestry representing one of the largest pediatric and African American whole genome datasets generated to date. Electronic medical record-derived deep phenotyping data is available on all participants, however, the primary aim of this work was not disease association, but rather a population genetic overview. All samples were sequenced to a depth of ~30X using a PCR free library prep on an Illumina HiSeq. FASTQ reads were aligned to the human genome reference (GRCh37) using BWA (version 0.7.12). The resulting BAM files were position sorted and duplicate reads marked by sambamba and samblaster. For SNP calling, we found a combination of FreeBayes, GATK HaplotypeCaller, and GATK UnifiedGenotyper yields the most sensitive integrated callset. For short insertions and deletions, a combination of FreeBayes, GATK HaplotypeCaller, and Platypus produced the best genotype callset. LUMPY and DELLY2 were used to call structural variants. ANNOVAR, SnpEff, and Variant Effect Predictor (VEP) were used in the pipeline to annotate and classify the effects of SNPs, short insertions and deletions, CNVs and SVs, using both Ensembl and RefSeq transcript sets. We present an overview of the 900 genomes, including the distribution and burden of rare variants in African Americans (AA) compared with the European ancestry (CEU) samples in our cohort, and with the variation and frequencies previously reported in ExAC, the distribution of rare variation in functional and coding sequence between AA and CEU, the distribution of structural variation and CNVs. We present an overview of non coding variation under negative selection in AAs by analyzing site frequency spectra in ENCODE elements, as well as structural and non coding variants under positive selection. Finally, we take advantage of deep sequence coverage on the Y and mitochondrial chromosomes to construct a detailed phylogeny of the AA individuals in the study. .

1070T

Y-chromosomal sequencing and screening reveal both stability and migrations in North Eurasian populations. *O. Balanovsky[1,2], V. Zaporozhchenko[2,1], A. Agdzhoyan[1,2], I. Alborova[5], M. Kuznetsova[2], V. Urasin[3], M. Zhabagin[4], M. Chukhryaeva[2,1], Kh. Mustafin[5], C. Tyler-Smith[6], E. Balanovska[2].* 1) Vavilov Institute of General Genetics, Moscow, Russian Federation; 2) Research Centre for Medical Genetics, Moscow, Russia; 3) YFull service, Moscow, Russia; 4) National Laboratory Astana, Nazarbayev University, Astana, Republic of Kazakhstan; 5) Moscow Institute of Physics and Technology (State University), Moscow, Russia; 6) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom.

Y-chromosomal markers exhibit the highest interpopulation diversity in the genome and thus form one of the most informative tools for tracing population history. However, their information value depends on discovering SNPs which subdivide haplogroups with broad geographic distribution into branches revealing fine population structure. Progress in such discoveries has recently moved from a slow linear phase to a rapid exponential phase due to NGS. We applied this approach to the Y-chromosomal pool of North Eurasian populations and concentrated on haplogroups C, G1, G2, N1b, N1c, and R1b. We sequenced 181 Y-chromosomes (capturing 11 Mb from each sample), developed the *NGSConv* software for calling Y-chromosomal SNPs, and identified roughly 2,500 SNPs, most of which were new. Then we constructed phylogenetic trees and dated dozens of their branches using our estimates of the mutation rate. The last – but not the least – step included screening branch-defining SNPs in the entire Biobank of indigenous North Eurasian populations (led by prof. Elena Balanovska), which includes 26,000 samples from 260 populations. This screening resulted in frequency distribution maps of 29 branches of haplogroups R1b and C, thus increasing the phylogenetic resolution by an order of magnitude compared to the two initial haplogroups. For haplogroup R1b, we identified a previously unstudied "eastern" branch, R1b-GG400, found in East Europeans and West Asians and forming a brother clade to the "western" branch R1b-L51 found in West Europeans. The ancient samples from the Yamnaya archaeological culture are located on this eastern branch, showing that the paternal descendants of the Yamnaya population – in contrast to the published autosomal findings - still live in the Pontic steppe and were not an important source of paternal lineages in present-day West Europeans. For haplogroup C-M217 - the predominant paternal component in Central Asians - we found signals of simultaneous expansion in two independent branches. Both expansion times and gene geographic maps of the expanded lineages indicated the emergence of the Mongol Empire as the likely trigger. We conclude that simply discovering new SNP is not enough, but in combination with screening for the branch-defining SNPs in large biobanks of indigenous populations, it allows comprehensive reconstruction of male population history. The study was supported by the Russian Science Foundation grant 14-14-00827 to OB. .

1071F

The borrowed fitness: Adaptation of modern Eurasians through archaic hominin introgression. *Y. Hu, Q. Ding, L. Jin.* Department of Human Genetics and Anthropology, Fudan University, Shanghai, China.

The standing or *de novo* mutations are considered as the primary source of genetic variations for adaptation especially to a changed environment. Ancestors of modern Eurasians, after their exodus from Africa continent, had to face the challenges of new environment and their fates were largely dictated by the genetic variations that they carried until they encountered the residents in Eurasia continent such as Neanderthal and Denisovan. The signatures of positive selection to the introgressed fragments of Neanderthal found in all sampled modern Eurasians suggest that such introgressions may have contributed to the adaptation of modern human in Eurasia, referred to as the borrowed fitness. In the Eurasian samples of 1000 Genomes Project, the contribution of Neanderthal introgressions to modern human was estimated at about 1% across the genome. High frequency (>30%) of Neanderthal alleles was found in genes related to inflammatory response, epigenetic modification, spermatogenesis, neuron development, etc. In particular, Neanderthal introgression in *HYAL2* and *MC1R*, two genes related to response to sunlight, may have helped adaptation of East Asians in local sunlight. In addition to Eurasians, we found that 78 Denisovan introgressions in Papuans and aboriginal Australians were under positive selection, involving genes related to evolutionarily important functions, such as fertilization, cold acclimation, brain development, immunity, etc. Denisovan introgressions also introduced at least 121 missense alleles into Papuans, which are also located in evolutionarily important genes involved in female pregnancy, visual and smell perception, wound healing, and metabolism. These findings suggested that introgressions of archaic hominins may have substantially facilitated adaptions of modern humans to the new environments. Thus, in addition to mutations, borrowing adaptive variations (i.e. borrowed fitness) from archaic hominin who had already adapted to Eurasia environment served as a supplementary and efficient mechanism to the adaptation for newcomers. .

1072W

The genetic origin of Tibetans and Sherpas of China as revealed by 15 autosomal STRs. *L.L Kang[1,2,3], L.F. Ma[1,2,3], Y.D Zhao[1,2,3], Y. Zhang[1,2,3], Z.Y. Zhang[1,2,3], N. He[1,2,3], Y. Zhang[1,2,3], L.J Liu[1,2,3].* 1) School of Medicine, Xizang Minzu University, Xianyang,Shaanxi, Shaanxi, China; 2) Key Laboratory for Molecular Genetic Mechanisms and Intervention Research on High Altitude Disease of Tibet Autonomous Region;;; 3) Key Laboratory of High Altitude Environment and Gene Related to Disease of Tibet Ministry of Education, School of Medicine, Xizang Minzu University, #6 East Wenhui Road, Xianyang 712082, Shaanxi, China.

Tibet Autonomy Region of China is host to a number of populations that are different in their language, culture and ethnic identity. However, rare genetic studies based on STR markers on Tibetans and Sherpas have been reported in literature till date. In the present study, we genotyped 15 forensic autosomal STRs from 225 unrelated Tibetan individuals and 181 Sherpas from Zhang-mu Town of Shigatse Prefecture. Genetic studies of Tibetans and Sherpas were performed using Cervus 3.0, Arlequin (version3.5) software package, Structure (version 2.3.4) software, R statistical software version 3.0.3 and Distruct version 1.1. The results suggested that Tibetans and Sherpas are most similar to the East Asians, and the least similar to Africans. Our results suggested that Tibetans and Sherpas shared very similar membership with other East Asian populations, especially revealing a common genetic makeup. **Key Words:** Tibetans; Sherpas; STRs; Genetic origin   **Acknowledgments** We are grateful to all of the individuals in this study who made this work possible. .

1073T

Genetic diversity and natural selection footprints of the glycine amidinotransferase gene in various human populations. *A. Khan[1,2], L. Tian[1], C. Zhang[1], K. Yuan[1], S. Xu[1].* 1) Biochemistry, Abdul Wali Khan Unversity Mardan, Mardan, KPK, Pakistan; 2) CAS-MPG Partner Institute of Computational Biology, Shanghai Institute of Biological Sciences, Shanghai China.

The glycine amidinotransferase gene (GATM) plays a vital role in energy metabolism in muscle tissues and is associated with multiple clinically important phenotypes. However, the genetic diversity of the GATM gene remains poorly understood within and between human populations. Here we analyzed the 1,000 Genomes Project data through population genetics approaches and observed significant genetic diversity across the GATM gene among various continental human populations. We observed considerable variations in GATM allele frequencies and haplotype composition among different populations. Substantial genetic differences were observed between East Asian and European populations (FST = 0.56). In addition, the frequency of a distinct major GATM haplotype in these groups was congruent with population-wide diversity at this locus. Furthermore, we identified GATM as the top differentiated gene compared to the other statin drug response-associated genes. Composite multiple analyses identified signatures of positive selection at the GATM locus, which was estimated to have occurred around 850 generations ago in European populations. As GATM catalyzes the key step of creatine biosynthesis involved in energy metabolism, we speculate that the European prehistorical demographic transition from hunter-gatherer to farming cultures was the driving force of selection that fulfilled creatine-based metabolic requirement of the populations.

1074F

Baboon *Alu* elements gone wild. *M.K. Konkel, V.E. Jordan, C.J. Steely, T.O. Beckstrom, J.A. Walker, E. Bennett, B. Clement, A. Robichaux, M.A. Batzer for the Baboon Genome and Analysis Consortium.* Department of Biological Sciences, Louisiana State University, Baton Rouge, LA.

The habitat of baboons (genus *Papio*) ranges across sub-Saharan Africa and the southern Arabian Peninsula. Baboons, which radiated ~2-3 million years ago, have a complex speciation history with evidence for introgressive hybridization. A consensus of six distinct baboon species has recently emerged: olive, yellow, Guinea, hamadryas, chacma, and kinda. However, the evolutionary relationships among baboons continue to be controversial, likely in part due to the presence of stable hybrid zones where baboons meet in the wild and produce fertile offspring. The divergence of baboon-macaque occurred at a comparable time to the human-chimpanzee-gorilla divergence, allowing for direct comparisons of mobile element evolution and amplification. Our analyses of the *Papio anubis* (olive baboon) [papAnu2] draft genome assembly support a rapid expansion of *Alu* elements, suggesting a roughly 6-fold increase in the *Alu* insertion rate compared to the human lineage. To investigate the population and phylogenetic relationships within genus *Papio,* we analyzed a panel of 80 diverse baboons (encompassing all six baboon species) using 497 informative polymorphic *Alu* insertions. To reduce ascertainment bias we selected candidate insertion loci from both the baboon reference genome assembly and from high-throughput sequencing data of a diversity panel that contained individuals of six species within *Papio*. Our analyses reveal a high degree of *Alu* insertion polymorphism across different baboon species and unexpectedly high homoplasy rate, in agreement with extensive incomplete lineage sorting, recent speciation, and/or ongoing introgressive hybridization. Depending on marker selection, our data reflect previously reported evolutionary discrepancies. Analysis of older *Alu* insertion data supports a Northern (olive, Guinea, and hamadryas) and Southern (yellow, chacma, and kinda baboon) species divergence . Inclusion of all the polymorphic *Alu* loci suggests a more complicated history, in agreement with continuous introgression. Our Structure analyses clearly distinguish the six baboon species. Furthermore, yellow baboons from the Southwest primate center show varying degrees of admixture with olive baboons. Two distinct population clusters each are supported within kinda and yellow baboons. In summary, our analyses support the complex relationships among baboons, and provide evidence for rapid mobile element expansion possibly linked to inter-species hybridization.

1075W

Leveraging large-scale clinical biobanks to study recent human evolution. *C. Simonti[1], W. Bush[2], J.A. Capra[1,3], eMERGE Consortium.* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Department of Biological Sciences, Vanderbilt University, Nashville, TN.

We introduce a high-throughput approach for characterizing the function of variation arising in recent human history and use it to demonstrate that introgressed Neanderthal alleles and targets of recent positive selection influence diverse biological systems, including psychiatric, skeletal, and immunological traits. Thousands of variants throughout the human genome have been identified in studies of positive selection, archaic admixture, and other evolutionary events that have shaped human populations. However, the functions of the vast majority of these variants are unknown. Historically, determining the phenotypic effects of such variants relied on indirect evidence, such as the function of nearby genes, or low-throughput methods for functional characterization, such as enhancer reporter assays. To overcome this limitation, we mined the electronic health records (EHRs) of ~28,000 genotyped individuals from seven hospitals across the US to identify cases and controls for ~1,500 clinical phenotypes. To illustrate the use of this resource in evolutionary studies, we identified phenotypic associations for two classes of variants important in recent human history: those present in the human genome due to admixture with Neanderthals, and variants appearing on the human lineage that have risen to near fixation since divergence with chimp (hominin-derived). We discovered and replicated associations of introgressed Neanderthal alleles with neurological, psychiatric, immune, skin, and reproductive phenotypes. Among the hominin-derived variants, we found several that influence risk for skeletal disorders, including bone fracture. Using functional genomics data from ENCODE/Roadmap Epigenomics and gene expression data from GTEx, we demonstrate that many of these variants influence the regulation of nearby genes relevant to the associated phenotypes. Our results establish that recent hominin-derived variants and archaic admixture influence risk for diverse, evolutionarily-relevant phenotypes in modern humans, and demonstrate the utility of EHR data linked to genotypes to inform evolutionary analyses.

1076T

Down-regulation of *EPAS1* transcription explains genetic adaptation of Tibetans to high-altitude hypoxia. *B. Su[1,2], Y. Peng[1], C. Cui[2], Y. He[1,8], T. Ouzhuluobu[2], H. Zhang[1], D. Yang[1], Q. Zhang[1], T. Bianbazhuoma[3], L. Yang[1], Y. He[1,8], K. Xiang[1], X. Zhang[1], S. Bhandari[1], P. Shi[1], T. Yangla[2], T. Dejiquzong[2], T. Baimakangzhuo[2], T. Duojizhuoma[2], Y. Pan[2], T. Cirenyangji[2], T. Baimayangji[2], T. Gonggalanzi[2], C. Bai[2], T. Bianba[2], T. Basang[4], S. Xu[5,9,10], H. Chen[6], T. Wu[7], X. Qi[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 2) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa 850000, China; 3) The Municipal People's Hospital of Lhasa, Lhasa 850000, China; 4) People's Hospital of Dangxiong County, Dangxiong 851500, China; 5) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai; 6) Center for Computational Genomics, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China; 7) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining 810012, China; 8) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100101, China; 9) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 10) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

Tibetans are well adapted to the hypoxic environments at high altitude, yet the molecular mechanism of this adaptation remains elusive. We reported comprehensive genetic and functional analyses of EPAS1, a gene encoding hypoxia inducible factor 2α (HIF-2α) with the strongest signal of selection in previous genome-wide scans of Tibetans. We proved that the Tibetan-enriched EPAS1 variants down-regulate expression in human umbilical endothelial cells and placentas. Heterozygous EPAS1 knockout mice display blunted physiological responses to chronic hypoxia, mirroring the situation in Tibetans. Furthermore, we found that the Tibetan version of EPAS1 not only accounts for the relatively low hemoglobin level as a polycythemia protectant, but also enables low pulmonary vasoconstriction response in Tibetans. We demonstrated that the down-regulation of EPAS1 serves as the molecular basis of Tibetans' adaption to high-altitude hypoxia.

1077F

Selection and functional consequence of Neanderthal introgression. *N. Telis[1], K. Harris[2], J.K. Pritchard[2,3,4].* 1) Biomedical Informatics, Stanford, CA; 2) Genetics, Stanford, CA; 3) Biology, Stanford, CA; 4) Howard Hughes Medical Institute.

There has been a long history of interest in the extent of, and selective consequences of, human interbreeding with other hominins. Existing knowledge of Neanderthal introgressed variants across the genome has revealed several deserts, as well as a general depletion in the vicinity of genes, with weak association with testes expression. Introgression has been linked with risk for multiple diseases, and therefore it seems prudent to understand broader trends in functional consequences of introgression. Moreover, characterizing sources of strong pressures against presumably deleterious Neanderthal introgression can shed light into broader functional trends of modern selection. By examining distributions of introgression across regulatory elements, we identify genome-wide depression of introgression in promoters and enhancers. This genome-wide depletion increases in extremity with increasing allele frequencies. Alongside this broader depletion of Neanderthal introgression in regulatory regions, we confirm independent depletion signals in brain, embryonic stem cell, and muscle tissues. Regulatory variants with likely severe effects (due to binding site alteration) also appear depleted, with variants affecting lymph node as a notable exception. Based our our analysis, we present a model linking regulatory consequence with fitness and distribution of existing Neanderthal variations.

1078W

Determinants of mutation rates in germline and soma. *C. Chen[1], H. Qi[3], Y. Shen[3], J.K. Pickrell[1,2,4], M. Przeworski[1,3,4].* 1) Department of Biological Sciences, Columbia University, New York, NY; 2) New York Genome Center, New York, NY; 3) Departments of Systems Biology and Biomedical Informatics, Columbia University Medical Center, New York, NY; 4) Co-supervised this project.

Understanding the determinants of fine-scale mutation rates is of crucial importance for many efforts in human and evolutionary genetics. Recent studies of somatic mutations in tumors and of germline mutations in pedigrees have led to the identification of a number of factors that influence mutation rates, including CpG methylation, expression levels, replication timing and GC content. Interestingly, some of these effects appear to differ between somatic and germline tissues. Notably, while mutation rates have been reported to decrease with expression levels in tumors, no such effect has been detected in the germline, leading to speculation that error or repair processes may differ among tissues. It is hard to know whether these apparent discrepancies are real, however, as different approaches were taken to analyze the data in different studies, with distinct sets of variables included as covariates and varying window sizes as the unit of analysis. To enable a cleaner comparison, we considered a model in which the mutation rate in a gene is predicted by GC content, expression levels, replication timing and two histone repressive markers. We applied this model to a large set of germline mutations identified in exomes from recently published studies on developmental disorders, and to exonic somatic mutations in four types of tumors from the Cancer Genome Atlas. Our analysis indicates that germline and soma share most determinants of mutations, but with some differences. Most intriguingly, the degree of strand asymmetry and germline mutation rate increase with expression in ovaries and testes, whereas the degree of strand asymmetry and somatic mutation rates have opposite dependencies on expression levels in somatic tissues matched to the tumor type. We speculate that the source of this difference between germline and soma is the occurrence of domain-associated repair (DAR) in somatic cells. .

1079T

Admixture inference of African Americans and Latinos in the United States through time. *M.L. Spear[1], D.G. Torgerson[2], R.D. Hernandez[1,3,4].* 1) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 2) Department of Medicine, University of California, San Francisco, San Francisco, CA; 3) California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, San Francisco, CA; 4) Institute for Human Genetics, University of California, San Francisco, CA.

The study of admixed populations has provided important insights into medical genetics and population history. The genomes of admixed individuals are mosaics of segments originating from different ancestral populations. At the genome-wide level, the proportion of one's genome deriving from each ancestral population is referred to as "global ancestry proportions". However, modern statistical methods enable inference of the ancestry at individual SNPs within a genome, "local ancestry", which allow us to reconstruct the mosaic pattern of ancestry tracts across an individual's genome. Local ancestry inference is critical for the analysis of admixed genomes and has been widely studied in the fields of medical genetics and human demographic history. Local ancestry tracts can be used to infer migration histories but the question remains how these histories have shaped ancestry proportions over time, particularly in the United States, a "melting pot" country that has faced changing societal norms over the past century. It has yet to be determined how the length distribution of ancestry tracts in admixed individuals has changed over decades as well as how the variation in ancestry proportions across chromosomes and individuals may differ. Thus, we estimated local ancestry for 4,600 Latinos and 2,100 African Americans from the Genetic Epidemiology Research on Adult Health and Aging (GERA) dataset using RFMix. With these local ancestry tracts, we used TRACTS to compare the observed length of the ancestry tracts to predictions of different demographic models of migration scenarios. Individuals were grouped by 5-year birth year categories, and comparisons were made between the demographic models generated from each birth year category. Overall, the local ancestry tracts of African Americans and Latinos from the United States have provided insights into the change in complexity of their genetic structure throughout the 20th century. .

1080F

Genomic and transcriptomic diversity in Xinjiang's Uyghurs. *L. Tian[1,2], C. Zhang[1,2], Y. Yuan[1], X. Wang[1,2], D. Lu[1,2], M. Shi[1,2], Y. Lu[1], K. Yuan[1,2], H. Lou[1], Q. Feng[1,2], R. Fu[1,2], X. Zhang[1,3], A. Khan[1], Y. Yang[4], Y. Guan[5], S. Xu[1,2,3,6].* 1) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai; 2) University of Chinese Academy of Sciences, Beijing 100049, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 4) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 5) Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, P. R China; 6) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

Whole-genome sequencing simultaneously DNA (30–60×) and mRNA in blood of Uyghurs and Han Chinese allows us, for the first time to our knowledge, to systematically investigate genomic and transcriptomic diversity in a Eurasian admixed population. We observed substantial splicing variation in Uyghurs and alternative splicing explains a substantial proportion of gene expression diversity. We identified a considerably large number of eQTLs and sQTLs, and differentiated contribution of the different ancestries to the gene expression patterns or levels. We argue that analyzying admixed populations facilitates to our understanding of the nature of genetic variants that affect gene expression and provide insights into elucidating the causes of phenotypic variation.

1081W

Assessing the value of population specific reference genomes by the '1000 Swedes project'. *A. Ameur[1], R. Karlsson[2], H. Che[1], I. Bunikis[1], P. Olason[1], I. Höijer[1], S. Häggqvist[1], I. Jonasson[1], P. Magnusson[2], U. Gyllensten[1]. 1) SciLifeLab, Uppsala University, Uppsala, Sweden; 2) Swedish Twin Registry, Karolinska Institutet, Sweden.*

We are undertaking the largest human whole-genome sequencing (WGS) project in Sweden to date; the generation of a reference database that reflects the genetic structure of the population. 1000 individuals were included in the study, selected to cover the genetic variation within Sweden based on results from PCA analysis of available SNP array data. All of these 1000 individuals have been subjected to short-read WGS using Illumina technology at 30X coverage. Two of these individuals were also chosen for long-read single molecule real-time (SMRT) sequencing on the PacBio RSII platform, generating ~50X coverage of the human genome with sequence reads of >10kb length on average.   A principal aim of this project is to establish a reference database of genetic variants in the Swedish population for use in disease association studies and as a look-up resource for the evaluation genetic variants identified in sequencing of patients. This database will be the first of its kind in Sweden and is needed both for GWAS-based disease association studies and for evaluating the clinical relevance of genetic alterations detected during clinical diagnostic analyses. It will also provide a unique dataset for research on the structure of genetic variation within Sweden, as well as for international collaborative projects.   Within this project we also investigate the benefits of creating novel reference genome sequences for the Swedish population. From the PacBio data, we have *de novo* assembled draft reference genomes using the FALCON software. This resulted in more than 2.8Gb of sequence data, distributed on ~7000 contigs, for each of the two individuals investigated. Our *de novo* assemblies showed very high similarity to the available human genome reference and 98.6% of the bases in our contigs could be mapped to GRCh38. Using the software Sniffles, we detected a substantial number of large insertions (~4000), deletions (~3700), inversions (~900) and duplications (~1300) with respect to the GRCh38 reference. Interestingly, we also found several mega bases of novel sequence data in our local PacBio genome assemblies, i.e. sequences not present in GRCh38. We are now using the Illumina data to validate these findings and to estimate the prevalence of the complex rearrangements in the Swedish population.

1082T

Large protein-coding polymorphisms in the human genome. *S. Cantsilieris[1], J. Huddleston[1,2], M.J.P. Chaisson[1], L. Harshman[1], K. Munson[1], M. Scofield-Sorensen[1], E.E. Eichler[1,2]. 1) Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Howard Hughes Medical Institute, Seattle, WA, USA.*

Massively parallel sequencing of whole genomes and exomes has provided remarkable insight into the spectrum of both rare and common genetic variation that underlies protein-coding DNA. Here, we focus on regions of the genome that are inaccessible to short-read whole-genome shotgun sequencing with a specific focus on protein-coding sequence variation. Leveraging long sequence reads generated by single molecule, real-time (SMRT) sequencing of human genomes, in combination with methods to detect indels and structural variants up to 30 kbp, we identified >30 structural variant calls >1 kbp that intersect protein-coding genes. Selecting the largest and most complex events, we sequenced and assembled large-insert clones (>250 fosmid/BAC) from eight diverse humans in order to understand the extent of human variation and haplotypic differences among genes and gene families. For example, our investigation of 11 mucin gene family members identifies individual haplotypes where protein-coding sequences differs by tens of kilobases due to expansions of protein-encoding higher-order repeat structures. The expansions maintain frame and suggest that different human haplotypes produce protein molecules that can differ by as much as 5000 amino acids or 550 kilodaltons. Additionally, we observe similar hyperexpanded protein-coding sequence genes associated with immune response, lipoprotein levels and epidermal barriers (hornerin, filaggrin, and repetin). Using this approach, we also generated a 220 kbp alternate haplotype of the amylase gene cluster, a group of genes embedded within a series of tandem duplications thought to be important for human adaptation to starch-rich diets. Sequence analysis reveals signatures of positive selection and interlocus gene conversion during human and primate evolution. Comparison with the human reference assembly shows that many of these regions are enriched for "black tag" annotations or contain sequences that are missing or incompletely assembled within the human reference genome. Our analysis has characterized a striking level of protein-coding polymorphism achieved by large-scale expansion of tandem repeats within coding exons that include a subset of human genes associated with environmental interaction and host defence. These high-quality alternate reference haplotypes provide insight not only into the mutability and complex patterns of human genetic coding variation but also a starting point for association with human disease.

1083F

Identifying adaptive variants and tracing their ancestral origins in Tibetan highlanders by whole-genome deep sequencing. *L. Deng[1,2], C. Zhang[1,2], H. Lou[1], K. Yuan[1,2], D. Lu[1,2], X. Wang[1,2,3], Y. Lu[1], R. Fu[1,2], Y. Wang[1,2], X. Yang[1,2], Y. Hu[4], Q. Ding[4], Y. Zhou[1,2], Q. Feng[1,2], Z. Wu[1,2,3], X. Zhang[1,2,3], L. Tian[1,2], Y. Yang[1,2], S. Li[4], L. Jin[4], Y. Guan[5], B. Su[6], L. Kang[7], S. Xu[1,2,3,8].* 1) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai; 2) University of Chinese Academy of Sciences, Beijing 100049, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 4) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 5) Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, P. R China; 6) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 7) Key Laboratory for Molecular Genetic Mechanisms and Intervention Research on High Altitude Disease of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang Shaanxi 712082, China; 8) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

Current knowledge of the genetic basis of Tibetans' adaptation to altitude is largely based on genotyping known variants or sequencing single candidate genes. Here, we report the first whole-genome deep sequencing (30–60×) study of 38 Tibetan highlanders living at seven different altitudes above 2,500 meters and 39 lowland Han Chinese. We identify 11.57 million variants with 2.22 million novel to known human variation. We build a list of putative adaptive genetic variants (AGVs) including 48 missense and 7 loss-of-function variants which affect 29 genes related to hypoxia adaptation. The majority of the AGVs are significantly correlated with altitude. Notably, we observe two highly differentiated genomic regions which show their elevated archaic ancestry and typical anatomically modern human origin, respectively. In both genomic regions, the sequences harboring AGVs in Tibetan highlanders trace their most recent common ancestor (MRCA) back to the time predating the Last Glacial Maximum. Our results indicated a much more complicated mechanism of human adaptation to altitude than previously thought, with key adaptions to the extreme environment contributed by pre-LGM populations.Current knowledge of the genetic basis of Tibetans' adaptation to altitude is largely based on genotyping known variants or sequencing single candidate genes. Here, we report the first whole-genome deep sequencing (30–60×) study of 38 Tibetan highlanders living at seven different altitudes above 2,500 meters and 39 lowland Han Chinese. We identify 11.57 million variants with 2.22 million novel to known human variation. We build a list of putative adaptive genetic variants (AGVs) including 48 missense and 7 loss-of-function variants which affect 29 genes related to hypoxia adaptation. The majority of the AGVs are significantly correlated with altitude. Notably, we observe two highly differentiated genomic regions which show their elevated archaic ancestry and typical anatomically modern human origin, respectively. In both genomic regions, the sequences harboring AGVs in Tibetan highlanders trace their most recent common ancestor (MRCA) back to the time predating the Last Glacial Maximum. Our results indicated a much more complicated mechanism of human adaptation to altitude than previously thought, with key adaptions to the extreme environment contributed by pre-LGM populations. .

1084W

Whole-genome sequencing in French Canadians from Quebec. *C. Low-Kam[1], D. Rhainds[1], K.S. Lo[1], S. Provost[1], I. Mongrain[1], A. Dubois[1], S. Perreault[2], J.F. Robinson[3], R.A. Hegele[3,4], M.-P. Dubé[1,5], J.-C. Tardif[1,5], G. Lettre[1,5].* 1) Montreal Heart Institute, Montréal, Québec, Canada; 2) Faculté de Pharmacie, Université de Montréal, Montréal, Québec, Canada; 3) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 4) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 5) Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada.

We present the first large-scale low-pass whole-genome sequencing (WGS) effort of the French-Canadian population. These genomes were collected as part of a larger project that focuses on the relationship between the functionalities of high-density lipoproteins (HDL) and coronary artery disease in French Canadians. Our work builds on previous studies that targeted specific subpopulations to discover population-specific genetic variation, in particular rare and low-frequency coding DNA variants. We sequenced at ~5.6x coverage the whole-genome of 1,970 French Canadians recruited by the Montreal Heart Institute Biobank and identified 29 million bi-allelic variants (51% novel), including 19 million variants with a minor allele frequency (MAF) <0.5%. Genotypes from the WGS data are highly concordant with genotypes obtained independently by exome array on the same individuals (99.8%), even when restricting this analysis to rare variants (MAF <0.5%, 99.9%) or only heterozygous sites (98.9%). Likely detrimental DNA sequence variants are enriched at the rarer end of the allele frequency spectrum, consistent with purifying selection. For instance, whereas 17% of synonymous SNVs are common, 4% of nonsense and 11% of frameshift indel variants have a MAF >5% . To validate the utility of our dataset to discover genetic associations, we showed that despite its modest sample size, we can effectively replicate associations with myocardial infarction risk and blood lipid levels. For instance, 51 of the 69 (binomial P=9x10^{-5}) SNPs previously shown to be associated with HDL cholesterol levels have an effect in the right direction, and eleven of these are nominally significant (P<0.05). To explore the value of our WGS dataset beyond association analysis, we built a French-Canadian-specific imputation reference panel and showed that it slightly outperforms the 1000 Genomes Project reference haplotypes for rare variants in independent French-Canadian samples (mean r^2 of 0.76 *vs.* 0.71). Although most of the participants were recruited in the Montreal area, we used information on their ancestry to illustrate how this WGS dataset reflects the population structure in the Province of Quebec. This collection of results shows that low-pass WGS is a promising resource to analyze the genetics of human diseases in the founder French-Canadian population.

1085T

Detection of small indels in 133 rhesus macaques (*Macaca mulatta*) by whole genome sequencing. *M. Raveendran[1], R. Alan Harris[1], L. Cox[2], G. Fan[3], B. Ferguson[4], J. Horvath[5,6,7], S. Kanthaswamy[8,9], H. Michael Kubisch[10], D. Liu[11], M. Platt[12,13], D.G. Smith[8], B. Sun[11], E.J. Vallender[10,14], R.W. Wiseman[15], D.M. Muzny[1], R.A. Gibbs[1], J. Rogers[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Southwest National Primate Research Center, San Antonio, TX; 3) Univ. of California, Los Angeles; 4) Oregon National Primate Research Center; 5) North Carolina Museum of Natural Sciences; 6) North Carolina Central University; 7) Dept. of Evolutionary Anthropology, Duke University; 8) California National Primate Research Center; 9) Arizona State University; 10) Tulane National Primate Research Center; 11) Anhui University, Anhui, China; 12) Dept. of Neurobiology, Duke University; 13) Univ. of Pennsylvania; 14) New England National Primate Research Center; 15) Wisconsin National Primate Research Center.

Among closely related species (e.g. humans and chimpanzees) small insertion-deletion differences (indels) account for more genetic divergence than single base substitutions. Indels are also known to explain a number of damaging mutations associated with human disease. Rhesus macaques (*Macaca mulatta*) are one of the most commonly used nonhuman primates in biomedical research, and genetic variation among rhesus can be useful in studying and understanding human diseases. However, little is known about the frequency or functional consequences of small indel polymorphisms among rhesus macaques. Here we describe the identification of small indels (1 to 60 bases in length) in a sample of 133 rhesus macaques (124 Indian-origin and 9 Chinese-origin). Using GATK software to call indels from whole genome sequence data (average coverage 26.7x), and excluding variant calls that do not pass quality control filters or are likely the result of errors in the reference genome, we identified 7,926,645 indels across the 133 animals. This includes 3,026,673 insertions, 3,951,905 deletions and 948,067 more complex sequence alterations with ambiguous mechanism. Out of these 7.92 million indels, 1.68 million are found only in Chinese rhesus and 2.94 million are Indian-origin specific. Across the full dataset, we observed 2.6 indels per 1kb of the rhesus genome. 48.25% of bases affected are due to insertions and 51.75% are due to deletions. Most of the indels (74.5%) are 1 to 4 basepair in length. The average number of indels per animal is 1,249,646, which is higher than observations for individual human personal genomes. To determine potential effects on gene function or protein sequence, we analyzed all rhesus indels using the Ensembl Variant Effect Predictor (VEP) tool and identified stop codons altered (365), coding sequence frameshift indels (7,664), inframe insertions (1,563), inframe deletions (2,165) and splice region indels (4,317). In order to identify possible disease models, we queried the DisGeNET database (http://www.disgenet.org) with genes containing frameshift indels. This identified several genes associated with specific human diseases, including muscular dystrophy (DMD), microcephaly (MCPH1), cleft palate X-linked (TBX22) and Tangier disease (ABCA1).

1086F

Mutation load is compensated by allele specific expression per individual and across populations. *I. Alves[1], H. Edgington[1], V. Bruat[2], J-C. Grenier[2], M-J. Favé[2], P. Awadalla[1,2,3].* 1) Ontario Institute for Cancer Research, OICR, Toronto, Ontario, Canada; 2) Sainte-Justine University Hospital Research Centre, Department of Pediatrics, University of Montreal, Quebec, Canada; 3) Department of Molecular Genetics, University of Toronto, Ontario, Canada.

Recently, several studies have shown that individual genomes and populations harbor varying numbers of putatively deleterious mutations, most of which are present at low-frequencies in the population. For example, the mutation load appears to vary depending on the size of the population, being particularly more pronounced in very small or founding populations. However, inferred mutation load from genome sequence data does not integrate information with respect to transcriptional or translational variation, including relationship to phenotypes. By combining whole-blood transcriptomic and genome-wide SNP-chip data from ~ 650 individuals with European and French-Canadian ancestry from the CARTaGENE Project, we measured individual allele specific expression (ASE) across all heterozygous SNP positions and incorporated ASE proportions across individuals to understand whether those variants showed a derived-, ancestral- or no biased expression profile at the population level. We then summarized the population derived allele frequencies, using the site-frequency spectrum (SFS), by independently conditioning on variants falling into the three expression profiles. We found that sites with derived-biased expression across the population show enrichment for derived alleles near fixation, but no bias towards functionally relevant alleles was found. On the other hand, sites showing an average allele expression bias towards the ancestral allele are enriched for rare derived alleles and carry an increased proportion of constrained, putatively damaging and missense alleles when compared to those variants with no bias in expression levels. Finally, when the burden carried by a particular derived allele is weighted according to its transcript abundance, the distribution of the mutation load per individual is inferred to be slightly smaller, but still significant, than one retrieved without accounting for ASE. Overall, our results suggest that allele specific expression tends to compensate for inferred genomic mutation load by reducing the expression of potentially damaging alleles at the level of the individual transcriptome, thereby possibly reducing the overall cost of deleterious mutations on individual' fitness. .

1087W

The NHGRI Sample Repository: A collection of cell lines and DNA for human genetic research. *A.M. Resch, F. Rahimov, M. Berkowitz, L. Brown.* Coriell Institute for Medical Research, Camden, NJ.

Knowledge of the spectrum of genetic variation across diverse human populations is important for reconstructing human evolutionary history and understanding the genetic basis of disease. Established in 2006, the NHGRI Sample Repository maintains and distributes a collection of well-characterized cell line and DNA samples to support studies of the human genome, including genetic variation in human populations. The NHGRI Repository hosts over 3,700 unique cell cultures and DNA samples collected for the International HapMap and 1000 Genomes Projects. Whole blood samples were collected from anonymized and de-identified donors and were contributed with consent to broad data release for use in future studies, including genotyping and sequencing, gene expression and proteomic studies, and all other types of genetic variation research. The samples include no identifying or phenotypic information. The HapMap and 1000 Genomes collections include samples from 27 distinct populations spanning 5 continents, making it one of the most comprehensive and unique resources of its kind. The 1000 Genomes Project is one of the few large-scale open access genomic datasets with companion cell lines and DNA. By providing the research community with a standard set of biomaterials of known provenance, invesitgators and clincians around the world can compare and contrast results. These standardized resources are critical elements in the efforts to take full advantage of the Human Genome Project for understanding the genetic origin of disease. Cell line and DNA samples from these collections are available to researchers around the globe and may be used by academic, non-profit and commercial entities. The full catalog of samples from the HapMap and 1000 Genomes collections can be accessed at https://catalog.coriell.org/1/NHGRI.

1088T

Ancestral origins and genetic history of Tibetan highlanders. *C. Zhang[1,2], D. Lu[1,2], H. Lou[1,2], K. Yuan[1,2,3], X. Wang[1,2], Y. Wang[1,2], Y. Lu[1,2], X. Yang[1,2], L. Deng[1,2], Y. Zhou[1,2], Q. Feng[1,2], Y. Hu[4], Q. Ding[4], Y. Yang[4], L. Jin[4], Y. Guan[5], B. Su[6], L. Kang[7], S. Xu[1,2,3,8].* 1) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai; 2) University of Chinese Academy of Sciences, Beijing 100049, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 4) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 5) Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, P. R China; 6) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 7) Key Laboratory for Molecular Genetic Mechanisms and Intervention Research on High Altitude Disease of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang Shaanxi 712082, China; 8) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

The origin of Tibetans remains one of the most contentious puzzles in history, anthropology and genetics. Analyses of deep sequenced genomes (30–60×) of 38 Tibetan highlanders and 39 Han Chinese lowlanders together with available archaic and modern human data allows us, for the first time, to comprehensively characterize the ancestral makeup of Tibetans and uncover their origins. Tibetans arise from a mixture of multiple ancestral gene pools but much more complicated and ancient than previously suspected. They share the largest proportion of genetic ancestry with modern human populations from East Asia, and to a lesser extent with those from Central Asia, Siberia, South Asia, western Eurasia and Oceania. Moreover, non-modern human sequences, possible inherited from Neanderthal, Denisovan and other unknown ancient people, comprise ~6% of the Tibetan gene pool and elevated in some genomic regions. In particular, highly differentiated sequences harbored in the highlanders' genomes were likely inherited from multiple non-modern human ancestral origins, and maintained in high frequency by natural selection.

1089F

Modeling human population separation history using physically phased genomes. *S. Song[1], E. Sliwerska[2], S. Emery[2], J. Kidd[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Phased haplotype sequences are a key component in many population genetic analyses since variation in haplotypes reflects the action of recombination, selection, and changes in population size. Haplotypes are typically estimated from unphased sequence or genotyping data using statistical models applied to large reference panels. To assess the importance of correct haplotype phase on population history inference, we performed fosmid pool sequencing and resolved phased haplotypes of five individuals from diverse African populations (including Yoruba, Esan, Gambia, Massai and Mende). We physically phased 98% of heterozygous SNPs into haplotype-resolved blocks, obtaining a block N50 of 1 Mbp. We combined these data with additional phased genomes from San, Mbuti, Gujarati and CEPH European populations and analyzed population size and separation history using the Pairwise Sequentially Markovian Coalescent (PSMC) and Multiple Sequentially Markovian Coalescent (MSMC) models. We find that statistically phased haplotypes yield an earlier split-time estimation compared with experimentally phased haplotypes. To better interpret patterns of cross-population coalescence, we implemented an approximate Bayesian computation (ABC) approach to estimate population split times and migration rates by fitting the distribution of coalscence times inferred between two haplotypes, one from each population, to a standard Isolation-with-Migration model. We inferred that the separation between hunter-gather populations and other populations happened around 120,000 to 140,000 years ago with gene flow continuing until 30,000 to 40,000 years ago; separation between west-African and out of African populations happened around 70,000 to 80,000 years ago, while the separation between Massai and out of African populations happened around 50,000 years ago.

1090W

Fine-mapping the favored mutation in an ongoing positive selective sweep. *A. Akbari[1], A. Iranmehr[1], S. Mirarab[1], V. Bafna[2].* 1) Electrical & Computer Engineering, University of California San Diego, La Jolla, CA; 2) Computer Science & Engineering, University of California San Diego, La Jolla, CA.

Motivation. Methods for detecting genomic regions under selective constraints are heavily studied and have been successful in identifying many ongoing selective sweeps. However, detecting the favored mutation itself remains an important but difficult problem. Here, we present a statistic for Selecting Alleles Favored in Evolution (SAFE). **Method.** SAFE combines scores derived from an extended version of Multivariate Ewen's Distribution on the number of distinct haplotypes in the sub-population of carriers of a mutation (Conditional Coalescent Tree) and the distribution of the Haplotype Allele Frequency (HAF) score. The HAF score, assigned to individual haplotypes in a sample, has previously been shown to successfully separate carriers of the favored allele from non-carriers in ongoing selective sweeps. **Result.** We tested SAFE on both simulated and real data including both hard and soft sweep (standing variation) as well as different demographic models, with very promising results. As an example, we ranked ~300 SNPs (50Kbp region) in each of 1,000 samples from populations simulated under selection with moderate selection coefficient (s=0.01), and standard population genetic parameters (effective population size Ne=10k, mutation rate u=2.5e-8, and recombination rate r=1.25e-8). The median of predicted rank of the favored mutation was 2. The favored mutation was ranked 1 in 46% of samples, and in the top 5% (respectively, 10%) in 94% (respectively, 99%) of samples. We also applied SAFE to Phase 3 of 1000 Genome Project data in regions where the selective sweep is well characterized, and the favored mutation is known with high confidence. Table 1 shows the excellent performance of our method on well-known selective sweeps. SAFE thus provides a sharper tool for studying evolutionary process at work, and complements methods based on functional analysis of variants in the region.

Table 1. Preliminary Result of detecting favored mutations of well-known selective sweeps.						
Genes	Population	Favored SNP	Frequency	Rank	Total # SNPs	Percentile (%)
HBB	YRI	rs334	0.13	1	512	0.2
LCT	CEU	rs4988235	0.71	1	119	0.8
TRPV6	CEU	rs4987682	0.92	8	173	4.6
PSCA	YRI	rs2294008	0.77	1	381	0.3
ADH1B	CHB+JPT	rs1229984	0.72	2	159	1.3
ABC11	CHB+JPT	rs17822931	0.93	1	43	2.3
SLC24A5	CEU+YRI	rs1426654	0.5	3	171	1.8
TLR5	YRI	rs5744174	0.88	9	359	2.5

1091T

Genomic identification of recent positive selection in populations from Andean highlands and Southern Chile. *R. Verdugo[1], P. Contreras[1], A. Blanco[1], A. Moreno-Estrada[2,4], K. Sandoval[2,4], C. Eng[5], S. Huntsman[5], E. Burchard[5,6], C. Gignoux[2,3], C. Bustamante[2,3], E. Llop[1], M. Moraga[1].* 1) Programa de Genetica Humana ICBM, Universidad de Chile, Santiago, RM, Chile; 2) Center for Computational, Evolutionary and Human Genomics, Stanford University; 3) Department of Genetics, Stanford University, Stanford, California; 4) Centro de Investigación y Estudios Avanzados, Instituto Politécnico Nacional, Mexico; 5) Department of Medicine, University of California, San Francisco, California; 6) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California.

Human migrations into South America reached the southern area of the continent by at least 14,800 BP. Therefore, populations inhabiting different regions in the continent may have developed local adaptations to their environment, adding to the genetic variation among indigenous populations. We evaluated genomic evidence of microevolutionary processes in three Amerindian populations from the Andes highlands in Peru (77 individuals from Puno) and Southern Chile (13 Pehuenche and 9 Huilliche) by genotyping with the Axiom LAT1 platform. We found sizable levels of genetic divergence between these populations. The Fixation Index Fst between the Highlands and Southern Chile was 0.046 and between Pehuenche and Huilliche was 0.025. To identify recent selective sweeps (less than 30,000 BP), we calculated iHS and XP-EHH scores in 7600~ windows (200 kb). We found that windows with the highest scores in these tests (p-value <0.05) are enriched in genes related with biological processes as Vitamin D metabolism in Andean individuals, and MHC class II protein complex in Southern Chile. Among these high scored windows, we highlight a region on chromosome 6 (51.8-52.6 Mb) that showed strong evidence of selection in all the three groups analyzed, as previously reported in other Amerindian populations. This region harbors genes associated with regulation of inflammatory responses and calcium homeostasis, among others. These results suggest that selective pressures on these biological mechanisms may have been present during the recent history of Native South American populations and represent a potential source of genetic variation in admixed individuals from South America.

1092F

Demography-aware method for inferring strength of selection on polygenic traits. *C. Antelope, R. Nielsen.* University of California Berkeley, Berkeley, CA.

Detecting selection acting on a polygenic trait is challenging because no single locus affecting the trait may alone show strong evidence of selection. However, the signal from multiple loci affecting the trait can be combined to provide strong signals of selection. We develop a maximum likelihood method for estimating the selection differential of a quantitative trait. The method combines the marginal likelihood functions for selection coefficients across multiple individual loci, using a classic quantitative genetics result from Lande and Arnold. The method assumes the availability of effect size estimates and their variances from GWAS studies, but does not assume complete ascertainment of causal loci. This facilitates formal testing of the hypothesis of selection acting on human polygenic traits in a specific population using a likelihood ratio test, and it provides maximum likelihood estimates of the trait-level selection differential. The method can also be used to predict the average change in the phenotypic value per generation due to selection. We validate the method using simulations and apply it to human GWAS data from Europeans to test previously proposed hypotheses regarding selection acting on height and Type II Diabetes. The method looks at selection in a single population, which complements previously published methods that take advantage of population differentiation to detect polygenic selection.

1093W

Painting by evolutionary history: Inference of local ancestry in admixed genomes. *A. Berens, J. Lachance.* School of Biology, Georgia Institute of Technology, Atlanta, GA.

Local ancestry inference is important for understanding demographic history and predicting hereditary disease risks in admixed populations, which are increasingly becoming the norm as human migration around the world accelerates. Current local ancestry painting methods require reference sequences from ancestral (or at least closely related proxy) populations. Unfortunately, some populations, including Native Americans, are underrepresented in the publically available global population-level genomic datasets due to deficient sampling or absence in modern genomes because of recent admixture. Additionally, current ancestry inference methods overlook potentially relevant information about the historical relationships between these ancestral populations. We propose a new hierarchical, reference-free chromosome painting method to infer individual local ancestries. From whole genome sequence data of admixed individuals, we estimate ancestry proportions on a chromosomal scale. These proportions are used to weight the contribution of individual alleles to ancestral population haplotype frequencies. Using a Hidden Markov Model, we estimate ancestry of each individual at the most informative sites between each population. Local ancestry is painted based on a weighted majority vote from multiple overlapping sliding windows that have been traversed in both the forward (5' to 3') and backward (5' to 3') directions. Our approach provides highly accurate (>90%) local ancestry predictions for simulated genomic data and produces summed local ancestry estimates that successfully recapture global ancestry proportions inferred by ADMXITURE (R^2 = 0.99) using real data from the 1000 Genomes Project. Future applications include iteratively increasing the number of ancestral populations by splitting a single population into two daughter populations and repainting local haplotypes to include these new ancestral populations. This hierarchical approach of painting chromosomes captures evolutionary relationships between ancestral populations.

1094T

The African Genome Resource. *T. Carstensen[1,2], D. Gurdasani[1,2], M.O. Pollard[1,2], C. Pomilla[1], AGR Investigators[1,2,3,4,5].* 1) Global Health and Populations, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 3) Stonybrook University, New York, United States of America; 4) MRC Uganda, Entebbe, Wakiso District, Uganda; 5) University of KwaZulu-Natal, KwaZulu-Natal, South Africa.

Despite large scale sequencing efforts such as 1000G, DNA and RNA sequence data collected from the African continent and made readily available to researchers is still limited. Capturing representative genetic diversity is important to provide a framework for medical genetics in Africa, as well as a resource for imputation and fine-mapping. In addition to facilitating GWAS, resources are needed to interpret associations in African populations, given that many of these arise from population specific variation. Existing resources examining the transcriptomic landscape in African populations, specifically focusing on genetic variants associated with gene expression are inadequate. The African Genome Resource here presented encompasses approximately 3000 whole genome sequences, 600 transcriptomes and an infrastructure for computation and capacity development in the form of a new informatics centre in Uganda. The three elements of the AGR are as follows: The African Diversity Reference Panel (ADRP). Whole genome DNAseq data at 4x-8x depth from approximately 3,000 Africans of 14 distinct ethnicities, which is the largest and most diverse set of haplotypes for the continent comprising more than 90 million SNPs of which many (more than 20 million) are novel and unique to the continent. This large count of unique SNPs highlights the importance for the development of reference panels, SNP arrays and RNA sequencing resources specific to the continent. The African Transcriptomics Resource. RNAseq data from 6 different populations is the first large-scale resource to examine eQTLs within Africa, as well as diversity of splicing isoforms, and will help to identify novel transcripts, and exons across the genome. Such a resource will be invaluable for the interpretation of findings from large-scale GWAS being conducted in Africa, supporting initiatives such as the H3A consortium. The Uganda Medical Informatics Centre (UMIC). A computational facility in Entebbe enabling local storage and handling of data from across the continent and supporting long term capacity building. Together, these three tangible resources will enable substantial advances in medical genetics in Africa. The large number of private variants across the continent and within individual populations highlights the future need to further sequence indigenous populations across the continent.

1095F

Fine-scale population structure in France: Loire River as genetic barrier.

C. Dina[1,2], J. Giemza[1], M. Karakachoff[1,2], F. Simonet[1,2], K. Rouault[3], E. Charpentier[1,2], S. Lecointe[1,2], P. Lindenbaum[1], J. Violleau[1,2], H. Le Marec[1,2], C. Férec[3], S. Chatel[1,2], S. Hercberg[4], P. Galan[4], J-J. Schott[1,2], E. Génin[3], R. Redon[1,2]. 1) Thorax Inst, INSERM-CNRS, Nantes, France; 2) CHU Nantes, Nantes University; 3) Inserm UMR 1078, CHRU Brest, University Bretagne Occidentale, EFS, Brest France; 4) Université Paris 13, Equipe de Recherche en Epidémiologie Nutritionnelle, Centre de Recherche en Epidémiologie et Statistiques, Inserm (U1153), Inra (U1125), Cnam, COMUE Sorbonne Paris Cité, F-93017, Bobigny, France.

Background The genetic structure of human populations varies throughout the world, being influenced by migration, admixture, natural selection and genetic drift. Human population structure has first been investigated at broad scales, between and within continents. Currently researchers focus on finer scales, examining genetic structure within countries. Characterising such genetic variation is of interest as it provides insight into demographic history and informs research on disease association studies, especially on rare variants. We here explored the genetic structure of a population living on the French territory (hereafter called French population) both on the whole territory and then on Western part where interesting stratification was identified.   Methods and Results We genotyped genome-wide ; 2276 individuals with known department of origin from French Population (SU.VI.MAX study) using Illumina Chip; 456 individuals (PREGO study) from Western France Atlantic Coast, from *Finistère* to *Vendée*, with at least three of their grandparents born within a 15 kilometres distance using Axiom CEU Chip. With EEMS software we visualised areas with low effective migration rates - the migration barriers, which match with geographical features, with particularly strong barrier on the lower course of Loire in Western France. We then focused on the PREGO study and Principal Components analysis revealed that individuals from the same departments form clusters. In both datasets we observed a high correlation between geographical position and components (p-value < 2e-16). Many independent methods support the hypothesis that Loire River is a genetic barrier. The two groups of individuals, from north or south of Loire, are well differentiated along PC1 axis. ADMIXTURE estimated different ancestry proportions for the two groups. The first split of hierarchical clustering returned by fineSTRUCTURE, and the one based on normalized counts of identity-by-descent segments is between north and south of Loire.   Conclusion   We here report genetic stratification at the level of continental French territory. The migration pattern is following the geographical structure. A specific pattern is noticed around the Loire River. We confirm both evidence for isolation by distance and existence of a genetic barrier, the Loire River. The discovered fine-scale population structure may have consequences in association analyses, especially for rare variants which tend to be geographically clustered. .

1096W

Genomic diversity and affinities in major population groups of Punjab, India: An analysis of microsatellite loci.

M. Kaur[1], B. Doza[2]. 1) Viral Research and Diagnostic Laboratory, Government Medical College, Amritsar, Punjab, India; 2) AMU Murshidabad Centre, Jangipur Barrage, Murshidabad, West Bengal, India.

The present study is motivated by the fact that negligible study is available regarding genetic structure of Punjabi population. In purview of this the present study was conducted to find out the genetic similarity and phylogenetic position of five endogamous population groups, namely Jat Sikh, Mazhbi Sikh, Brahmin, Ramdasia and Muslim from the Northwest districts of Punjabi population, with respect to admixture of other Indian and international populations in the past history. A total of 1246 subjects, including 252 Jat Sikh, 255 Mazhbi Sikh, 248 Brahmin, 251 Ramdasia and 240 Muslim samples were analyzed for six autosomal microsatellite markers (THO1, TPOX, CSF1PO, vWA, D7S820 and FGA). Statistical analyses were performed to interpret allele frequencies, heterozygosities, gene diversity, analysis of molecular variance (AMOVA), phylogenetic analysis (neighbor joining (NJ) method) and multidimensional scaling (MDS). The average heterozygosity for each locus in each population as a whole was substantial. Significant deviations from HWE were observed in almost all the populations and all the loci. In the entire analyses the Brahmin and the Mazhbi Sikh formed same cluster, while Jat Sikh and Ramdasia were placed close to their cluster, whereas, Muslim population group maintained the genetic distance from all the other population groups. Jat Sikh, Mazhbi Sikh, Brahmin and Ramdasia groups formed a clear and distinct separate cluster when compared with other Indian populations. The present study has also examined the affinities of Northwest Punjabi population with international populations, and some elements of central Asia, Middle East, Caucasoid and European populations were evident among the present studied groups. In conclusion**,** overall, a low level of genetic differentiation was observed in the studied population groups, especially, Jat Sikh, Mazhbi Sikh, Brahmin and Ramdasia indicating that genetic drift might have been small or negligible in shaping the genetic structure of the Northwest Punjabi populations.

1097T

The genetic structure of the world's first farmers. *I. Lazaridis[1,2], D. Nadel[3], G. Rollefson[4], D.C. Merrett[5], N. Rohland[1], S. Mallick[1,2,6], D. Fernandes[7,8], M. Novak[7,9], B. Gamarra[7], K. Sirak[7,10], S. Connell[7], K. Stewardson[1,6], E. Harney[1,6,11], Q. Fu[1,12,13], G. Gonzales-Fortes[14], S. Alpaslan Roodenberg[15], G. Lengyel[16], F. Bocquentin[17], B. Gasparian[18], J.M. Monge[19], M. Gregg[19], E. Vered[20], A.-S. Mizrahi[20], C. Meiklejohn[21], F. Gerritsen[22], N.A. Hovhannisyan[23], N. Patterson[2], R. Pinhasi[7], D. Reich[1,2,6].* 1) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; 3) The Zinman Institute of Archaeology, University of Haifa, Haifa 91905, Israel; 4) Dept. of Anthropology, Whitman College, Walla Walla, WA 99362, USA; 5) Dept. of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada; 6) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA; 7) School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland; 8) CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal; 9) Institute for Anthropological Research, Zagreb 10000, Croatia; 10) Dept. of Anthropology, Emory University, Atlanta, Georgia 30322, USA; 11) Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge 02138, USA; 12) Dept. of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany; 13) Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China; 14) Dept. of Biology and Evolution, University of Ferrara, Ferrara I-44121, Italy; 15) Independent Researcher, Santpoort-Noord, The Netherlands; 16) Department of Prehistory and Archaeology , University of Miskolc , 3515 Miskolc-Egyetemváros , Hungary 17) Centre National de la Recherche Scientifique, 75794 Paris, France; 18) Institute of Archaeology and Ethnology,National Academy of Sciences of the Republic of Armenia, 0025 Yerevan, Republic of Armenia; 19) University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA 19104, USA; 20) Israel Antiquities Authority, Jerusalem 91710, Israel; 21) Dept. of Anthropology, University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada; 22) Netherlands Institute in Turkey, Istiklal Caddesi, Nur-i Ziya Sokak 5, Beyoğlu, Istanbul, Turkey; 23) Dept. of Ecology and Nature Protection, Yerevan State University, 0025 Yerevan, Republic of Armenia.

We report the first genome-wide DNA from the ancient Near East, including some of the world's first farmers, their hunter-gatherer predecessors, and their successors. The early farmers of the Levant and of Western Iran were highly differentiated and each descended from local hunter-gatherers. All ancient Near Eastern populations derived part of their ancestry from a 'Basal Eurasian' lineage that had less Neanderthal ancestry than other non-Africans. Over time, Near Eastern farmers admixed with each other and with the hunter-gatherers of Europe to transform West Eurasia from a region of high genetic differentiation into one of relative homogeneity by the Bronze Age. We find that many populations outside West Eurasia also trace part of their ancestry to some of the same sources as West Eurasians: East Africans to the Levant, East Asians to north Eurasia, and South Asians to populations related to the ancient peoples of both Iran and the Eurasian steppe.

1098F

The paternal demographic history of African-derived Brazilian *quilombo* **populations inferred by high density SNPs array.** *R. Mingroni-Netto[1], L. Kimura[1], K. Nunes[1], L.I. Macedo-Souza[1], J. Rocha[2], D. Meyer[1].* 1) Dept. of Genetics and Evolutionary Biology, University of São Paulo, São Paulo , Brazil; 2) Dept. of Biology, Sciences Faculty, University of Porto, Portugal.

Quilombo remnants are relics of communities founded by runaway or abandoned slaves, but often with subsequent extensive and complex admixture patterns with European and Native Americans. In the present study we combine genetic studies and anthropological surveys in order to obtain a portrait of *quilombo* population structure and history, a region that gathers the largest number of *quilombo* remnants in São Paulo State. Samples from 289 individuals from *quilombo* remnants were genotyped using a set of 17 microsatellites (AmpFlSTR-Yfiler). A subset of 82 samples was also genotyped using SNPs array (Axiom-Affymetrix). We estimated haplotype and haplogroup frequencies; haplotype diversity and sharing; and pairwise genetic distances through F_{ST} and R_{ST} indexes.We identified 95 Y chromosome haplotypes, classified into 15 haplogroups. The most common were: R1b (European, 34.2%), E1b1a (African, 32.3%), J1 (European, 6.9%) and Q (Native American, 6.2%). Genetic differentiation among communities was high (F_{ST}=0.0983; R_{ST}=0.0768), and haplotype sharing was extensive. Available data (genetic, genealogical and oral survey) allows us to detect five main founder haplotypes, which explain a total of 27.7% of the Y chromosome lineages in *quilombos*. Our results showed a high European patrilinear genetic contribution among the founders of *quilombos*, high amounts of gene flow and a recent common origin of these populations. Predominant haplotypes and genealogical data indicate a pattern of origin of *quilombos* from few male individuals. As such, our study reinforces the importance of a dual approach, including both anthropological and genetic data. .

1099W

Genetic architecture of five endogamous groups from Punjab, Northwest India – A study based on autosomal markers. *G. Singh[1,2], I. Talwar[1], K. Matharoo[2], A.J.S. Bhanwer[2].* 1) Department of Anthropology, Panjab University, Chandigarh, India; 2) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India.

India represents a pool of culturally and genetically diverse populations. The region of Punjab (India) has an extreme importance in understanding the population diversity of India as it acted as a passage to the foreign invaders from Eurasia and Central Asia. The present study was designed to explore the genetic diversities and affinities among the five major endogamous groups using autosomal markers. A total of 1021 unrelated samples belonging to Banias, Brahmins, Jat Sikhs, Khatris and Scheduled Castes were genotyped for four Ins/Del (*ACE, APO, PLAT, D1*) and six restriction fragment length polymorphisms [*ESR (PvuII), LPL (PvuII) T2 (MspI*) and three *DRD2* (Taq1A, Taq1B, Taq1D)]. All the loci were found to be polymorphic among the studied populations. The average heterozygosity among the studied populations is fairly substantial ranging from 0.3816 in Banias to 0.4311 in Scheduled Castes. The genetic differentiation among studied populations ranged from 0.0418-0.0022 for the *PLAT* and Taq1A loci, respectively, with an average value being 0.0136. Phylogenetic analysis revealed that Banias and Khatris are genetically closest to each other. The Jat Sikhs are genetically close to Brahmins and are distant from the Banias. The Jat Sikhs, Banias, Brahmins and Khatris are genetically very distant from the Scheduled Castes. Overall, a uniform allele frequency distribution pattern and low level of genetic differentiation was observed in the studied population groups indicating that genetic drift might have been small or negligible in shaping the genetic structure of North-West Indian Populations. **Keywords**: *Alu* Insertion/Deletion, North-West India, Genetic diversity, endogamous group, SNP.

1100T

Genetic structure in the Sherpa and neighboring Nepalese populations. *A.M. Cole[1], S. Cox[2], C. Jeong[3], N. Petousi[4], D. Aryal[5], Y. Droma[6], M. Hanaoka[6], M. Ota[7], N. Kobayashi[7], P. Gasparini[8,9], H. Montgomery[2], P. Robbins[4], A. Di Rienzo[3], G.L. Cavalleri[1].* 1) Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Irelnad, Dublin 2, Ireland; 2) Institute for Human Health and Performance, University College London; 3) Department of Human Genetics, University of Chicago; 4) Department of Physiology, Anatomy and Genetics, University of Oxford; 5) Paropakar maternity and Women's hospital, Thapathali , Kathmandu, Nepal; 6) First Department of Medicine, Shinshu University School of Medicine, Matsumoto, Japan; 7) Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Japan; 8) University of Triests, Italy; 9) Division of Experimental Genetics, Sidra, Doha, Qatar.

Introduction: Nepal, located on the southern slope of the Himalayan arc, has an intricate demographic history. Today Nepal is an ethnically diverse country with 125 recognised ethnic groups, including the Sherpa, whom migrated from Tibet 400-600 years ago. **Aim and Methods:** We set out to describe the fine-scale population structure across the eastern region of Nepal using dense genotype data. We sampled five indigenous Nepalese ethnic groups, and the Nepalese Sherpa from eight villages in the Khumbu region of Nepal. We complemented this dataset with dense genotype data representing a variety of different populations resident across the greater Himalayan region including Tibet, China, India, Pakistan, Kazakhstan, Uzbekistan, Tajikistan and Kirghizstan, establishing a cohort of 1117 individuals. Using this dataset, we performed analysis of principal components, admixture, F4 ratio estimation, D-statistic and homozygosity. **Results and Conclusion:** We identified clear substructure across populations resident in the Himalayan arc. The genetic structure broadly mirrors the geographical features of the region. Ethnic subgroups within Nepal show distinct genetic structure. We detected differential proportions of ancestry from northern Himalayan populations across Nepalese subgroups, with the Nepalese Rai, Magar and Tamang carrying the greatest proportions of Tibetan ancestry. The Sherpa appear as a remarkably isolated population, with little gene flow from surrounding Nepalese populations. We confirmed the presence of an ancestral component that appears specific to high altitude populations of the Himalayas and is enriched in the Nepalese Sherpa, with particularly high proportions identified in individuals from a specific Sherpa village in the Khumbu valley of eastern Nepal.

1101F

Contributions of LINE-1 retrotransposon ORF0 to genomic output in the primates. *A.M. Denli, M. Pena, A. Saghatelian, F.H. Gage.* Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, CA.

What are the molecular underpinnings of diversity and evolution? The answer to this question has important implications for human health and biology: variations in disease predisposition, response to pathogens and drug therapies play roles in lifespan and quality of life. A comprehensive molecular understanding of diversity and evolution requires the availability of population-level genomic sequences, well-annotated transcriptomes and proteomes, as well as experimental models for functional genomics. This talk will summarize our recent efforts in the latter two. Transposable elements (TEs) are mobile genetic elements that can alter their chromosomal locations in the host genomes and occupy nearly 45% of the human genome. It has been estimated that ~2,000 common TE polymorphisms exist in human populations. LINE-1s, as the sole autonomously active retrotransposons in humans, are a major class of TEs that diversify our genomes. We have recently shown that primate LINE-1 5'UTR contains a primate-specific open reading frame (ORF) in the antisense orientation that we named ORF0. This ORF is present in more than 3,000 loci across human and chimpanzee genomes and forms fusion proteins with proximal exons, leading to insertion-specific novel contributions to the proteome. ORF0 expression is elevated in pluripotent stem cells as well as a subset of cancers. Our most recent analyses suggest that transcribed ORF0 polymorphic loci exist in human genomes and, in addition to the direct proteomic contributions, ORF0 also influences gene expression. The influence of TEs on the host cells will be discussed.

1102W

Conserved microRNA targeting of dosage-sensitive genes on mammalian X and avian Z chromosomes. *S. Naqvi[1,2], D.W. Bellott[1], D.C. Page[1,2,3].* 1) Whitehead Institute, Cambridge, MA; 2) Department of Biology, Massachusetts Institute of Technology, Cambridge MA; 3) Howard Hughes Medical Institute, Whitehead Institute, Cambridge MA.

The mammalian X and Y chromosomes arose from a pair of autosomes. Genes originally present on these autosomes evolved independently along a multistep path characterized by Y gene loss and the subsequent acquisition of X chromosome inactivation (XCI) in females. We assessed repression by microRNAs (miRNAs) to identify dosage constraints that led to three stable evolutionary fates along the pathway. X-linked genes at the beginning (with a surviving Y homolog) and end (with no Y homolog and subject to XCI) of the pathway have been under greater pressure to maintain miRNA targeting than evolutionary intermediates (with no Y homolog but escaping XCI). Such pressure has also acted upon avian Z-linked genes with surviving homologs on the female-specific W chromosome. We conclude that in addition to sex chromosome-wide mechanisms, purifying selection maintained ancestral X- or Z-linked gene dosage though retention of miRNA targeting. We propose that a combination of upward and downward constraints on ancestral dosage determined the evolutionary fates of sex-linked genes..

1103T

Key differences and similarities between human and mouse genes that escape X inactivation. *A. Slavney[1,2], A. Keinan[2], A. Clark[1,2].* 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

In eutherian mammals, X chromosome inactivation (XCI) silences transcription from one ChrX homolog in each female cell to equalize ChrX gene dosage between XX females and XY males. However, some genes are never fully inactivated, and are termed XCI escapers. XCI escape introduces variation in gene expression level between females and males, making these genes intriguing candidates for involvement in the many complex diseases that show substantial sex biases. Approximately 15% (n~200) and 7% (n=38) of human and mouse ChrX genes escape XCI, respectively. Analysis of the differences between human and mouse XCI escapers may elucidate the evolutionary drivers of XCI escape, and by extension the role of XCI escapers in normal and disordered phenotypes. Toward this end, we compared XCI status for orthologous human and mouse ChrX genes, and examined functional annotations and patterns of evolutionary conservation for XCI escapers in each species. Using a consensus XCI data set generated from 5 previous studies, we observed strikingly poor conservation of XCI status between human and mouse ChrX orthologs. In particular, only 13 orthologs escape XCI in both species. Additionally, several genes escaping XCI in one species have translocated to the autosomes in the other, consistent with the common hypothesis that dosage-sensitive genes are prone to escaping XCI. When we compared PANTHER gene ontology annotations of human and mouse XCI escapers, we observed high overlap in top biological process hits. However, the proportions of genes in the same functional categories differ greatly between species. In particular, human XCI escapers show a higher proportion of genes with metabolic functions than mouse XCI escapers. Lastly, we compared the distributions of exonic phyloP scores between XCI escapers and X-inactivated genes within each species to assess how conserved these gene groups are across vertebrates. In agreement with previous studies, we observed that human XCI escapers show a greater degree of sequence conservation than X-inactivated genes. While XCI escapers are also more conserved than X-inactivated genes in mouse, the difference is much less dramatic than in humans. Taken together, these results suggest that certain biological functions and strong evolutionary conservation are common features of XCI escapers in mammals, despite the fact that the XCI status of individual genes is highly lineage-specific.

1104F

Improvements in genotype imputation by using a population specific reference panel in Africa. *J. Chen[1], M. Sun[2], T. Carstensen[1], A. Morris[2], F. Pirie[3], A. Motala[3], M. Sandhu[1], M. McCarthy[2], I. Barroso[1], E. Wheeler[1], A. Mahajan[2].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridge, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) University of KwaZulu-Natal, Nelson R Mandela School of Medicine, Durban, South Africa.

Genome-wide association study (GWAS) has proven to be a powerful tool to detect disease/trait associated variants. An important step to boost power in GWAS is genotype imputation where untyped variants are inferred from the genotype data and known haplotypes in a reference panel. As discovery efforts in non-European ancestry populations increase, it has become popular to supplement existing imputation panels with population specific reference panels. In our study, we conducted a GWAS of type 2 diabetes (T2D) in African subjects of Zulu descent (1,602 cases/976 controls), and compared the imputation quality of two different imputation schemes. In scheme 1, we utilized the 1,000 genomes phase 3 reference panel (Panel 1) with uniformly distributed 1Mb chunks. Panel 1 was merged with 2,298 African samples with sequence data from the African Genome Variation Project (AGV) and the Uganda 2,000 Genomes Project (UG2G) into 1000Gp3+AGV+UG2G panel (Panel 2). In scheme 2, we used panel 2 with chunks containing 200 more variants in both the genotyped data and panel 2, avoiding large gaps and shorter than 5Mb. The imputation quality was evaluated in three aspects: number of variants imputed, imputation accuracy and imputation information score. Imputation accuracy was assessed using 97 Zulu samples with both genotype and sequence data. The sequence data was treated as the truth set and four metrics (genotype concordance, non-reference sensitivity, non-reference genotype concordance and precision) were applied to evaluate the similarity between the genotype data and the sequence data. As expected, 20,262,808 bi-allelic variants were imputed in scheme 2 only because of the additional African specific variants present in panel 2. When comparing variants common to both schemes, variants imputed in scheme 2 were more accurate (more similar to the sequence data) than those imputed in scheme 1 for all chromosomes and all minor allele frequency bins (common, low frequency and rare). We also observed an improvement in the imputation quality from the chunking strategy utilized in scheme 2, evident when considering Chromosome X, for which panel 2 is the same as panel 1. Overall, in our data, panel 2 both increased the number of imputed variants and improved the imputation accuracy and information score.

1105W

Inferring local demographic history and the age of mutations from haplotype cluster graphs. *G. Gliner[1], D.C. Aguiar[2], B.E. Engelhardt[2,3].* 1) Department of Operations Research and Financial Engineering , Princeton University, Princeton, NJ; 2) Department of Computer Science , Princeton University, Princeton, NJ; 3) Center for Statistics and Machine Learning , Princeton University, Princeton, NJ.

The age of a genomic mutation, as well as its frequency within a population, provides information about selective forces acting on a genomic locus, including susceptibility to disease, population structure, and demographic history. With the ability to rapidly sequence, genotype, and impute vast catalogues of genetic variation, developing novel data-driven statistical methods to estimate the age of a genetic mutation is possible. Here we present a statistical framework to estimate the age of genomic mutations across multiple individuals leveraging information contained in the haplotype cluster graph, a genetic model for haplotype sequences. Rather than relying singularly on population based statistics such as allele frequencies or attempting to reconstruct full ancestral recombination graphs, our approach explicitly models recombination and mutation deduced from the haplotype clusterings; this approach provides a much finer resolution of shared ancestry than methods that rely on genotypes or conservative linkage disequilibrium thresholding (e.g. 4-gamete test) in haplotypes. From the haplotype cluster graph, we generate marginal trees describing the ancestry of the extant subsample inheriting the mutation across unrecombined regions that, in turn, provide estimates for the shared physical and genetic distances and the number of mutations on each branch of the marginal tree following a coalescent statistical model to estimate the age of the locus. The novelty of our model originates from the ability to compute variant age estimates across multiple individuals and the synthesis with a robust statistical haplotype cluster graph model to estimate accumulation of ancestral mutations along ancestral branches and shared genetic and physical distances. We simulate whole-genome data with empirically estimated recombination patterns and compare results on inferring the age of mutations from a number of related methods showing both greater flexibility in application and more accurate estimates of age. Further, we apply our method to estimating mutational age in the 1000 Genomes Project data and show patterns of enrichment for specific functional categories..

1106T

Identification and characterization of common haplotypes found in a database of one million human genomes. *H. Guturu[1], K. Noto[1], J. Byrnes[1], S. Song[1], P. Carbonetto[1], R.E. Curtis[2], E. Elyashiv[1], J.M. Granka[1], E. Han[1], E.L. Hong[1], A.R. Kermany[1], N.M. Myres[2], K.A. Rand[1], Y. Wang[1], C.A. Ball[1], K.G. Chahine[2].* 1) Ancestry.com DNA, LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT.

Introduction: A common DNA-based method to detect relatives and ancestors ("cousins") is to identify and match shared portions of chromosomes (haplotype blocks) between an individual and their potential relatives. Identifying and matching the shared haplotype blocks is challenging due to the non-uniform halving of genetic information that takes place during the meiosis events of each generation. As the number of generations increases, the average size of matching haplotype blocks shrink, due to successive chromosomal recombination. Additionally, genetic drift, flow and selection establish population structure that skews the distribution of frequency and size of some haplotype blocks. We aim to characterize haplotype blocks based on their frequency profiles and link haplotypes to ancestral communities ("genetic ethnicities") and more recent admixed communities. **Methods:** Using a novel haplotype block matching algorithm, we identify haplotype blocks that occur frequently in a database of over one million samples genotyped by Ancestry.com DNA, LLC. We review the frequency profiles of each haplotype, and associate them with metadata inferred from global and local estimated admixture ("genetic ethnicity") as well as aggregated family history data from public family trees associated with some of the genotypes. **Results:** Common SNP windows have been characterized as identifying signatures of the gamut from ethnicities to more recent admixed communities resulting from migration. Further, we show that these signals of ethnic populations and communities can be used to improve the accuracy of identifying distant "cousin" matches by correcting for matches that are predominately generated due to more ancient signals of ancestry. **Conclusion:** By linking common haplotype blocks to ancestral groups of varying age of origin, we can improve the accuracy of ancestor identification for the desired task – ancient haplotype blocks for ethnicity admixture detection to more recent haplotype blocks that reflect recent cousins. Additionally, our characterization of haplotype blocks by ancestral groups reveals interesting candidates for further study and interpretation of their functional implications in various ethnic and community groups.

1107F

Maps of effective migration as a summary of human genetic diversity. *B. Peter, D. Petkova, M. Stephens, J. Novembre.* University of Chicago, Chicago, IL.

A dominant pattern of genetic diversity in humans is that geographically proximal populations are generally more genetically similar to one another; however, there are exceptions to this rule. Persistent geographical features such as mountains, oceans, or deserts, have allowed excess genetic differences to accumulate in some regions more than others. Conversely, historical migrations and population movements have led to cases where exceptional levels of similarity persist across large geographic distances. To provide more insight into how genetic differentiation is distributed geographically in humans, we examine the fine-scale genetic structure of humans. We produce maps that represent the spatial structure of human genetic diversity using a recently developed, spatially explicit method (EEMS, Estimation of Effective Migration Surfaces). We apply EEMS on global, continental, and sub-continental scales, analyzing genetic data from 8,740 individuals from 469 geographically localized populations, obtained from 24 different source studies. In addition to the major, well-known barriers such as the Sahara, Himalayas and Mediterranean, we detect barriers that correlate with historic language group boundaries (boundaries of Slavic and Bantu speakers with their neighbors), mountain ranges (Zagros, Caucasus, Ural) and marine features (English Channel, Adriatic Sea, Wallace line). We also identify regions showing high connectivity despite having geographic separation (Britain and Scandanavia, Iceland and Denmark, among the Lesser Sunda Islands). Simultaneously, we find that levels of diversity vary more smoothly, decreasing gradually with distance from Africa. Overall, our results suggest that diversity patterns are consistent and primarily shaped by the signature of the Out-of-Africa expansion, but that migration rates are strongly influenced by geography and local events.

1108W

Thousands of adaptive polymorphisms in human proteins and their prevalence in disease. *L. Scheinfeldt, R. Patel, T. Lanham, S. Kumar.* Temple University, Philadelphia, PA.

Hundreds of thousands of polymorphic missense mutations have been discovered in the human genome. However, fewer than 100 of these polymorphisms are known to carry signatures of positive selection. If true, this would make the incidence of adaptation through protein sequence changes to be an extremely rare phenomenon. Through an evolutionary time-series analysis, which uses between-species differences to generate neutral expectations and discover candidate adaptive polymorphisms (*caps*) based on the discordance between neutral expectations and observed allele frequencies, we have discovered more than 18,000 missense *caps*. Using available genome-wide association data, we have validated >6,000 *caps* to be involved in phenotypic traits, which are further validated by their disease-association (>5,000 *caps*). Therefore, we have identified 100-times more *bona fide caps* (*bf-caps*) than known previously. Missense *cap*-containing proteins are over-represented in several biological processes (e.g. sensory perception, immunity, and metabolism), and disease associated *caps* are enriched for several auto-immune diseases (including lupus, multiple sclerosis, rheumatoid arthritis, and asthma). Our new multidimensional approach integrates inter-species, intra-species, and phenotypic information in one framework to discover adaptations that have shaped human phenotypic variability and the disease landscape.

1109T

Dynamics of chromatin accessibility and transcription factor binding in human and chimpanzee pluripotent stem cells. *I. Gallego Romero[1,2], S. Gopalakrishnan[3], Y. Gilad[4].* 1) School of Biological Sciences, Nanyang Technological University, Singapore, Singapore; 2) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore; 3) Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark; 4) Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Many human-specific traits have long been hypothesised to be driven by gene regulatory differences between ourselves and our close evolutionary relatives. To test this hypothesis, we have generated maps of genome-wide chromatin accessibility using ATAC-seq in induced pluripotent stem cell (iPSC) lines derived from 6 humans and 7 chimpanzees (Pan troglodytes, our closest living relative), and quantified patterns of transcription factor (TF) binding activity in nearly 240 million putative sites across 1,229 different TFs. As expected, we find that sharing of chromatin accessibility patterns between the two species is strongest near orthologous transcription start sites (orthoTSS, r^2 = 0.949), and decreases with distance from orthoTSS. Combining these results with RNA-sequencing data from the same cell lines we find that significant inter-species differences in chromatin accessibility near orthoTSS occur more often than expected ($p < 10^{-5}$) at differentially expressed genes. Similarly, when we focus on transcription factor binding patterns in the two species, we find that TF binding sites most likely to be bound in both species are preferentially located close to orthoTSS and tend to have high position weight matrix (PWM) scores ($p < 2.2*10^{-16}$). Intriguingly, some of the transcription factors with the most divergent inter-species binding patterns have been implicated in early developmental processes, suggesting that the differences we observe at the pluripotent stage might underlie other interspecies cellular-level, and potentially even organismal-level, differences between humans and chimpanzees. Taken together, our results suggest that changes in chromatin accessibility and transcription factor activity are a likely gene regulatory mechanism through which human-specific traits can arise.

1110F

Deep sequencing of the human MHC region reveals long structural variation of ancient origins. *A.Q. Fu[1,2,3], E.Y. Kim[4], Y.I. Li[2], M. Stephens[1,5], S. Wolinsky[4].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Statistical Science, University of Idaho, Moscow, ID; 4) Division of Infectious Diseases, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) Department of Statistics, University of Chicago, Chicago, IL.

The major histocompatibility complex (MHC) region in humans plays a central role in immunity and exhibits high genetic diversity, which makes it difficult to study, or even to sequence this region. Here we used a combination of targeted sequence capture and high-coverage pyrosequencing (Roche 454) to characterize genetic variation in 63 people. We captured 73% of this region of 4.75 Mb (chr6:28.70-33.45 Mb; hg19), and achieved good coverage (55X on average) across the region. Unexpectedly, de novo assembly of the sequencing reads (median 450 bp) revealed that, in addition to HLA genes, 30 people carried extra DNA segments (median 2 kb) that are more similar to the MHC of the great apes (chimp, gorilla, or orangutan) than to the eight standard major human MHC reference haplotypes. Most (76%) of the ape-like segments share the highest levels of sequence identity with gorilla HLA-A (Gogo-A). We confirmed the presence of these gorilla-like segments (GLSs) also in the DNA of individuals from the 1000 Genomes Project. The GLSs are present in 20 of our samples, and tend to co-occur with several human HLA-A alleles that are closely-related in phylogeny (A*30:01, A*31:01, A*32:01, A*33:01/02/03, and A*68:02; p<1e-3), indicating close proximity of the GLSs to HLA-A in the sequence. Interestingly, the GLSs harbor a previously characterized pseudogene HLA-BEL, the first two exons of which are highly similar to a functional allele (Gogo-A*05:01) in gorilla. Although HLA-BEL has been reported to be untranscribed in humans, we validated its transcription through cDNA sequencing in all three individuals carrying this pseudogene. We also found HLA-BEL-like reads in the RNA-seq data of at least 6 (2%) blood samples from the Genotype-Tissue Expression (GTEx) cohort, thus validating its expression in a more general population. Additionally, we found an unknown gene-like sequence that is <70% similar to pseudogene HLA-W, 10 kb downstream of HLA-A. We named this sequence HLA-WLK, but did not find evidence for its transcription. These ape-like segments suggest additional complexity in the evolutionary history of the MHC region. Whereas ancestral, functional copies of the GLSs likely existed before the human-gorilla split, it is difficult to determine when these additional segments arose near HLA-A. However, transcription of HLA-BEL suggests that its promoter has not yet eroded through genetic drift, while HLA-WLK may have been pseudogenized further in the past.

1111W

Repeated expansion of human amylase genes create multiple independent CNV series. *N.A.A. Shwan[1], S. Louzada[2], F. Yang[2], J.A.L. Armour[1].* 1) School of Life Sciences, University of Nottingham, Nottingham , United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge,United Kingdom.

The human amylase gene family is highly copy number variable, and the salivary (*AMY1*) and pancreatic (*AMY2A* and *AMY2B*) amylase genes encode the starch-digesting enzyme expressed in the salivary gland and pancreas, respectively. High *AMY1* copy number (CN) has been shown to be correlated with adaptation to human dietary starch intake, and low *AMY1* CN reported to be a predisposition factor to obesity. These findings have not been replicated independently, and reliable measurement methods and accurate structural characterisation of the region are important to address such findings. Here we show independent allelic series of amylase CN variants (CNV) in sub-Saharan Africans using high-resolution measurement methods and segregation analysis in trios. To confirm the haplotype compositions we used fibre-FISH (Fluorescence *in situ* hybridization), which unveiled some structural complexity difficult to observe using conventional methods. Our work showed rearrangements including multiple expansion of a unit containing one copy each of *AMY1*, *AMY2A* and *AMY2B*. Overall, our data demonstrated at least five independent rearrangements of pancreatic amylase genes, in which the region has undergone homologous and non-homologous rearrangements to create new haplotypes, some of which contain five copies each of the *AMY2A* and *AMY2B* genes. The structural features shared by fundamentally distinct lineages strongly suggest that the common ancestral state for the human amylase cluster contained more than one, and probably three, copies of *AMY1*. These additional rearrangements we have discovered in the pancreatic amylase genes suggest that the pancreatic amylase genes should be taken into account when evaluating the adaptive significance of variation in this gene cluster. Furthermore, our results suggest that the CNV typing procedures we have developed constitute an accurate, reliable and high throughput method for measuring amylase CN in a large set of samples.

1112T

Association of toll-like receptor-3 gene polymorphism in patients with dengue as compared to controls: A case-control study. *A. Singh, A. Jain.* King George's Medical University, Lucknow, Select a Country.

Background: Toll-like receptors (TLR) are expressed by immune cells engaged in signalling cascades that culminates in inflammatory and immune defence response. Therefore, we aimed to assess the association of rs3775291, rs3775290, rs3775296 in cases with Dengue as compared to controls in Indian population. **Material and methods:** It was a hospital based case-control study, 98 Dengue cases and 98 age and sex matched controls were recruited. Cases included were laboratory conformed cases of dengue virus (DV) infection, confirmed by Either NS-1 antigen positivity or DV-RNA positivity or Demonstrating antibody in sera. Children with no past and present viral infection history were included as control. We determined TLR3 genotypes by polymerase chain reaction and followed by sequencing using chain termination method (PCR-sequencing) using ABI 3130 genetic analyser. **Results:** In the present study, **i**ncreased risk was observed in rs3775291 polymorphism in cases as compare to control (TT genotype OR=5.74, 95 %CI =1.47-2.42, p = 0.0001) (CT+TT genotype OR=2.1, 95 %CI =1.28-3.80, p = 0.0001). Significantly higher risk was found in individuals carrying (CT+TT) genotype of rs3775296 polymorphism among cases and controls (OR=1.94, 95 %CI =1.13-3.31, p = 0.0001). No significant difference was observed in rs3775290 polymorphism of TLR gene. **Conclusion:** TLR3 gene polymorphism might confer host genetic susceptibility to Dengue in Indian population.   .

1113F

The African Genome Resource Project: Patrilineal and matrilineal inheritance through the Y chromosome and the mitochondrial genome. *F. Abascal, D. Gurdasani, T. Carstensen, M. Pollard, C. Pomilla, M. Sandhu on behalf of AGR investigators.* Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Background The Y chromosome and the mitochondrial genome are inherited from the paternal and maternal lines, respectively. The lack of recombination in the mitochondrial genome and in large part of the Y chromosome leads to evolution almost in isolation from the autosomal genome. As a result, the Y chromosome and the mitochondrial genome offer a unique perspective on human demographic processes. Y chromosome (Y-) and mitochondrial (mt-) haplogroups can be very informative about human origins, migrations and admixture, as well as about potential sex biases during these processes. Further characterisation of the diversity of Y- and mt-haplogroups within Africa is essential to understand human history. Here, we present the mitochondrial and Y chromosome diversity among ~5000 individuals from the African Genome Resource panel. **Methods** We predicted the mt- and Y-haplogroups for 4,990 individuals and 2,399 males, respectively, representing diverse ethno-linguistic groups from Ethiopia, Uganda, South Africa, Egypt, and 5 African populations sequenced within the 1000 Genomes project. Mitochondrial and Y haplogroups were predicted with Haplogrep and YFitter, respectively. We called the mitochondrial genome and the Y chromosome for each sample and reconstructed their phylogenetic relationships with FastML. **Results** We found evidence for Eurasian admixture among several populations across sub-Saharan populations. Eurasian mt haplogroups appeared in 23% of the Ethiopians and 0.8% of the Ugandans. No Eurasian mt haplogroups were detected for the Zulu and Nama. We identified 13% Ethiopians, 0.5% Ugandan, and 43% Nama/Khoe-Sans with Eurasian Y haplogroups. Eurasian admixture is prevalent in Ethiopia but it is not distributed homogenously. Whereas the Gumuz show no Eurasian haplogroups, the Amhara show the highest frequencies. Within the Nama/Khoe-San there is not a single Eurasian mitochondrial haplogroup but up to 43% of Eurasian Y haplogroups, revealing a strong sex bias (p=1e-12). Consistent with previous reports, the oldest haplogroups are found in highest frequencies within the Khoe-Sans. **Conclusions** We present the largest panel of mt and Y chromosome sequences across Africa, including highly diverse Khoe-San populations from South-Africa. Our findings suggest substantial variation in Y chromosome and mt haplogroups across Africa, and provide evidence for extensive Eurasian admixture among several populations across Africa.

1114W

Whole-genome sequence analyses provide new insights into the demographic history and local adaptation of African populations. *S. Fan[1], D.E. Kelly[1], M.H. Beltrame[1], M.E.B. Hansen[1], S. Mallick[2,3,4], T. Nyambo[5], S. Omar[6], D. Meskel[7], G. Belay[7], A. Froment[8], N. Patterson[3], D. Reich[2,3,4], S.A. Tishkoff[1,9].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; 3) Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 4) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA; 5) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dares Salaam, Tanzania; 6) Kenya Medical Research Institute, Center for Biotechnology Research and Development, Nairobi, Kenya; 7) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 8) UMR 208, IRD-MNHN, Musée de l'Homme, Paris, France; 9) Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA.

Africa is the origin of modern humans within the past 200,000 years. There are more than 2,000 ethnolinguistic groups in Africa, which encompass around one-third of the world's languages. To infer the complex demographic history of African populations and adaptation to diverse environments, we sequenced the genomes of 94 individuals from 44 indigenous African populations using high coverage Illumina sequencing technology. Phylogenetic analysis confirms that the San lineage is basal to all other modern human population lineages. The location of other African populations in the phylogenetic tree correlates with geographical location, with the exception of the Central Africa rainforest hunter-gatherer (RHG) populations, who group with Southern African populations. We characterize ancient African population structure by inferring the effective population size and divergence time between populations. A common population bottleneck for all African populations was observed at ~200 thousand years ago (kya), corresponding with paleobiological evidence for modern human origins. Since then, the San and RHG populations have maintained the largest effective population size compared to other populations prior to 10 kya. Using MSMC analysis, we infer that the San population split from the RHG and the East African Khoesan-speaking Hadza and Sandawe hunter-gatherers within the past 66-82 kya, suggesting these populations could have originated from a historically more widespread population of hunter-gatherers. By contrast, the San diverged from all non-Khoesan speaking populations ~100-120 kya The divergence times of Niger-Kordofanian, Nilo-Saharan and Afroasiatic speaking populations were within the past ~22 to 41 kya. In the RHG populations, the oldest divergence was found between Eastern and Western RHG at ~36-51 kya; the time of divergence of the western RHG populations was inferred to be ~12-18 kya. Based on the ADMIXTURE analysis, Niger-Kordofanian and RHG populations were pooled for analyses of natural selection. We observed signatures of positive selection at genes involve in muscle development, bone synthesis, reproduction, immune function, energy metabolism, cell signaling, and neural development. This work is supported by NIH grants 1R01DK104339-01, 1R01GM113657-01, and DP1 ES022577-04 to SAT. The sequencing was funded by the Simons Foundation (SFARI 280376) and the U.S. National Science Foundation (BCS-1032255) grants to DR.

1115T

Investigating the evolution of human metabolic traits using primate iPSCs. *S.C. Makohon-Moore[1], G.A. Wray[1,2,3].* 1) Biology Department, Duke University, Durham, NC; 2) Evolutionary Anthropology Department, Duke University, Durham, NC; 3) Center for Genomic and Computation Biology, Duke University, Durham, NC.

Humans differ from our closest living relatives in a variety of important metabolic traits including disease susceptibilities. Understanding the evolutionary basis and molecular underpinnings of these uniquely human traits has previously been hampered by the limited ability to perform experiments on and access to samples from non-human primates. The development of induced pluripotent stem cell (iPSC) lines allows us to control genetic background, design rigorous experimental manipulations and remove environmental variability. iPSCs can be differentiated into a variety of cell types allowing for the characterization of specific cellular traits and responses to environmental stimuli. We are using iPSCs to identify genetic, epigenetic and environmental effects on the evolution of physiology during human origins. In particular, we are focusing on adipocytes, the key component cell of fat, which allows for the identification of differences in molecular, cellular and metabolic traits. Changes in these traits were likely necessary during human evolutionary origins as shifts in metabolic function were of particular importance. These questions would remain impossible to answer using any other existing approach. Using adipocytes differentiated from human and chimpanzee iPSCs along with multi-omic approaches we can infer how specific metabolic pathways have changed throughout evolution. Our results demonstrate that stem cells provide valuable insights into key metabolic traits that distinguish humans from our closest relatives and may be important for human specific disease susceptibilities.

1116F

The evolutionary history of human-specific gene duplications. *P. Moorjani[1,2], B. Handsaker[2,3,4], F. Polleux[5], S. McCarroll[2,3,4], M. Przeworski[1,6].* 1) Department of Biological Sciences, Columbia University, New York, NY; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 5) Department of Neuroscience, Zuckerman Mind Brain Behavior Institute and Kavli Institute for Brain Science, Columbia University, New York, NY; 6) Department of Systems Biology, Columbia University, New York, NY.

While gene duplications are known to have led to evolutionary novelty in many species, the functional consequences of duplications that arose on the human lineage are unknown. Intriguingly, human-specific duplications (i.e., duplicates present in humans but not in other primates) appear to be enriched for genes expressed in the developing brain. A subset (notably *SRGAP2*, *ARHGAP11B*) has been shown to have important functions in spine maturation and neocortex expansion and may contribute to the risk of neuropsychiatric diseases. Motivated by these observations, we decided to use the high quality primate genomes now available, to systematically characterize the repertoire of human-specific duplications, and to learn about when and under what selective pressures the duplicates arose. As a first step, we inferred the genomic location and copy number of the ancestral and duplicated gene copies from high coverage, whole genome sequences for ~280 humans sampled across the world, two archaic humans (Denisova and Neanderthal), and samples of great apes (chimpanzees, gorillas and orangutans). After identifying a set of duplications that arose in humans after separation from chimpanzees and gorillas and that are fixed in our relatively large sample of humans, we applied an Approximate Bayesian Computation method to estimate the timing of each gene duplication event. Importantly, our dating method accounts for paralogous gene conversion between the gene copies, leading to significantly older dates than had previously been reported, in particular for *SRGAP2* that had been previously dated as coincident with the transition from *Australopithecus* to *Homo* and the beginning of the neocortex expansion. Together, these analyses allow us to gain a comprehensive picture of the chronology and evolution of human-specific duplications.

1117W

A critical review and reassessment of archaic admixture inference. *S. Gopalan, E. Atkinson, B.M. Henn.* Ecology and Evolution, Stony Brook University, Stony Brook, NY.

   The controversial question of whether humans and archaic hominins interbred over the past 100,000 years has been a focus of population geneticists and paleoanthropologists alike for decades and which could not, until recently, be definitively answered. The earliest studies of archaic hominin ancient DNA were of mitochondrial (mt) sequence, and these largely concluded that such interbreeding did not occur. However, theoretical work which involved extensive modeling and simulations of demographic scenarios were more mixed; some supported the conclusions drawn from the genetic data while others contended that even relatively high rates of interbreeding could not be ruled out from mtDNA evidence alone. Recent advances in DNA sequencing and ancient DNA technology in the last 10 years have allowed geneticists to test the archaic introgression hypothesis more thoroughly by using multilocus genetic data from both archaic hominins and *Homo sapiens*. The sequencing of the Neandertal genome led to the conclusion that introgression into anatomically modern humans did occur, and that modern 'Out of Africa' human genomes contain a small percentage of this archaic DNA (1%-4%). Further ancient DNA studies, including the sequencing of the Denisovan genome, have continued to support the occurrence of archaic admixture, which is now the consensus among geneticists. However, there continue to be inconsistencies in the results between studies, and the methods used to obtain them are often opaque to even specialized audiences. In this research, we critically review the statistical methods used to infer archaic admixture in modern humans, the timing of admixture, and estimates of species divergence. Additionally, we reanalyze estimates of archaic introgression after controlling for reference genome bias and differential coverage, and testing for contamination using monophyletic loci.

1118T

The phenomes, genomes, and microbiomes of thirteen indigenous populations from the Himalaya. *A. Jha[1,2], Y. Gautam[1], D. Bhandari[3], S. Tandukar[3], J. Sherchand[3,4], C. Bustamante[1,2,5].* 1) Stanford Center for Computational, Evolutionary, and Human Genomics (CEHG), Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Publich Health Research Laboratory, Institute of Medicine, Tribhuvan University Teaching Hospital, Kathmandu, Nepal; 4) Department of Microbiology, Institute of Medicine, Tribhuvan University Teaching Hospital, Kathmandu, Nepal; 5) Department of Biomedical Data Science, Stanford University, Stanford, CA.

   Understanding human histories and identifying functional variants in the human genome that influence human adaptation and heritable traits are some of the central goals in human genetics. More recently, studies have indicated that human microbiomes may have influenced human evolution and they continue to affect human health. Despite the progress in characterizing genetic and microbial diversity and their functions in humans, those in many human populations remain to be cataloged. The Himalaya, which extends 2,337 miles from Myanmar to Afghanistan, is geographically, culturally, and ethno-linguistically diverse. Despite being home to five hundred ethnic populations, relatively little is known about the genomic and microbial diversity in this region. We have, for the first time, collected human DNA samples along with phenotypic measurements of 13 traits (including height, weight, BP, heart rate, skin pigmentation, and bitter tastes), dietary information, and household characteristics from ~500 individuals across thirteen Himalayan ethnic groups, many of which are underrepresented even in anthropological literatures. We have also, for the first time, extracted microbial DNA from stool, saliva, dental crevice, and tongue from a subset of individuals across four populations (N > 10 for each tissue), including recently settled nomads, primitive (slash and burn) agriculturalists, and commercial agriculture. Here, using our preliminary data, we present a comprehensive characterization of the genomic, microbial, and phenotypic variation in diverse Himalayan populations and discuss the potential approaches using which we intend to understand their population histories, investigate the genetic basis of local adaptations, identify genetic variants that influence morphological variations. Comparison of microbial diversity across these populations may show shifts in microbial populations in the mouth and gut, which may reveal key insights into diet-mediated adaptations in our species.

1119F

Non-allelic gene conversion is ten times faster than the rate of point mutations in humans. *X. Lan[1,3], A. Harpak[2], J.K. Pritchard[1,2,3].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Biology, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

Gene conversion is the unidirectional transfer of genetic sequence from a "donor" region to an "acceptor". In one of its modes, non-allelic gene conversion (NAGC, also known as interlocus gene conversion), the donor and the acceptor are homologous sequences on the same chromosome. Despite the implication of NAGC as the cause of various genetic diseases, and its role in the concerted evolution of many human gene families, the rates and contributing factors of NAGC are not well-characterized. Recent gene duplications are of focal interest in studying NAGC as NAGC is contingent on high sequence similarity between donor and acceptor. Notably, NAGC events are expected to distort the genealogy of a gene family at an affected region. Here, we develop a machine learning tools to survey duplicate gene families across primates in search of such local genealogy distortions, and identify converted regions in 44% of duplicate gene families surveyed. In addition, we estimate the parameters governing NAGC in humans. We estimate a tenfold higher rate of NAGC than point mutations in humans, with a median NAGC tract length of 525bp. Finally, we quantify the effects of genomic features which determine NAGC rates, including GC content, methylation levels and homology between donor and acceptor sequences. This work improves our understanding of the mechanisms behind NAGC and of the role NAGC plays in shaping sequence evolution in humans.

1120W

Impacts of Neandertal-introgressed sequences on the landscape of human gene expression. *R.C. McCoy[1], J. Wakefield[2,3], J.M. Akey[1].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Statistics, University of Washington, Seattle, WA; 3) Department of Biostatistics, University of Washington, Seattle, WA.

Approximately 2% of the genome of each non-African modern human traces ancestry to Neandertals. Recent studies suggest that Neandertal-inherited haplotypes contribute to variation in modern human traits, yet the mechanistic bases by which these phenotypes manifest remain poorly characterized. To this end, we leveraged RNA-seq data from GTEx (214 individuals and 52 tissues) to systematically quantify the regulatory impacts of introgression revealed by patterns of allele specific expression (ASE). We developed a Bayesian generalized mixed model approach to combine information across individuals and tissues and quantify allelic effects at single nucleotide polymorphisms (SNPs) that tag Neandertal-introgressed haplotypes. A total of 5055 tag SNPs were identified in 2034 genes and compared to 581,124 non-introgressed control SNPs in 26,437 genes. At a false discovery rate of 10%, 1236 introgressed SNPs (24.5%) exhibited significant ASE, as compared to 161,590 non-introgressed SNPs (27.8%). Stratifying by allele frequency to account for power differences, the proportion of introgressed SNPs exhibiting ASE and the distribution of allelic effects were comparable to non-introgressed SNPs. Several introgressed variants previously associated with human phenotypes showed evidence of ASE, suggesting regulatory mechanisms driving these associations. Notable examples include rs3765107 and linked SNP rs1385374 (r^2 = 1.0) in *SLC15A4*, which have a robust association with systemic lupus erythematosus risk and show strong down-regulation of the Neandertal allele. Patterns of ASE varied across tissues [χ^2(df=51) = 281.3, $P < 10^{-10}$], potentially reflecting differences in the rates of regulatory divergence. Driving this signal, we observed significant down-regulation of Neandertal alleles in brain tissues compared to non-brain tissues (β = -0.0168, $P < 10^{-10}$). One representative gene with significant down-regulation of the Neandertal allele is *NTRK2*, which encodes a neurotrophic tyrosine receptor kinase that regulates neuron survival and differentiation as well as synapse formation. Testis-expressed SNPs exhibited a similar pattern of down-regulation (β = -0.0145, $P = 0.0010$), which is particularly intriguing given that genes with high expression in the testes are significantly depleted of Neandertal ancestry. Together, our analyses reveal measurable impacts of Neandertal introgression on human gene expression that contribute to variation in modern human phenotypes.

1121T

Complex ancient genetic structure and cultural transitions in southern African populations. *F. Montinaro[1], G. Busby[2], M. Gonzalez-Santos[1], O. Oosthuizen[3], E. Ooosthuizen[3], P. Anagnostou[4], G. Destro-Bisol[4], V.L. Pascali[5], C. Capelli[1].* 1) Zoology, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK; 3) School of Medicine, University of Namibia, Windhoek, Namibia; 4) Dipartimento di Biologia Ambientale, Universit` a "La Sapienza", Rome, Italy; 5) Istituto Italiano di Antropologia 00185, Rome, Italy.

The characterization of the structure of southern Africa populations has been the subject of numerous genetic, medical, linguistic, archaeological and anthropological investigations. Current diversity in the subcontinent is the result of complex episodes of genetic admixture and cultural contact between the early inhabitants and the migrants that have arrived in the region over the last 2,000 years, with some of the variation present in the past being now lost as the result of cultural and demographic assimilation by surrounding populations. Here we analyze 1,856 individuals from 91 populations, comprising novel and available genotype data to characterize the genetic ancestry profiles of southern African populations. Combining local ancestry and allele frequency analyses we identify a tripartite, ancient, Khoesan-related genetic structure, which correlates with geography, but not with linguistic affiliation or subsistence strategy. The fine mapping of these components in southern African populations reveals admixture dynamics and episodes of cultural reversion involving several Khoesan groups and highlights different mixtures of ancestral components in Bantu speakers and Coloured individuals.

1122F

Evolutionary and functional plasticity of the nucleophilic elbow promotes novel activities in alpha/beta hydrolases. *J.J. O'Byrne[1], S-J. Reilly[2], M. Fares[3], M.C. Hunt[4], S.E.H. Alexson[2].* 1) Department of Clinical Genetics, Our Lady's Childrens Hospital Crumlin, Dublin 12, Ireland; 2) Karolinska Institutet, Department of Laboratory Medicine, Division of Clinical Chemistry, C1-74, Karolinska University Hospital at Huddinge, SE-141 86 Stockholm, Sweden; 3) Department of Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland; 4) School of Biological Sciences, Dublin Institute of Technology, Dublin 8, Ireland.

Enzymes are proteins that exhibit an astonishing diversity of functions, being a paradigm for the evolution of biological novelties. Many such enzymes emerged as a result of largely diversified gene families generated by gene duplication. Since organisms are well adapted to their environment, the mechanism that underlies the origin of novel, likely un-adapted, functions is not fully understood. In this study we explore the evolution of a small, highly conserved gene family belonging to the alpha/beta hydrolase superfamily of acyl-CoA thioesterases/acyltransferases expressed from bacteria to man. This gene family has evolved a number of different functions on a wide variety of substrates, including catalyzing the hydrolysis of acyl-CoAs to free fatty acids and CoASH (acyl-CoA thioesterases) and the conjugation (or amidation) of fatty acids or bile acids to taurine or glycine (acyl-CoA:amino acid *N*-acyltransferases). Phylogenetic, evolutionary and syntheny analyses of members of acyl-CoA thioesterases/acyltransferases, including acyl-CoA thioesterase 1 (*ACOT1*), bile acid-CoA:amino acid *N*-acyltransferase (*BAAT*) and acyl-CoA:amino acid *N*-acyltransferase (*ACNAT*), supports the hypothesis that these genes originated via duplication of a common ancestral gene, which was followed by the specialization of gene copies in the catalyses of distinct substrate-specific reactions. Structural and evolutionary analyses identified a common protein structure known as the nucleophilic elbow, which contains the active site nucleophilic residue, as the source of the enzymatic functional promiscuity. Our results point to the nucleophilic elbow as an ancestral reservoir of multiple coexisting functions, each of which became fixed after gene duplication in some gene copies, leading to the partition of ancestral functions. All members of this gene family have evolved subtle differences in the strand helix motif of their active sites that allows for the various activities and in part also substrate specificities to occur. The molecular engineering of the strand helix motif results in the redesign of these enzymatic functions. In summary this study neatly demonstrates "genomic enzymology", an expansive strategy for understanding the structural basis for functional plasticity and catalysis as well as the design principles nature employs to develop/evolve/redesign existing enzymes into new catalysts for new or even unnatural reactions.

1123W

Estimating the timing of adaptation for a beneficial allele. *J. Smith[1], M. Stephens[2,3], G. Coop[4,5], J. Novembre[2,1].* 1) Ecology and Evolution, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL; 4) Center for Population Biology, University of California-Davis, Davis, CA; 5) Department of Evolution and Ecology, University of California-Davis, Davis, CA.

A beneficial mutation carries linked information about its history which can shed light on the ecological context and putative cause for its increase in frequency. In particular, its age relies on the pattern of decay that mutation and recombination impose on its linked background. We provide a method to exploit this pattern and infer the time to the common ancestor of a positively selected allele following a rapid increase in frequency. We do so using a hidden Markov model which accounts for the accumulation of derived mutations on the ancestral background, its length distribution, and the surrounding haplotype diversity to infer the time to the common ancestor. Using simulations, we demonstrate how the inclusion of information from both mutation and recombination events increases estimate accuracy. We also show the behavior of estimation in cases where data do not conform to model assumptions, and provide some diagnostic routines for improving inference. Using the 1000 Genomes Project panel, we leverage population specific patterns to provide a global perspective on the timing of adaptation for several variants which show evidence of recent selection and functional relevance to diet, skin pigmentation, and morphology in humans. .

1124T

Whole-genome screen for disease-associated variants with extreme local allele frequency differences. *A. Sulovari[1,2], J. Hudziak[3], D. Li[1,4,5].* 1) Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT 05405, USA; 2) Cell, Molecular and Biomedical Sciences Graduate Program, University of Vermont, Burlington, VT 05405, USA; 3) Vermont Center for Children, Youth, and Families, Department of Psychiatry, University of Vermont, Burlington, VT 05405, USA; 4) Department of Computer Science, University of Vermont, Burlington, VT 05405, USA; 5) Neuroscience, Behavior, and Health Initiative, University of Vermont, Burlington, VT 05405, USA.

Variants with extreme allele frequency difference (EAFD) between two populations of the same continent may underlie human health disparities. To identify novel disease-associated EAFD loci, we systematically surveyed 81 million genomic variants from 2,504 unrelated individuals of 26 world populations (phase III of the 1000 Genomes Project). We mapped the identified EAFD variants to genes, and tested for enrichment of gene pathways, diseases and traits in each of five continental populations. Our results revealed known disease or trait-associated EAFD genes, such as *LCT* (lactose tolerance) and *SLC24A5* (skin pigmentation). More importantly, we identified novel EAFD genes, such as *GRM5* (synaptic plasticity), *TRIM40* (gastrointestinal inflammation), and *MICU1* (mitochondrial calcium uptake). We found several complex diseases influenced by EAFD, including alcohol consumption (adjusted $P = 0.0002$) and attention-deficit/hyperactivity disorder ($P = 0.003$). Our findings strongly support that variants with EAFD among populations contribute to genetic susceptibility of many complex human diseases.

1125F

Genome-wide detection of directional selection in Chileans, an admixed population. *L. Vicuña[1,2], F.I. Martinez[2], F. Crespo[1,2], S. Eyheramendy[1].* 1) Department of Statistics, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago 6904411, Chile; 2) Interdisciplinary Center for Intercultural and Indigenous Studies, Anthropology Program, Institute of Sociology, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago 6904411, Chile.

   The advent of genomics has brought along a radical transformation in the study of natural selection in humans, changing from a single-gene hypothesis-driven paradigm to a new paradigm that is exempted from a priori assumptions regarding candidate genes for selection. Most of the research so far has focused in native populations that have been geographically isolated for thousands of years, e.g. African, Inuit and Andean populations, and therefore exhibit clear genetic differentiation. There are, however, few studies reporting directional selection in recently admixed populations, where natural selection is more difficult to detect, and most of them have been done in African Americans and Latino populations from North America. The Chilean population represents an interesting model for population genetics studies. This population from South America is the result of an admixing process that started in the 16th century with the arrival of male Spanish conquistadores to the Chilean territory who mixed with Native American women, a process that continued with several minor migratory events (mainly from Europe) at later times. In this study, we aimed at detecting footprints of positive selection in the Chilean genome, represented by a cohort of 195 individuals, using a panel of 633744 autosomal SNPs. With this purpose, we used a slightly modified version of the Composite of Multiple Signals (CMS, described by *Grossman et al 2010*), a composite of 5 test statistics that account for three parameters of population genetics: the presence of long-range haplotypes (iHS, ΔiHH and XP-EHH tests), population differentiation (F_{ST} test) and the presence of high-frequency derived alleles (ΔDAF test). Through a rigorous analysis that included simulated genetic and demographic data, we found novel selected variants associated with genes responsible for important physiological and pathophysiological processes such as lipid metabolism, immunity and pathogen resistance. Hereby, we incorporated a novel demographic model for the peopling and admixture in the Chilean territory, which closely reflects real genetic data and historical records. The results of this study contribute to understanding the genetic basis underlying adaptation and disease, shed light into human demographic history and have implications in anthropology and medicine. .

1126W

Diversity across the pseudoautosomal boundary varies across human populations. *D. Cotter[1], T. Webster[1], M. Wilson Sayres[1,2].* 1) School of Life Sciences, Arizona State University, Tempe, AZ, USA; 2) Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, AZ, USA.

   Unlike the autosomes, recombination on the sex chromosomes is limited to the pseudoautosomal regions (PARs) at each end of the chromosome. PAR1 spans approximately 2.7 Mb from the tip of the proximal arm of each sex chromosome, and a pseudoautosomal boundary between the PAR1 and non-PAR region is thought to have evolved from a Y-specific inversion that suppressed recombination across the boundary. In addition to the two PARs, there is also a human-specific X-transposed region (XTR) that was duplicated from the X to the Y chromosome. Genetic diversity is expected to be higher in recombining than nonrecombining regions for many reasons, including that recombination reduces the effects of linked selection, allowing neutral variation to accumulate. We previously showed that diversity decreases linearly across the previously defined pseudoautosomal boundary (rather than drop suddenly at the boundary), suggesting that the pseudoautosomal boundary may not be strict as previously thought. In this study, we analyzed data from 1271 genetic females to explore the extent to which the pseudoautosomal boundary varies among human populations (broadly, African, European, South Asian, East Asian, and the Americas). We find that, in all populations, genetic diversity is significantly higher in the PAR1 and XTR than in the non-PAR regions, and that diversity decreases linearly from the PAR1 to finally reach a non-PAR value well past the pseudoautosomal boundary in all populations. However, we also find that location at which diversity changes from reflecting the higher PAR1 diversity to the lower non-PAR diversity varies by as much as 500 kb among populations. The lack of genetic evidence for a strict pseudoautosomal boundary and the variability in patterns of diversity across the pseudoautosomal boundary suggest that the boundary itself may vary across populations, or that population-specific demographic histories have shaped diversity spanning the pseudoautosomal boundary.

1127T

The Genome Diversity in Africa Project: A deep catalogue of genetic diversity across Africa. *D. Gurdasani[1,2], J.P. Martinez[1], M.O. Pollard[1,2], T. Carstensen[1,2], C. Pomilla[1,2], GDAP Investigators[1,2].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) Department of Medicine, University of Cambridge, Cambridge.

Background While recent efforts have greatly extended our understanding of genetic diversity in Africa, current sequence panels are limited in their capture of African genetic variation. Deeper sequencing with sampling of diverse indigenous populations is needed to capture diverse haplotypes across Africa. The Genome Diversity in Africa Project (GDAP) aims to characterise diversity from representative populations across all of Africa, including from several indigenous hunter-gatherer populations across the region. This would provide an important global resource to understand human genetic diversity and provide insight into population history and migrations across Africa in recent times. The project has completed sequencing of 575 samples across 23 populations in Africa, including populations from the Gambia, Ghana, Morocco, South Africa, Sudan, Chad, Kenya, South Africa, Uganda, Egypt and Ethiopia. Here, we present preliminary results from the project on 133 samples from 5 ethno-linguistic groups groups from Morocco, Ghana (Ashanti), Nigeria (Igbo), Kenya (Kalenjin) and South Africa (Zulu) sequenced on the Hiseq X platform (30x). **Methods** Reads were mapped to the GRCh38 reference. Following quality control, variant sites were called using HaplotypeCaller v3.5 for each sample to generate gVCFs. GenotypeGVCFs was run across all samples for joint calling. VCFs were filtered using VQSR calibrated on DP, QD, FS, SOR, ReadPosRankSum and MQRankSum annotations. A tranche sensitivity threshold of 99.5% was applied for filtering of SNPs and 99% for indels. Only sites called in >90% of individuals were included. **Results** We identified 25.1M SNPs and 2.9M indels among 133 individuals in the GDAP pilot phase, with 25% and 47% of SNPs and indels being novel (not in dbSNP141), respectively. A large proportion of variants per population were private, varying from 12-18%, being greatest among the Kalenjin and Zulu. We found the highest level of heterozygosity and genetic variation among the Zulu, consistent with reported Khoe-San admixture in this group. **Conclusions** We present the pilot phase of the Genome Diversity in Africa Project, identifying a high level of diversity across 5 populations from Africa. Inclusion of indigenous population groups, such as the Hadza, Twa Pygmies, and Ju/'hoansi in the next phase will materially advance the understanding of genetic diversity across African populations, and provide an invaluable resource to researchers worldwide.

1128F

Sequence diversity of *AZFc* genes on the human Y chromosome. *M.T. Oetjens[1], J.M. Kidd[1,2].* 1) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, 48109, USA; 2) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.

The AZFc region of the human Y-chromosome contains five multi-copy gene families arranged in long stretches of repetitive DNA termed amplicons. Amplicons introduce a susceptibility for amplification or deletion of the interstitial sequence through non-allelic homologous recombination. The AZFc gene families: *DAZ*, *RBMY*, *BPY2*, *PRY*, and *CDY* have critical roles in spermatogenesis, as large AZFc deletions associate with idiopathic male infertility. Recent studies have captured widespread structural diversity of Y-linked amplicons across global reference populations. Yet with respect to selection, the significance of structural diversity in Y-chromosomes is unclear. In contrast, nucleotide level variation of the Y-linked single-copy genes exhibit a strong bias against protein coding variation, suggesting strong purifying selection on the human Y-chromosome. However, little is known about AZFc nucleotide diversity and whether or not purifying selection of the AZFc has contributed to the low nucleotide diversity found across human Y-chromosomes remains an open question. To address this question, we called AZFc gene nucleotide variation in nine 1000 Genomes samples representing three haplogroup clades: E, O, and R. Here, we extracted reads within the boundaries of AZFc genes and psuedogenes with high identity and remapped them to intervals of a single copy where nucleotide variation between members was low and alignment quality was high. We filtered regions that exhibited extreme sequence depth, interspersed repeats identified by RepeatMasker, and indels identified in our alignment. Our callset includes 51,089 intronic, 2,852 non-synonymous, and 826 synonymous unique sites within the AZFc. To confirm the sensitivity of our analysis, we performed a sequence alignment of members within families of AZFc genes to identify single-family variants (SFVs) in the hg19 reference genome (haplogroup-R). Of 153 SFVs identified in hg19, only one was not seen and 40 are found to be polymorphic across our sample set. Expanding our results to non-reference SFVs, we in total call 223 intronic, 5 synonymous, and 2 non-synonymous polymorphic SFVs in the AZFc. We are currently applying this method to a globally diverse population of Y-chromosomes representing all major haplogroups and utilizing long-read sequencing technology to phase psuedogenes from fully functional genes to infer purifying selection of the AZFc region.

1129W

Searching for multiple sources of archaic introgression in African-American genomes. *S. Vattathil, J.M. Akey.* Genome Sciences, University of Washington, Seattle, WA.

When modern humans migrated out of Africa, they encountered sister hominin populations that had spread into Europe and Asia several thousand years prior. The presence of Neanderthal and Denisovan DNA segments within the genomes of contemporary humans of non-African descent is proof that ancient modern humans interbred with members of these closely-related but genetically distinct populations. Observations of introgressed fragments have now been collected from hundreds of individuals of diverse ancestry. These catalogs provide the basis to test models of the timing, location, and extent of the ancient interactions, and therefore refine our understanding of the prehistoric migration patterns of the human population. In addition, some of the introgressed alleles show strong evidence of selection, and therefore reveal variation important to human adaptation and evolution. Analogous interbreeding is reasoned to have occurred in Africa between ancient modern humans and local hominin populations, and we postulated that introgressed sequences persist in the genomes of contemporary individuals of African ancestry. We examined whole-genome sequence data from 61 African-American genomes for evidence of archaic introgression using the S* framework. Previous introgression studies using S* employed a 2-step approach, first identifying putative introgressed sequences using S* scores and then filtering this initial set based on similarity to the Neanderthal or Denisovan genome to derive a set of high-confidence introgressed sequences. Using this approach, we identify Neanderthal introgressed sequences in the European-derived chromosomes at a rate consistent with previous results, and effectively no Neanderthal introgression in African-derived chromosomes. Since using DNA sequence similarity is not an option for filtering for archaic African sequences (because no sequenceable archaic African DNA has been recovered), we explored other sequence characteristics such as divergence and haplotype length to define criteria for identifying likely introgressed sequences. In contrast to studies of Eurasians, the choice of reference population when using S* to study African chromosomes is not straightforward. We assessed the impact of using different reference populations.

1130T

Whole-genome reference panel of Tohoku Medical Megabank Organization (ToMMo) and allele frequency of pathological variants in the Japanese population. *Y. Yamaguchi-Kabata, Y. Kawai, K. Kojima, T. Mimori, F. Katsuoka, K. Kinoshita, S. Kure, Y. Suzuki, N. Fuse, H. Kawame, M. Nagasaki, J. Yasuda, M. Yamamoto.* Tohoku Medical Megabank Organization, Tohoku Univ, Sendai, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) have sequenced whole genomes of 1,070 cohort participants, and constructed the whole-genome reference panel (1KJPN) as a catalogue of genomic variants for foundation for genomic medicine in the Japanese population. We opened a website "integrative Japanese Genome Variation Database (iJGVD; http://ijgvd.megabank.tohoku.ac.jp/)", and publicly released allele frequency data of SNVs obtained from the 1070 individuals. We are annotating variants of 1KJPN with biological and medical information to identify variants having possible biological or pathological effects. By using a database of known pathological variants (the Human Gene Mutation Database, HGMD), we identified 4,368 HGMD variants (including 1,002 disease-causing variants) in 1KJPN. On the other hand, based on gene-based annotation, we identified loss-of-function variants including more than 3000 stop-gained SNVs. Among these stop-gained SNVs in 1KJPN, only a small proportion (4.5%) were annotated as known pathological variants, and biological effects are unknown for most of stop-gained variants. Individual variant load was calculated for disease-causing variants and stop-gained variants. The estimates were very similar with those in East Asian populations (Japanese in Tokyo, Han Chinese in Beijing, and Southern Han Chinese) in the 1000 Genomes Project (1KGP), therefore we did not find any signature of strong bottleneck in history of the Japanese population. However, incidence rates for inherited diseases vary among populations and it is known that incidence rates of inherited metabolic disorders were lower in the Japanese population compared to those of European ancestry. We compare allele frequencies of the pathological variants in 1KJPN with those in other population. In comparison to each of 14 populations of 1KGP, 2,638 HGMD variants showed significantly different allele frequency (P value<10⁻⁵; Fisher's exact test) in comparison to at least one population of the 1KGP. These differences in variant frquency between populations may partially explain the difference in the disease prevalence. .

1131F

Genetic diversity and founder effects in the Sea Island Gullah African Americans. *K.D. Zimmerman[1], S.P. Sajuthi[1], W. Chen[2], J. Divers[1], J.K. Fernandes[3], G.S. Gilkeson[3], K.J. Hunt[3], D.L. Kamen[3], U. Nayak[2], W.T. Garvey[4], M.M. Sale[2], C.D. Langefeld[1], P.S. Ramos[3].* 1) Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest University, Winston-Salem, NC; 2) Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Department of Medicine, Medical University of South Carolina, Charleston, SC; 4) Department of Nutrition Sciences and Birmingham VA Medical Center, University of Alabama, Birmingham, AL.

The Gullah form a unique population of African ancestry in the U.S., with lower European admixture (less than 11% on average) and higher ancestral homogeneity from the Sierra Leone area in Far-West Africa supported by historical evidence. Given their relative genetic and environmental homogeneity and ties to Sierra Leone, it is often questioned if the Gullah have reduced genetic diversity and increased frequencies of some disorders as the result of founder events and population bottlenecks. The goal of this study was to compare the genetic diversity of the Gullah and Sierra Leone populations in order to elucidate whether population bottlenecks and genetic isolation can account for the current genetic structure of the Gullah population. We used genotype data on 277 healthy Gullahs and 400 Sierra Leoneans (SL) to compute, for each population, the mean heterozygosity (HET) per SNP and inbreeding co-efficient (F) per individual based on expected and observed heterozygous calls, with $F = (HET_{exp} - HET_{obs})/(HET_{exp})$. After pruning SNPs in high LD ($r2 > 0.5$) with PLINK, the remaining 395K shared SNPs between populations were used to calculate the HET and F. A standard two-tailed t test was used to calculate the significance of the difference in HET and F between populations. To explore the amount of genetic variation within the Gullah and Sierra Leone (SL) populations, we first measured the mean heterozygosity. We observed a lower level of heterozygosity among Gullah individuals (HET=0.3326) compared with SL (HET=0.3336). This difference was statistically significant (P=0.036) even after pruning SNPs that were in high LD. This lower diversity in the Gullah compared with SL is paralleled with a higher inbreeding coefficient (F=-0.0018 in Gullah vs. F=-0.0045 in SL), which was very close to statistically significant (P=0.057). Collectively, these results suggest that the SL population is more outbred and genetically diverse than Gullah, and provide the first evidence supporting founder events in Gullah African Americans. Further analyses are underway to elucidate if the Gullah carry other evidence for founder effects. Understanding the genetic diversity of the Gullah contributes to a better understanding of the natural history and disease risks in this population and it helps inform future medical genetics research in African Americans.

1132W

The whole genome sequences and experimentally phased haplotypes of 182 personal genomes. *Q. Mao[1], S. Ciotlos[1], R.Y. Zhang[1], M.P. Ball[2,3], R. Chin[1], P. Carnevali[1], N. Barua[1], S. Nguyen[1], T. Clegg[2,4], A. Connelly[2,4], W. Vandewege[2,4], A.W. Zaranek[2,4], P.W. Estep[2], G.M. Church[2], R. Drmanac[1,5], B.A. Peters[1,5].* 1) Complete Genomics, Inc., 2071 Stierlin Ct., Mountain View, CA 94043; 2) Harvard Personal Genome Project, Harvard Medical School, NRB 238, 77 Avenue Louis Pasteur, Boston, MA 02115; 3) PersonalGenomes.org, 423 Brookline Avenue, #323, Boston, MA 02215; 4) Curoverse Inc, 212 Elm St, 3rd Floor, Somerville, MA 02144; 5) BGI-Shenzhen, Shenzhen 518083, China.

Background Since the completion of the Human Genome Project in 2003 it has been estimated that over 200,000 individual whole human genomes have been sequenced; a stunning accomplishment in such a short period of time. However, most of these were sequenced without experimental haplotype data and as such are missing an important aspect of genome biology. In addition, much of the genomic data generated is not available to the public and lacks phenotypic information. **Findings** As part of the Personal Genome Project (PGP), 182 participants' blood samples were collected and processed with Complete Genomics' Long Fragment Read (LFR) technology. Here we report the results for the experimental whole genome haplotyping and sequencing of these samples to an average read coverage depth of 100X. This level of coverage is approximately 3 fold higher than the read coverage applied to most whole human genome assemblies and was done to ensure the highest quality results. Currently 114 genomes from this data set are freely available through the PGP and *GigaScience* and are associated with rich phenotypic data. All 182 participants' genomes will be made freely available in the near future. 10 genomes were sequenced at least two times using independently made LFR barcoded libraries for reproducibility analyses. Additionally, 7 genomes were also sequenced using Complete Genomics' standard non-barcoded library process. **Conclusions** These genomes represent a unique source of haplotype and phenotype data for the scientific community, an order of magnitude larger than existing sets of experimentally phased genomes. In total, the dataset represents over 200 million phased variants in contigs with an average N50 of greater than 500kb. In addition, we report over 2 million high quality rare variants not previously identified in dbSNP or Phase 3 1000 Genomes Project (1KG) data. The data presented here should help expand our understanding of the human genome evolution and functioning.

1133T

High-coverage sequencing of the Human Genome Diversity Project (HGDP-CEPH) Panel. *S. McCarthy[1], A. Anders Bergström[1], Y. Xue[1], Q. Ayub[1], S. Mallick[2,3,4], M. Sandhu[1], D. Reich[2,3,4], R. Durbin[1], C. Tyler-Smith[1].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Howard Hughes Medical Institute, Boston, MA.

We discuss the completion of high coverage (>30x), whole-genome sequencing of all 952 core individuals in the Human Genome Diversity Panel (HGDP-CEPH), with the results being made available as an open access population data resource. This widely used panel contains samples from 52 populations spanning Africa, the Middle East, Europe, Asia, Oceania and the Americas, and previous genotype data from these samples have been an important reference resource for human genetic diversity. As seen in the 1000 Genomes Project, having fully open access data, unencumbered by managed access restrictions and other hurdles, is an invaluable driver for democratized data analysis and methods development Building on previous sequencing efforts by the Simons Genome Diversity Project, we have completed sequencing of the panel and are making the data available via the ENA and the 1000 Genomes Project data management successor, the International Genome Sample Resource (IGSR) (www.internationalgenome.org). All data has moved to the new GRCh38 reference and we present preliminary results on the call set derived from this data. We have GATK HaplotypeCaller and fermikit primary calls, are making mpileup and freebayes calls, and will present an integrated call set that has been computationally phased, together with initial population genetic analyses. A small number of samples are being experimentally phased using 10X Genomics technology which will allow evaluation of phasing accuracy, and also unbiased use of haplotype-based analyses such as MSMC.

1134F

Interleukins 4 and 13 gene polymorphism increases asthma risk: A meta analysis including 14,527 cases and 17,696 controls. *P. Dixit, S. Awasthi.* King George`s Medical University, Lucknow, India.

Background: Studies on the association of cytokines in asthma pathogenesis have yielded conflicting results. Therefore, we aimed to assess the association in north Indian children with asthma and include an updated meta-analysis. **Methods:** Odds ratios (ORs) with corresponding 95% confidence intervals (CIs) were computed to calculate the association. **Data Sources:** Pub Med (till July 2015) **Results:** Of total sixty nine case-control studies were included in this meta-analysis. Included were 14,527 cases and 17,696 controls. Significant association of IL13 -1112C/T and IL 4 C589T gene polymorphism was observed (OR = 1.12, 95% CI 1.01–1.25, P = 0.03) (OR=1.43, 95%CI= 1.13-1.81, P= 0.003) in a dominant genetic model. Significant association was observed in IL13 -1112C/T OR = 1.39, 95% CI 1.10–1.74, P=0.00) and IL 4 C589T (OR= 1.95, 95%CI=1.52-2.52, P= 0.00) polymorphism among Caucasians. Moreover, significant association of IL 4 C589T was also observed in African –American (OR= 1.61, 95%CI=1.17-2.22, P=0.00) population. **Conclusion:** IL 4-C589T and IL13 -1112C/T increases the risk of asthma development. .

1135W

Polymorphisms associated with European skin pigmentation in Oceanic populations. *I. Naka[1], N. Nishida[2], R. Kimura[3], K. Yamaguchi[4], T. Furusawa[5], T. Yamauchi[6], K. Natsuhara[7], T. Ishida[1], T. Inaoka[8], Y. Matsumura[9], R. Ohtsuka[10], J. Ohashi[1].* 1) Biological Sciences, The university of Tokyo, Tokyo, Tokyo, Japan; 2) Research Center for Hepatitis and Immunology, International Medical Center of Japan Konodai Hospital, Ichikawa, Japan; 3) Faculty of Medicine, University of the Ryukyus, Nakagami, Okinawa, Japan; 4) School of Natural Sciences and Psychology Liverpool John Moores University, U.K; 5) Graduate School of Asian and African Area Studies, Kyoto University, Kyoto, Japan; 6) Health Sciences, Hokkaido University School of Medicine, Sapporo, Hokkaido, Japan; 7) The Japanese Red Cross Akita College of Nursing, Akita, Akita, Japan; 8) Human Ecology, Faculty of Agriculture, Saga University, Saga, Saga, Japan; 9) Faculty of Health and Nutrition, Bunkyo University, Chigasaki, Kanagawa, Japan; 10) Japan Wildlife Research Center, Sumida, Tokyo, Japan.

The Oceania is geographically classified into three areas: Melanesia, Micronesia and Polynesia. Although Oceanic people are genetically close to Asians, their skin color is darker than Asians. In Oceania, people living in Melanesian have the darkest skin color, which is comparable to that of people living in Sub-Saharan Africa. To understand the genetic basis of skin pigmentation in Oceanic people, we investigated eight SNPs previously reported to be associated with self-reported tanning ability in European ancestry using Digitag2 and TaqMan assay in five Oceanic populations. The genotype data of YRI (African), CEU (European ancestry), JPT (East Asian), and CHB (East Asian) populations were obtained from the HapMap database. All the alleles associated with tanning ability were ancestral ones. The Fst analysis showed that some SNPs were highly differentiated between Melanesian and HapMap-YRI populations. We further calculated the tanning ability associated allele count and the tanning ability score ($\Sigma \beta_i x_{ij}$) for each individual. The number of allele counts was the highest in YRI and the lowest in CEU. Of particular interest, the mean of the tanning ability score is lower (i.e., lighter skin color) in Tongan (Polynesian) and Rawaki (Micronesia) populations than in HapMap-JPT and CHB (East Asian) populations. The present results suggest that the genetic variation that produces darker skin color of Oceanians is different from that of Asians.

1136T

Fine-scale identity-by-descent and birth records in Finland provide insights into recent population history. *A.R. Martin[1,2], S. Kirminen[3], A.S. Havulinna[4], A. Sarin[3], A. Palotie[1,2,3], V. Salomaa[4], S. Ripatti[3], M. Pirinen[3], M.J. Daly[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) National Institute for Health and Welfare (THL), Helsinki, Finland.

Finland provides unique opportunities to investigate both population and medical genomics because of its adoption of unprecedented uniformity in national electronic health records, concerted coordination of research centers across the country, detailed historical records, as well as recent population bottlenecks that drove specific disease alleles to high frequency. We investigate recent population history (up to ~50 generations ago), particularly relevant to rare, disease-conferring alleles, using identity-by-descent (IBD) haplotype sharing in >10,000 Finns. We compare IBD sharing in Finland to nearby Scandinavian countries with considerably different population histories, including >8,000 Swedes and >30,000 Danes. We find drastically more sharing on average in Finns, including many long tracts. By leveraging fine-scale birth record data, we find a non-linear decay of pairwise IBD sharing with increasing distance across Finland. This arises from pockets of excess IBD sharing; e.g. pairs of individuals from northeast Finland share on average several-fold more of their genome IBD than pairs from southwest regions containing the major cities of Turku and Helsinki. We demonstrate inference of recent migration patterns from IBD sharing patterns. For example, high IBD sharing in northeast Finland radiates from north to south rather than to the west, indicating that migration is restricted near the Russian border. We also investigate recent effective population size changes across regions of Finland and find evidence supporting the distinction between early and late settlement areas. However, our results indicate a more continuous flow of migration than previously posited, with a minimum N_e occurring ~12 generations ago in the northernmost Lapland region and moving further back in time to the south, with a bottleneck detectable in the early settlement area ~40 generations ago. Lastly, we leverage IBD sharing for genetic disease mapping and show that rare, functional haplotypes show more significant association via IBD mapping than single variants with linear mixed effect models.

1137F

Partitioning heritability and polygenic risk scores by hominid ancestry using genome-wide association summary statistics. *G. Athanasiadis, M.H. Schierup, B.J. Vilhjálmsson.* Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark.

  Researchers have previously attempted to describe the phenotypic legacy of Neanderthals to non-African anatomically modern humans using raw genomic data and a collection of relevant phenotypes. Here, we provide further insights into the Neanderthal contribution to the genetic architecture of 47 complex traits. We do so by calculating the proportion of heritability due to Neanderthal introgressed SNPs and the relative importance of those SNPs in polygenic risk scores. We also estimate the phenotypic impact of the Neanderthal introgressed SNPs, contrasting the predicting human phenotype with the predicted Neanderthal phenotype using LDpred polygenic risk scores. We base our analyses on summary statistics from large-scale publically available GWASs. We introduce a new statistic – conceptually similar to LD score – that uses the 1000 Genomes as reference panel to detect summary statistic inflation due to residual population structure. Using this statistic we can flag potentially biased heritability estimates and polygenic risk scores. Moreover, we account for the estimated summary statistic inflation due to residual population structure, arriving at more accurate results. The use of publically available GWAS summary statistics in our analysis allows for a more accurate assessment of the Neanderthal phenotypic legacy because SNP effects had been better estimated (i.e. the error of the effect is smaller.

1138W

Meiotic recombination is an important driver of sequence evolution at hotspots. *I. Tiemann-Boege[1], B. Arbeithuber[1], A. Heissl[1], A.J. Betancourt[2].* 1) Johannes Kepler University, Linz, Austria; 2) Vetmeduni Vienna, Vienna, Austria.

  Meiotic recombination, a key biological process during germ cell production, is clustered in recombination hotspots that undergo rapid sequence degeneration, as shown repeatedly by population data analysis. The exact mechanisms driving this sequence evolution are still not well known. By sequencing a large number of single recombinants obtained from human sperm, we have shown that recombination is mutagenic and crossovers are enriched for *de novo* CG to TA transitions, especially at methylated CpG sites, which could be the predominant mutational pattern in processes involving single stranded DNA. Another mechanism driving the sequence evolution at hotspots is biased gene conversion. Our large data set also provided new evidence that the transmission of GC-alleles is favored during crossing-over and showed that GC biased gene conversion (gBGC) is a strong driver of hotspot sequence evolution opposing mutation. We have also examined the evolution of short tandem repeats (STR) at hotspots by analyzing the transmission of polymorphic STRs and their effect on recombination. We observed that a long polymorphic tract of polyAs (A9/A19) located at the center of the hotspot can shift the hotspot center and reduce the overall crossover frequency. In addition, the transmission of STRs do not follow the Mendelian rules of segregation (50:50), but longer repeats are transmitted more frequently than shorter ones, and our data show for the first time with experimental evidence an insertion-biased gene conversion (iBGC) in a STR (A6/A7) within a hotspot. The molecular mechanisms inducing the observed patterns are not clear yet; however, the recurrent repair of double-strand breaks (DSB), required for the initiation of recombination, seems to play an important role.

1139T

Studies of canine breed development on the island of Sardinia recapitulate genomic features of human population isolates. *D.L. Dreger[1], B.W. Davis[1], R. Cocco[2], S. Sechi[2], A. di Cerbo[3], H.G. Parker[1], M. Polli[4], P. Crepaldi[4], E.A. Ostrander[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Dipartimento di Medicina Veterinaria, Sezione Clinica Medica, University di Sassari, Italy; 3) School of Specialization in Clinical Biochemistry, G. d'Annunzio University, Italy; 4) Dipartimento di Medicina Veterinaria, Universita di Milano, Italy.

The island of Sardinia has long been a focus for studies of human genetic disorders and complex traits due to the unique ancestral background of its inhabitants and the maintenance of population isolation due to restrictive geography as well as cultural and linguistic separation. Like most human population isolates, the inhabitants share decreased genomic diversity, increased linkage disequilibrium, and increased inbreeding coefficients. Canines have been the constant companions of humans since the earliest days of their domestication, with breeds often arising through human-driven selection of characteristics to meet an ideal standard of appearance and function. However, in certain regions, dogs and humans have been exposed to the same natural and human-driven forces of environment, migration, cultural development, industrial growth, population expansion, and geography. We therefore hypothesized that the events that result in isolated canine populations or breeds, would be mirrored in human populations from the same region. Sardinia is home to an endemic dog population, the Fonni's Dog, with documentation dating to the 19th century establishing the dogs as fierce protectors and guardians. The Fonni's Dog, however, has not been subjected to the intensive system of artificial selection of a typical dog breed, but rather has developed along side the human populations of Sardinia, influenced primarily by geographic isolation and unregulated selection based only on its adaption to the environment and aptitude for its owner-desired behaviors. Through analysis of whole-genome sequence and ~170K genome-wide single nucleotide variants, we have produced a genomic illustration of the Fonni's Dog in the context of 27 additional Mediterranean dog breeds. Patterns of homozygosity, relatedness, and private variation confirm within-breed similarity, while population structure and phylogenetic analyses provide spatial identity of Fonni's Dog to other Mediterranean breeds. Investigation of admixture and fixation indices reveal intriguing insights into the Fonni's Dog's involvement in breed development throughout the Mediterranean. We describe how characteristics of population isolates are reflected in dog breeds that have undergone artificial selection, and have been mirrored in the Fonni's Dog through traditional isolating factors that affect human populations. Lastly, we show that the genetic history of Fonni's Dog parallels demographic events in local human populations.

1140F

Patterns of germline mutation from exome sequencing of 8,000 parent-offspring trios. *J. Kaplanis[1,2], M. Hurles[1] on behalf of the DDD Study.* 1) Wellcome Trust Sanger Institute, Hinxton UK; 2) University of Cambridge, Cambridge UK.

Germline mutation is the fundamental root of evolution and disease causing variants. The characteristics of these mutations, and their flanking sequences, can reveal information about the source and timing of these events and improve our understanding of the mutational process We identified ~17,000 high confidence mutations in ~8,000 exome sequenced trios from the Deciphering Developmental Disorders (DDD) Study which consists of families with children with severe developmental disorders. De novo mutations (DNMs) are a major cause of developmental disorders. We phased these de novo mutations on to their parental haplotypes and built a statistical model to identify post-zygotic mutations. We found that ~3% of the de novo mutations arose post-zygotically in the child. We then explored parental effects with respect to differences in mutational rates and spectra. We found that the number of DNMs increases with both paternal and maternal age, with the paternal effect being considerably stronger. In general, the number of DNMs per child follows the expected Poisson distribution, however, I will also describe our investigation of outliers for potential germline mutator phenotypes within the DDD cohort. In summary, these analyses can help us build a better understanding of germline mutation which has important implications in genetic disease.

1141W

Comprehensive context-specific estimation of tandem repeat mutation rates and variant densities. *A. Tan, H.M. Kang.* University of Michigan, Ann Arbor, MI.

It is known that ~50% of all indels are found in the 2% low complexity regions of genome where most Indels are variable number of tandem repeats (VNTRs). Because VNTRs are highly affected by polymerase slippage dependent on sequence context, it has been challenging to precisely measure their mutation rates. Comprehensive measurement of indel mutation rates or variant density in a context-specific manner expands our understanding of mutational mechanisms and allows us to distinguish between germline tandem repeat variants and PCR or alignment induced variants. We produced a comprehensive catalog of context-specific tandem-repeats using a novel computational algorithm. Our method estimates the boundaries of repetitive regions of each VNTR by the left-right alignment of alleles to systematically classify sequence context of exact and inexact tandem repeats. Thus, we can estimate context-specific mutation rates or variant densities by comparing genome-wide count of specific repeat contexts with the observed number of (de-novo or population-based) variants and its allele frequency spectrum. We applied our method to characterize context-specific VNTR densities for NA12878 and other deeply sequenced genomes. We observed that poly-A stretches show the largest context-specific variant densities. Interestingly, the density of 2-bp A deletion in poly-A context is observed to be one third of 1-bp deletion in the same context, suggesting a higher likelihood of multi-nucleotide slippages. For most repeat-contexts, when the repeat tract length is 15bp or greater, there were >100-fold enrichment of variant density compared to non-repetitive regions. For moderate repeat lengths often ignored in tandem repeat studies (e.g. 5-10bp) we observed a significant >10-fold enrichment in variant densities across all repeat contexts, which was only 3-4-fold lower compared to 15bp repeats. We also observed striking under-representation of VNTRs in standard callsets like Genome-In-the-Bottle, reducing its utility as a measure for Indel sensitivity analysis. We find that filters exacerbate the inconsistencies between different callsets. For example, VNTRs in poly-A contexts are different between variant callers due to poly-A filtering, while poly-AAC VNTRs are more consistent between callers. Our approach provides another step towards understanding the complexities of calling the spectrum of indels. .

1142T

Unused Program Number

1143F

Pleiotropic noncoding regulatory elements are under purifying natural selection. *D. Radke[1,2,3], D. Balick[2,3], J. Sul[4], S. Akle[2,5], M. Maurano[6], R. Green[2,3], J. Stamatoyannopoulos[7], S. Sunyaev[2,3].* 1) Program in Genetics and Genomics, Harvard Medical School, Boston, MA; 2) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA; 5) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA; 6) Institute for Systems Genetics, New York University, New York City, NY; 7) Department of Genome Sciences, University of Washington, Seattle, WA.

   Assessing the role of natural selection on genetic variation in noncoding regulatory regions has been difficult up to the present, because most work has only utilized SNP variation, which has hampered functional interpretation because of the lack of clear loss-of-function (LoF) SNPs in noncoding loci. Genomic deletions, either partial or complete, provide a powerful LoF model in noncoding regions due to a deletion event removing the nucleotide component space altogether. To gain high-resolution deletion allele frequency information for population genetic analysis, we performed deletion discovery and genotyping from high-coverage WGS data within a cohort of >750 unrelated individuals, identifying ~15K unique deletions. Using regulatory annotations from primary tissues and cell-types characterized as part of the Roadmap Epigenomics Project, and accounting for covariates based on genomic regional context, we find a statistically shift in the allele frequency spectrum (AFS) towards reduced frequency of deletions overlapping noncoding regulatory loci versus deletions occurring in nonfunctional loci, indicating the action of purifying selection on regulatory elements. Since the functional outputs from the regulatory annotation data are derived from multiple distinct human tissues and cell-types of wide-ranging organismal function, yet share overlapping genomic coordinates, the overall functional epigenetic information present within a locus can serve as a proxy for the pleiotropy of the locus. We hypothesize that regulatory loci exhibiting highly-pleiotropic effects (i.e., loci with regulation across many diverse tissues) should be under stronger purifying selective pressure than cell-specific or nonfunctional loci. Analyzing the pleiotropy at each locus for each of the genotyped noncoding deletions, we find a very statistically significant shift in the AFS towards deletions being rarer which overlap highly-pleiotropic regulatory loci. We find that either the average pleiotropy or the maximum pleiotropy along the length of a deletion is sufficient to explain shifts in the AFS. We interpret these results as evidence of the action of purifying selection on regulatory elements, the strength of which is determined by the amount of pleiotropy. These findings open up the noncoding genome to more rigorous functional interpretation for use in medical or experimental studies.

1144W

CYP2D6 **allele frequencies in the Asian population.** *W. Chan[1], P.Y. Cheung[1], M.S. Li[1], S.K. Sundaram[1], B. Tomlinson[2], C.H. Tzang[1].* 1) Prenetics, Limited , Hong Kong; 2) The Chinese University of Hong Kong, Hong Kong.

   Background: Cytochrome P450 CYP2D6 plays a crucial role in drug metabolism. CYP2D6 is known to be highly polymorphic with enzymatic activity ranging from complete deficiency to ultra-rapid. The genetic variation of CYP2D6 has been associated with variability in drug responses. The frequency of CYP2D6 genetic polymorphisms varies significantly between different ethnic groups. While the frequencies of CYP2D6 alleles are generally known in Asian populations, data on frequencies of the copy number variations (CNV) in CYP2D6 and complex haplotype backgrounds in which they occur are less well studied in these populations. The objective of this study is to determine and report the observed CYP2D6 allele, genotype and derived phenotype frequencies in the Asian populations in the East and Southeast region. **Methods:** Genotyping was performed using Prenetics Limited (Prenetics) iGenes Pharmacogenomics Test (PGx) at Prenetics' ISO15189:2012 accredited laboratory using TaqMan genotyping, CNV, as well as digital PCR technology (Life Technologies, USA). CYP2D6 phenotypes of study subjects were then stratified into Ultra-rapid Metabolizers (UMs), Extensive Metabolizers (EMs), Intermediate Metabolizers (IMs) and Poor Metabolizers (PMs) based on predicted activity scores predicted from the genotypes. **Results:** Out of the 800 genotyped individuals, 735 (92%) were self-identified as Asians, 56 (7%) were Caucasians and 9 (1%) were mixed race. Among the Asian group, the prevalence rates of CYP2D6 alleles were as follows: CYP2D6*1 (23.5%), CYP2D6*2 (10.5%), CYP2D6*4 (0.5%), CYP2D6*5 (3.3%), CYP2D6*6 (0.2%), CYP2D6*10 (21.9%), CYP2D6*10-*36 Tandem Repeats (31.2%), CYP2D6*14A (0.4%), CYP2D6*14B (1.6%), CYP2D6*36 (4.1%), and CYP2D6*41 (2.7%). Duplicated CYP2D6 alleles were observed in 74 individuals (10.1%). The observed phenotype frequency of CYP2D6 UM, EM, IM and PM were 3.3%, 49.9%, 46.4%, and 0.4%, respectively. **Conclusions:** The present study provides a comprehensive analysis on different CYP2D6 alleles in an Asian population which is the largest study in terms of the number of individuals genotyped and the number of CYP2D6 alleles analyzed in this population. Our results contribute to the overall knowledge of pharmacogenomics specifically in the Asian population and may help the implementation of precision medicine in East and Southeast Asia.

1145T

Temporary pulses of accelerated mutagenesis in human and great ape population history. *K. Harris[1], J.K. Pritchard[1,2].* 1) Genetics, Stanford University, Palo Alto, CA; 2) Howard Hughes Medical Institute.

Like any complex physiological trait, the germline mutation rate has the potential to evolve as the result of selection and genetic drift. Understanding the dynamics of mutation rate evolution is technically challenging but fundamentally important to evolutionary biology. Present-day mutation rates can be estimated by counting de novo changes that distinguish parents from children, but ancestral mutation rates must be estimated via more uncertain phylogenetic methods. Here, we introduce a novel method for inferring recent mutation rate evolution from population genetic data, specifically by comparing the site frequency spectra of different mutation classes to determine whether the rates of any mutation classes have accelerated. We find several mutation spectrum changes that appear to have fixed early in primate history, for example, an A→T transversion rate increase in the human/chimp common ancestor. More surprisingly, we find several "mutation pulses" where a mutational process with a unique signature appears to have acted for a short time before subsiding. Such a pulse appears to have affected Europeans approximately 10,000 years ago, producing excess transitions in the context TCC→TTC.

1146F

Analysis of global ancestry, ethno-linguistics, and migration correlates ancestry with linguistic data. *J.L. Baker, D. Shriner, C.N. Rotimi.* Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Genetic studies have established a sub-Saharan African origin for anatomically modern humans with subsequent migrations out of Africa. Characterizing the genetic variation and relationships among global populations stemming from these events is challenging due to gene flow and admixture. Analysis of ancestry effectively removes the effects of recent instances of these confounding factors, revealing history in the distant past. Using the largest multi-locus data set for a study of genetic variation known to date, we investigated the genetic differentiation of early modern humans during global peopling, the extent of human admixture and migration events, and the relationships among ancestries and language groups. We compiled publicly available genome-wide genotype data on autosomal single nucleotide polymorphisms. The data set includes 5,966 individuals from 282 global samples genotyped at ~19K SNPs. Thirty of the world's 141 primary language families, accounting for 97.8% of people, are represented in this data set. The best evidence from these data supports 21 ancestries that delineate genetic structure of the global human population during the Upper Paleolithic, 17 of which are present in Africa. Independent of self-identified ethno-linguistic labels, we found that the vast majority (97.4%) of present-day individuals in this global dataset have mixed ancestry with evidence of multiple ancestries in 96.7% of samples and 100% of continents. We find evidence for migration events between Eastern and Northern Africa and between Omotic ancestry and the node leading to Northern European, Arabian, Northern African, Southern European, and Western Asian ancestries. We found a moderate to strong positive correlation between ancestries and language families or branches, such that ancestry data support or refute several proposed linguistic relationships and the linguistic data point to possible resolutions of heterogeneity in the ancestry data. Thus, ancestry data yield insight into a deeper past than linguistic data can, while linguistic data provide clarity to ancestry data. Finally, these data do not provide genetic support for the existence of distinct human racial groups. These data will be a public resource that can be used to describe genetic variation observed at multiple levels, from worldwide to continental to population to individual.

1147W

Y-chromosomal composition of mediaeval and contemporary populations in Norway and adjacent Scandinavian countries: Y-STR haplotypes and the rare Y-haplogroup Q. *B. Berger[1], S. Willuweit[2], H. Niederstätter[1], P. Kralj[1], L. Roewer[2], W. Parson[1,3].* 1) Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria; 2) Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin, Berlin, 13353, Germany; 3) Forensic Science Program, The Pennsylvania State University, PA, USA.

   In the framework of the project "Immigration and mobility in mediaeval and post-mediaeval Norway" molecular genetic analyses were performed on 97 pre-modern human remains including genetic sexing and Y-chromosomal DNA typing. All samples were subjected to molecular genetic analyses of the sex using "Genderplex" consisting of two different regions of the amelogenin gene, SRY and four X-STR loci. From 90% of the extracted remains (n=87) sex assignment was possible. Of these, 49 (56.3%) brought a genetically male result. All of these DNA extracts were subjected to Y-STR analysis using Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific) and/or PowerPlex Y23 System (Promega). At least partial Y-STR profiles were obtained from all samples. A detailed comparison between mediaeval/post-mediaeval and contemporary Y-chromosomes was performed by searching the obtained haplotypes (HTs) in the Y Chromosome Haplotype Reference Database (YHRD: https://yhrd.org) comprising 154,329 haplotypes from 991 populations in 129 countries at the time of query (Release 50). YHRD searches of the pre-modern haplotypes yielded full matches plus neighbor-matches differing at only one allele from the query HT. Matches are presented with geographical and ancestry information of the contemporary HTs. For samples without direct YHRD-matches, this information is provided through their neighbor HTs. AMOVA was performed using the YHRD online tool on pairwise R_{ST} values to create the corresponding MDS plots. The pre-modern HTs were grouped according to medieval and post-medieval origin and compared to contemporary populations from Scandinavian (Norwegian, Swedish and Danish), Northwest European, and Northeast European populations. Both pre-modern populations showed small genetic distances to contemporary Scandinavians and larger distances to Northeast Europeans with Northwest European populations in between. As expected, an initial assessment of the Y-chromosomal haplogroups (HGs) showed that most of the samples were attributable to the main European HGs I1, R1a and R1b. However, one of the HTs seemed to be associated with HG-Q which is rare in Europe and hitherto little evaluated in this region. Network analysis was applied for detecting similar HTs in contemporary samples from Norway and adjacent Northern European countries stored in the YHRD. The outcomes of this survey should initiate a detailed SNP based HG-assessment of HG-Q candidate samples.

1148T

Detecting polygenic adaptation using GWAS data. *C. Bhérer[1], J.K. Pickrell[1,2].* 1) New York Genome Center, New York, NY, USA; 2) Department of Biological Sciences, Columbia University, New York, NY, USA.

   Natural selection on a polygenic phenotype may drive local adaptation via small shifts in allele frequencies at many loci. To test for this mode of adaptation in humans, we introduce a hierarchical model for detecting selection on a phenotype from changes in allele frequency at loci identified by genome-wide association studies (GWAS). Our method uses the normal approximation to genetic drift to jointly model neutrality and selection in a three-population tree, and a MCMC to estimate the strength of selection. Simulations show that our three-population test has increased power to detect selection compared to a related sign test. We test for polygenic adaptation in contemporary human populations using GWAS data for over 40 traits (Pickrell et al. 2016), and genomic data from Bray et al. 2010, Lazaridis et al. 2014 and The 1000 Genomes Project. We confirm selection for increased height in northern European populations (and decreased height in southern European populations), and find suggestive evidence in populations from other continents. Intriguingly, we also find evidence for selection on unibrow in some European and African populations. Our study shows how to leverage results from GWAS studies to gain insights into the evolutionary history of alleles that contribute to current phenotypic variation, notably disease risk.

1149F

Evidence for detailed historical European population structure from large-scale, diverse genetic polymorphism data. *P. Carbonetto[1], J. Byrnes[1], J.M. Granka[1], Y. Wang[1], K. Noto[1], E. Han[1], A.R. Kermany[1], K.A. Rand[1], E. Elyashiv[1], H. Guturu[1], N.M. Myres[2], E.L. Hong[1], R.E. Curtis[2], K.G. Chahine[2], C.A. Ball[1].* 1) Ancestry.com DNA, LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT.

Despite the recent surge of interest in ancient genomes, we show that there is still much to be elucidated about human demography from contemporary genomes. Here, we demonstrate the use of genealogical data to generate demographic insights from analysis of a large-scale, heterogeneous genetic data set. Specifically, we show that an unsupervised ADMIXTURE analysis of genotypes from 131,293 primarily US-born individuals, followed by a simple statistical analysis of the 3 million pedigree records linked to these genotype samples, yields novel insights into European genetic diversity. In contrast to principal component analysis (PCA), which is the most widely used approach to investigating European genetic diversity, we use ADMIXTURE to infer genetically differentiated source populations reflecting more distant historical time periods. Unsurprisingly, among European-origin individuals, admixture is pervasive. Despite this, our ADMIXTURE analysis with $K = 12$ ancestral populations identifies 5 stable, genetically differentiated groups within Europe (with putative historical counterparts in parentheses): Ashkenazi Jewish, Irish (Celts), Eastern Europeans (Slavs), Scandinavians (Nordics) and Iberians, featuring Basques and Sardinians. The genealogical data also allow us to provide a detailed portrait of the genetic composition of contemporary peoples across North America (e.g., Iberians in Cuba), and other parts of the world. This work suggests the potential for drawing more detailed connections between present-day and ancient genetic variation by leveraging large, heterogeneous genetic data sets.

1150W

Extremely rare variants reveal sequence context and genomic features that shape the germline mutation rate in humans. *J. Carlson[1], J. Li[1,2], S. Zöllner[3,4], The BRIDGES Consortium.* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) Department of Psychiatry, University of Michigan, Ann Arbor, MI.

Mutation is a fundamental biological process that drives evolution and the incidence of heritable disease. The human germline mutation rate is known to correlate with various features of the genomic landscape, but the exact effect of these predictors is unknown. Existing approaches are insufficient to study the variation in mutation rate across the genome in an unbiased, high-resolution manner: comparative genomics methods are confounded by selection and biased gene conversion, whereas *de novo* mutation events ascertained by pedigree sequencing are too sparse to characterize mutation rate heterogeneity in great detail. Here, we present a novel approach to comprehensively assess predictors of mutability. We collect a dataset of over 36 million high quality singleton variants from 3765 whole-genome sequences. These extremely rare variants arose very recently in the population, and are thus largely unaffected by confounding evolutionary factors. Using these singletons, we develop statistical models to predict the effect of genomic context and adjacent nucleotides on the mutation rate at a single-base resolution. We find strong evidence for heterogeneity in the mutation rates of short sequence motifs, with up to a 60-fold difference between the mutation rates of non-CpG 7-mer motifs within a single mutation class. These sequence-specific mutation rates vary by an additional order of magnitude when we account for local genomic and epigenomic features, such as replication timing and chromatin organization, suggesting that the genomic landscape is important in modulating the efficiency of DNA damage and repair mechanisms. We apply this knowledge to develop a model that predicts the distribution of true *de novo* mutations with greater accuracy than competing models trained using common polymorphisms and/or ignoring the influence of genomic features. Adapting this strategy to model singleton density in genomic windows, we are able to explain over 97% of variance in the genome-wide distribution of recent mutations. This detailed atlas of germline mutation rates can be applied to many important topics in the field, ranging from population genetics inference to clinical sequencing.

1151T

Using whole-genome sequencing to shed insight on the complex pre-history of Sardinia. *C. Chiang[1], J. Marcus[2], C. Sidore[3], M. Zoledziewska[3], M. Steri[3], H. Al-asadi[2], S. Sanna[3], G. Abecasis[4], D. Schlessinger[5], F. Cucca[3], J. Novembre[2].* 1) Psychiatry, UCLA, Los Angeles, CA; 2) Human Genetics, University of Chicago, Chicago, IL; 3) Istituto di Ricerca Genetica e Biomedica, CNR, Cagliari, Italy; 4) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 5) Laboratory of Genetics, NIA, NIH, Baltimore, MD.

Genetic studies of complex traits using individuals from the island of Sardinia have been fruitful for decades. Understanding the population history of Sardinia may provide insights to how risk alleles arise in Sardinia and the expected degree of sharing with mainland populations. Current models for the peopling of Europe consider extant Europeans as having varying ancestral contributions from Paleolithic hunter-gatherers, Neolithic farmers, and Steppe pastoralists. In these analyses, extant Sardinians are inferred with the largest amount of Neolithic farmer ancestry among Europeans. However, previous studies of uniparental markers have highlighted the high frequency of the Y-chromosome haplotype I2a1a1 in Sardinia, which has been associated with Paleolithic ancestry in Europe. Here we aim to elucidate finer details of the population history of Sardinia. We study >3,500 whole-genome sequenced individuals across Sardinia together with reference European datasets and recently published ancient humans. We confirm that compared to mainland Europeans, Sardinians exhibit the greatest amount of shared drift with Neolithic farmers. Within the island of Sardinia, there is a demarcation of individuals from the Lanusei Valley in the geographically isolated province of Ogliastra and individuals from other provinces. In unsupervised analyses, a Sardinian-specific ancestry component correlates with shared drift parameters with both the Neolithic farmers and Paleolithic hunter-gatherers, and is consistent with supervised estimates of ancestry proportions in which the Ogliastra individuals have higher Neolithic farmer and Paleolithic hunter-gatherer ancestries, while individuals from the rest of the island show an infusion of the pastoralist ancestry. Finally, we find that the Sardinian people exhibit increased sharing of alleles with the Neolithic farmers on the X chromosome compared to the autosome (P < 1x10^{-4}), suggesting a sex-biased demographic history in Sardinia. Together, our results indicate that in addition to the strong Neolithic farmer component of ancestry, isolated regions of Sardinia also harbors significant ancestry components from Paleolithic Europe and that the Neolithic transition in Sardinia may have involved sex-biased demographic change. These results help provide more insight into the history of Sardinians and the frequency distribution of variants they carry.

1152F

Mutation load and health outcomes are modulated by allele specific expression and the environment. *H. Edgington[1], M.J. Fave[1,3], I. Alves[1], J.C. Grenier[3], V. Bruat[3], P. Awadalla[1,2,3].* 1) Ontario Institute for Cancer Research Informatics and Biocomputing 661 University Ave. Suite 510 Toronto, Ontario M5G 0A3 Canada; 2) University of Toronto Department of Molecular Genetics 1 King's College Circle Toronto, Ontario M5S 1A8 Canada; 3) Centre Hospitalier Universitaire Sainte-Justine 3175 Chemin de la Côte-Sainte-Catherine Montréal, QC H3T 1C4 Canada.

While mutation load is often inferred by counting the number of mutations in an individual genome or population, it does not incorporate expression or translation information, or impact on phenotypes or fitness. For example, allele-specific expression (ASE), the imbalanced expression of alleles at a particular locus, is a phenomenon that is frequent and pervasive. Not only are patterns of ASE heritable, they have important consequences for biological processes and functions. In this study we used genomic, transcriptomic and endophenotypic data captured from a population cohort of over 40000 individuals in Quebec to infer the impact of mutation, allele specific expression and regional variability including environmental exposures on phenotypic data and clinical outcomes. Allelic proportions from RNAseq data were combined with genotype data from approximately 1000 individuals to characterize the intersection of ASE, individual phenotype, and population exposures, genome-wide. Tests of associations between ASE summary statistics such as expression-weighted mutation load and three groups of phenotypes, demographic and lifestyle traits, health and disease related traits, and medication use, revealed significant associations between ASE and several phenotypes including bowel disease and arterial stiffness. Using coinertia analyses, we identified variants showing the strongest covariance between allelic expression patterns and phenotypic variation. Genes showing consistent bias towards the ancestral alleles were enriched for biological processes critical to cell function and replication, such as metabolic and localization processes and immune pathways. Furthermore, we found that the unique demographic history of the French-Canadian population has impacted ASE patterns in this cohort. French-Canadians are descendants of a founder population that has expanded by a series of sequential founder events following colonization of Quebec. We compared ASE between three subpopulations and found that individuals at the forefront of expansion exhibit increased ASE as well as reduced heterozygosity. We conclude from these results that patterns of ASE at a genomic scale are correlated with important measures of individual health, and describe novel associations between expression profiles at specific variants and health outcomes. Additionally, we found that like genetic variation, expression profiles vary among populations that have experienced rapid demographic change.

1153W

Genomic insights into the population structure and history of the Irish Travellers. *E.H. Gilbert[1], S. Carmi[2], S. Ennis[3], J.F. Wilson[4,5], G.L. Cavalleri[1].* 1) Molecular and Cellular Therapeutics , Royal College of Surgeons in Ireland, Dublin, Leinster, Ireland; 2) Braun School of Public Health, The Faculty of Medicine, The Hebrew University of Jeruslam, Jerusalem, Israel; 3) School of Medicine and Medical Science, University College Dublin, Dublin, Ireland; 4) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, Scotland; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, Scotland.

  Aims: The Irish Travellers are a population with a history of nomadism. Consanguineous unions are common, and as a population they are socially and genetically isolated from the surrounding, "settled" Irish population. A previous low-resolution genetic analysis suggested a common Irish origin between the settled and the Traveller populations. What is not known, however, is the extent of population structure within the Irish Traveller population, the time of divergence from the general Irish population, and the extent of autozygosity. **Methods**: We recruited Irish Travellers from across Ireland and the UK. To be included a participant had to have had at least three grandparents with a surname associated with the Irish Travellers. DNA was extracted from saliva samples, and genotypes were generated using the Illumina OmniExpress SNP genotyping platform. With this data, we investigated population structure using fineStructure, quantified the levels of autozygosity with PLINK, and estimated a time of divergence using a method based on Identity by Descent (IBD) segment identification. **Results**: We merged, cleaned, and analysed data from 42 Irish Travellers, 2232 settled Irish, 2039 British, 143 Roma Gypsies, and 931 individuals from 57 world-wide populations. We confirm an Irish origin for the Irish Travellers, demonstrate evidence for population substructure within the population, confirm high levels of autozygosity consistent with a consanguineous population, and for the first time provide estimates for a date of divergence between the Irish Travellers and settled Irish. **Conclusion**: Our findings have implications for disease mapping within Ireland, as well as on the social history of the Irish Traveller population.

1154T

A compendium of worldwide surveys of population-specific F_{ST} for forensic Y-STR markers. *T.O. Hall[1], B.S Weir[2].* 1) Institute of Public Health Genetics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA.

  Y-STR markers are particularly useful forensically for identifying the male contributor to a male-female DNA mixture. A population genetic approach to estimation of the match probability is dependent on an appropriate estimate of θ, which is calculated empirically as F_{ST}. This estimate depends on the choice of F_{ST} estimator, the markers included and how they are combined. This paper looks at the characteristics of estimates obtained using a population-specific F_{ST} estimator in three datasets: Y-Chromosome Haplotype Reference Database (YHRD), Human Genome Diversity Project (HGDP), and data published by Xu et al in 2014. Individuals with missing data were excluded. To emphasize that a population-specific estimate of θ can only be calculated relative to a total, our estimator is written as one minus the heterozygosity within population i at locus l, divided by the average heterozygosity between all populations in the total sample: $\beta_{il}=1-(H_{il}/H_{Bl})$. $H_{il} = 1-sum(p^2_{Ail})$ where p_{Ail} is the sample allele frequence of allele A at locus l in population i. Average heterozygosity between populations is $H_{Bl}= sum(H_{ijl})/(r(1-r))$, where $i \neq j$ and r is the number of populations. Conventional F_{ST} is the average of the β_{il}'s, here denoted as β_{Wl}, and can be calculated as: $\beta_{Wl}=1-(H_{Wl}/H_{Bl})$, where $H_{Wl} = sum(H_{il})/r$. Because markers on the Y-chromosome are assumed to be linked, multi-marker haplotypes are considered a single locus. β_{il}'s were calculated using the above equations for the population in each of the three datasets for l = haplotype combinations of Y-STR loci, where n=23 (YHRD), n=19 (HGDP) and n = 16 (Xu et al.) and k = 1, 2, …, 13. β_{Wl}'s's were calculated for each haplotype within regional ethnicity. Median worldwide β_{il}'s ranged from -0.01 to 0.26. Because Y-STRs are linked, we hypothesized that there would also be substantial dependency between pairs of single-locus β_{il}'s. However, we observed low to moderated correlation for all three datasets. There appears to be a weak relationship between β_{il} correlation and LD. As the number of loci included in a haplotype increased, mean β_{W} decreased, and the range of β_{W} values became smaller. The rate of decrease in mean β_{W} is not linear and related to correlation among single-locus β_{il}'s.

1155F

SMC²: Using particle filters to infer historic population sizes and migration rates from whole genomes and multiple samples. *D. Henderson, J. Zhu, G. Lunter.* Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

The sequentially Markovian coalescent model is widely used to infer historic population sizes from whole-genome sequence data from a few individuals. The complexity of the state space, a sequence of genealogies, severely limits the number of samples involved in the analysis and necessitates simplifications, such as discretization of time. Previous methods, such as MSMC, are able to consider samples from multiple populations enabling inference of a measure similar to migration rate. However, in MSMC the ability to consider multiple samples comes with a cost of reduced resolution in the distant past. We propose a simulation-based approach which allows for the explicit inference of migration rates and population sizes in the distant past. Our method, sequential Monte Carlo inference of the sequentially Markovian coalescent model (SMC²), simulates sequences of genealogies and weighs these according to their likelihood given the genotypes of the individuals in the sample. The weighted sample of sequential genealogies is used to update our parameter estimates for the recombination rate, historic population sizes, and historic migration rates. SMC² is able to accurately infer either bi-directional or uni-directional migration between two populations along with the effective population sizes. Due to its high resolution inference of the distant past, SMC² is an attractive method for investigating ancient populations and gene flow. As such, we have incorporated ancient samples into our analyses for comparing Neanderthal, Denisovan, and modern human population histories. We have developed a flexible tool for demographic inference from whole-genome sequences of a few individuals. The use of simulation-based techniques allows us to fit complex demographic models, and in particular enables nonparametric inference of migration rates.

1156W

Evolutionary history of Tibetans inferred from whole-genome sequencing. *H. Hu[1], N. Petousi[2], G. Glusman[3], Y. Yu[1], R. Bohlender[4], T. Tashi[5], J.M. Downie[6], J.C. Roach[3], A.M. Cole[7], F.R. Lorenzo[5], A.R. Rogers[4], M.E. Brunkow[3], G. Cavalleri[7], L. Hood[3], A.M. Alpatty[8], J.T. Prchal[5,6], L.B. Jorde[6], P.A. Robbins[9], T.S. Simonsons[10], C.D. Huff[1].* 1) Epidemiology, UT MD Anderson Cancer Center, Houston, TX; 2) Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 3) Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA; 4) Department of Anthropology, University of Utah, Salt Lake City, UT; 5) Department of Medicine, University of Utah School of Medicine and George E. Wahlin Veterans Administration Medical Center, Salt Lake City, UT; 6) Department of Human Genetics, University of Utah, Salt Lake City, UT; 7) Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland; 8) Skaggs School of Pharmacy and Pharmaceutical Science, UC San Diego, CA; 9) Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, OX1 3PT, UK; 10) Department of Medicine, Division of Physiology, University of California San Diego, La Jolla, CA.

The indigenous people of the Tibetan Plateau have been the subject of much recent interest because of their unique genetic adaptations to high altitude. Here we report the first analysis of whole-genome sequences from Tibetans. A total of 27 Tibetan genomes were sequenced, allowing us to infer a detailed history of demography and natural selection of this population. We demonstrate that Tibetans and Han Chinese likely first diverged in the middle Paleolithic period between 44 and 58 thousand years ago, but with substantial gene flow until between 3 and 11 thousand years ago; both populations have undergone large expansions since they diverged. Using a recently developed statistical approach to detect archaic admixture, we show that the Han and Tibetan genomes each contain about 0.4% admixture from ancient Denisovans. The Composite of Multiple Signals (CMS) test identified *EPAS1* and *EGLN1* as the top two positive selection candidates genome-wide, consistent with previous reports. Both genes encode components of the hypoxia-inducible factor (HIF) pathway. The CMS test identified additional high-altitude adaptation candidates, including *PTGIS*, *VDR*, and *KCTD12*. Our haplotype analysis found that the advantageous Tibetan *EPAS1* haplotype contained a large number of variants shared with the Denisovan genome, consistent with previous reports, but with an ancient gene tree divergence between the Tibetan and Denisovan *EPAS1* haplotypes of about one million years ago. With the exception of *EPAS1*, we observed no evidence of positive selection on Denisovan-like haplotypes, strongly suggesting that adaptive introgression from a Denisovan-like ancestor was a unique occurrence in the evolutionary history of Tibetans.

1157T

Parallels between processes of genetic and linguistic admixture in Cape Verde. *E.M. Jewett[1], P. Verdu[2], T.J. Pemberton[3], N.A. Rosenberg[4], M. Baptista[5].* 1) EECS and Statistics, UC, Berkeley, Berkeley, CA; 2) CNRS/MNHN, Paris Diderot/Sorbonne Paris Cite, Paris, France; 3) Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB; 4) Biology, Stanford University, Stanford, CA; 5) Linguistics and Afroamerican and African studies, University of Michigan, Ann Arbor, MI.

The Cape Verdean archipelago, located near the western coast of Africa, provides one of the earliest examples of admixture between European and African populations, resulting in a population that is highly diverse in both genetic and linguistic ancestry. Because present-day genetic and linguistic variation in Cape Verde arose from the same well-documented migration and admixture events beginning in the late 1400s, Cape Verde provides an important resource for studying similarities and differences between processes of linguistic and genetic transmission during admixture. We investigated patterns of genetic and linguistic diversity among 44 unrelated Cape Verdean individuals, sampling genotypes at ~2.5 million genome-wide SNPs and sampling the spontaneous speech of individuals in Cape Verdean Creole. We found evidence for the vertical transmission of speech patterns in Cape Verde, finding that variation in parental birthplaces was predictive of variation in the speech patterns of study participants. We also found that individual levels of African genetic ancestry in Cape Verde were significantly positively correlated with the number of words of putative African origin used by each individual. This finding suggests that similar processes of transmission, diffusion, and isolation have led to similar patterns of genetic and linguistic variation despite non-vertical transmission and potentially strong influences from sociocultural factors on individual speech patterns. Although the vertical transmission of speech patterns can partly account for the observed correlation between genetic and linguistic ancestry, we also found evidence for the influence of cultural practices on linguistic transmission, finding that maternal birthplaces were more predictive of speech patterns in offspring than paternal birthplaces, a result that is consistent with patterns of marriage and remarriage in Cape Verde. Our study provides an improved understanding of the processes that have shaped diversity in the present-day Cape Verdean population and the degree to which linguistic and genetic changes occur in parallel.

1158F

How molten is the pot? Analyzing self-reported and genetic ancestry of individuals and reproducing couples in the USA. *K.E. Kaseniit, S. Nazareth, C. Lo, K. Wong, J.D. Goldberg, E.A. Evans, I.S. Haque.* Counsyl, Inc., South San Francisco, CA.

Purpose: The 2010 US Census showed that 15% of US marriages occur between spouses of different ethnicities, a more than double increase over the previous two decades (Pew Research Center). Further, although prior studies in personal genetics have noted that self-reported ethnicity (SRE) is not fully concordant with genetic ancestry (GA), current medical genetics guidelines are based around SRE. We sought to characterize the increasing diversity of the US population and its impact on medical practice by exploring both the discordance between SRE and GA and current inter- and intra-ethnic mating patterns. **Methods:** Deidentified data was drawn from >100,000 individuals receiving carrier screening; all individuals self-reported ethnicity, including options for unknown and for mixed ethnicity. To evaluate discordance between SRE and GA, we adapt an existing ancestry analysis method [1] for use with sparse (<5000 reads/sample) off-target sequencing reads and analyze data from >60,000 individuals tested by next-generation sequencing. To evaluate mating patterns, we analyze the paired self-reports from 37,719 couples in which both members of the couple were tested. **Results:** We demonstrate the ability to robustly recover GA coefficients from extremely sparse sequencing data. We find that over one-sixth of the samples with unknown SRE have a true GA that would have prompted additional screening under existing medical guidelines. Further, over 4% of all patients have at least 25% GA from a different population than their self-reported ethnicity, suggesting at least one ancestrally different grandparent. We find that 39.7% of couples were of mixed ethnicity with significant divergence in behavior between males and females. Women who identify as as East Asian, Southeast Asian and Mixed/Other Caucasian are more likely to partner with men outside of their self-reported ethnic group than are men of the same groups. Men who identify as Ashkenazi Jewish, African American or European (Northern or Southern) are more likely to partner with women outside of their self-reported ethnic group than are women from those groups. **Conclusions:** SRE is not fully reliable in the clinical setting and a large fraction of couples cross ethnic boundaries, suggesting that existing medical standards that rely on single-ethnicity self-reports are likely to become untenable in the near future. [1] Frichot E et al. Genetics 2014.

1159W

Fine-scale genetic diversity of Japanese population. *Y. Kawai, K. Misawa, T. Mimori, K. Kojima, F. Katsuoka, M. Yamamoto, J. Yasuda, M. Nagasaki.* Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.

Japanese population is isolated from other east Asian populations due to its geological location. It is thought that the modern Japanese people are admixture of two ancestral populations; migrants in Jomon period (16 000–3000 year before present; YBP) and migrants in Yayoi period (3000–1700 YBP). However, despite current large census population size (approximately 127 million), demographic history after spread into the Japanese Archipelago is still not well understood. Thus, the study of detailed genetic structure within the population is important not only in human genetic studies but also from a biomedical point of view. We investigated the genetic history of the modern Japanese population using the genome-wide data of the cohort study of Tohoku Medical Megabank Organization. The whole genome sequencing and SNP genotyping data obtained from cohort participants some of whom all four grandparents originated from the same area were used for population structure and demographic inferences. The haplotype-based fine scale genetic structure is consistent with geographic variation. The analysis of cross coalescent rate using the whole-genome sequence data suggests that the genetic differentiation within Honshu island has begun after the last migration event in Yayoi period.

1160T

Characterization of local adaptation in Africa from whole genome sequence data. *D.E. Kelly[1,2], S.H. Fan[1], M.H. Beltrame[1], M.E.B. Hansen[1], S. Mallick[3,4,5], T. Nyambo[6], S. Omar[7], D. Meskel[8], G. Belay[9], A. Froment[10], N. Patterson[4], D. Reich[3,4,5], S.A. Tishkoff[1,2,10].* 1) Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA; 4) Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA; 5) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, 02115, USA; 6) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dares Salaam, Tanzania; 7) Kenya Medical Research Institute, Center for Biotechnology Research and Development, Nairobi, Kenya; 8) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 9) UMR 208, IRD-MNHN, Musée de l'Homme, Paris, France; 10) Department of Biology, University of Pennsylvania, Philadelphia, PA 19104.

Humans have lived in Africa for much of their evolutionary history, and the diversity of climates, diets, and pathogens across the African continent have imposed unique selective pressures. Different ethnic groups in Africa have undergone local adaptation in response to these pressures, selecting for genes and biological pathways in a population- or region-specific manner. A powerful method for detecting local adaptation is the population branch statistics (PBS), which partition measures of population differentiation along the branches of a given evolutionary tree. We applied this approach to whole genome sequencing data from 52 individuals representing 22 ethnic groups from across Africa, pooled into 6 ancestries based on *ADMIXTURE* analyses. Using pairwise F_{ST} measures between ancestral groups and a fixed tree topology, branch lengths which best explain the observed patterns of population differentiation were inferred at all variable loci. A sliding-window approach was used to identify localized regions of enriched PBS values, indicating putative selection on genes related to taste perception in the San and Rainforest Hunter Gatherers, keratinization in the San and Nilo-Saharan/Cushitic groups, and immunity in Bantu-speaking populations. This work is supported by NIH grants 1R01DK104339-01, 1R01GM113657-01, and Pioneer Award DP1 ES022577-04 to SAT. The sequencing was funded by the Simons Foundation (SFARI 280376) and the U.S. National Science Foundation (BCS-1032255) grants to DR.

1161F

A temporal perspective on the interplay of demography and selection on deleterious variation in humans. *E. Koch[1], J. Novembre[1,2].* 1) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL.

   Population genetic theory has long stressed the importance of population size in determining the abundance of mutations with small effects on fitness in populations. The effects of population size become complex when population sizes have been changing as part of a dynamic, non-equilibrium demographic history. Several recent papers in human population genetics have focused on this interplay of demographic history and deleterious variation. Here we investigate how deleterious variation responds to population size change by providing a temporal view and mechanistic explanation of changes in the site frequency spectrum. We focus on demographic events thought to have occurred during human history, specifically contrasting a history representative of an Out-of-Africa population experiencing a bottleneck and growth versus a relatively stable population history representative of an African population. We then analyze exome sequence data from the Exome Aggregation Consotrium (ExAC), including 33,370 and 5,203 individuals in empirical ancestry clusters of African and non-Finnish European individuals, respectively. Our theoretical results emphasize how even simple summaries of the site frequency spectrum can have potentially complex responses to demographic change. This problem grows as higher moments of the frequency spectrum are considered, and we use the proportion of segregating sites predicted to be deleterious as an example. Our results help explain the difficulty of inferring slight differences in the distribution of deleterious variation using next generation sequence data. These findings have implications for comparisons of deleterious variation between populations that are based only on summary statistics of the site frequency spectrum, indicating such comparisons can be ambiguous about differences in the efficacy of selection.

1162W

Personal ancestry inference at the finest scale reveals more sub-structure in the UK. *D. Lawson, G. Weyenburg.* Integrative Epidemiology Unit, University of Bristol, Bristol, UK, United Kingdom.

   Chromosome Painting has revealed genetic differences within the UK at a very fine scale [1], with structured genetic variation within a single county in some cases (such as Cornwall & South Wales). However, in that work, it was not possible to genetically distinguish much of England, which appeared as a single homogeneous group. Here, we describe an extension to the Fine-STRUCTURE [2] clustering that can further distinguish ancestry even within England; for example, identifying regions such as Norfolk, the Midlands and the South as genetically distinct. The approach works by using the known county locations to craft genetic features to use in unsupervised clustering. Specifically, we group individuals by their geographic sampling location into reference donor populations. This forms an ancestry profile - which can be viewed as a careful choice of feature vector - that still allows unsupervised genetic clustering for all individuals. Further, we describe how this approach allows individuals to be described as an admixture of the inferred geographical clusters. This allows ancestral information to be recovered for individuals who are not purely represented by a single geographical location. This also allows us to characterise the genetic relationship between the inferred clusters, several of which represent drift that is most strongly represented by a particular geographical region (including Cornwall, Wales, Scotland and the North of England) and others of which represent characteristic admixture proportions between these ancestral drifted populations. Beyond improving resolution, this approach facilitates personal genomics because individuals can be represented in terms of the fixed reference panel. We demonstrate the utility of the approach by describing the ancestry of the UK10K participants in terms of the new, high resolution POBI clusters. Previously, a similar analysis [3] without geographical information inferred little population structure in the UK from these samples, but now we have a rich representation of their population structure, including an assessment of admixture from outside the UK. This highlights the value in high quality fine-scale geographic sampling, which could now facilitate this level of ancestry identification for many other countries. [1] Leslie et al 2015, Nature 519:309–314 [2] Lawson et al 2012, PLoS Genet. 8:e1002453 [3] UK10K Consortium 2015, Nature 526:82-90.

1163T

Exploring detailed demographic histories using stairway plot 2. *X. Liu.* Human Genetics Center, UTHealth School of Public Health, Houston, TX.

Inferring human demographic history using genetic information can shed light on important prehistoric evolutionary events such as population bottleneck, expansion, migration, and admixture, among others. Recently we developed a model-flexible method called stairway plot (Liu and Fu, Nature Genetics, 2015), which infers detailed population size changes over time using SNP frequency spectra (SFS). This method can be applied to low-coverage sequence data, pooled sequence data or RAD-seq data from non-model organisms with hundreds of individuals. We have extended this method to be applicable to folded SFS, therefore will not be biased by incorrected inferred ancestral alleles. We also used an ensemble approach to control model overfitting. Using extensive simulation we showed that the accuracy of stairway plot with folded SFS is similar to that with correct unfolded SFS. We also compared our improved method (aka stairway plot 2) to the state of the art method for inferring demographic history, multiple sequentially Markovian coalescent (MSMC) (Schiffels and Durbin, Nature Genetics, 2014), and showed that our method outperforms the MSMC for inferring recent population size changes. We applied our method to whole genome sequence data of some human and non-human primate populations and showed their detailed demographic histories.

1164F

Testing directional selection on polygenic traits using ancient DNA. *J. Marcus, J. Novembre.* Department of Human Genetics, University of Chicago, Chicago, IL.

Understanding the evolutionary basis of trait variation requires the ability to distinguish between models where a trait has recently undergone directional selection as opposed to stabilizing selection or simply neutral evolution. Methods to distinguish amongst such models are still poorly developed, especially for highly polygenic traits. Recent progress in understanding the polygenic basis of trait variation, using genomewide association studies (GWAS), and the increasing availability of ancient DNA (aDNA) samples provide new opportunities. Here, we develop and compare several possible inferential procedures that intersect putative quantitative trait loci discovered via GWAS with aDNA data to test whether a phenotype has experienced directional selection on the basis of allele frequency change. We consider key challenges for methods development in this area, specifically, uncertainty in effect sizes and allele frequencies. We compare Bayesian procedures that take these uncertainties into account with approaches that use simple heuristic test statistics. We show that for several methods power exists even with small aDNA sample sizes when the signature of selection is distributed across many loci, and we apply these approaches to investigate signatures of selection on a panel of quantitative traits in humans (e.g. height, BMI, autoimmune disease risk) using published GWAS and aDNA data. While aDNA studies in some species are rapidly scaling upwards in sample size, even small samples will be helpful for shedding light on the evolution of polygenic traits.

1165W

Genome-wide ancient DNA from Europe's first encounter of farmers and hunter-gatherers. *I. Mathieson[1], N. Rohland[1,2], S. Mallick[1,2,3], D. Fernandes[4], M. Ferry[1,3], B. Gamarra[4], E. Harney[1,3], M. Michel[1,3], K. Sirak[4], K. Stewardson[1,3], A. Mittnik[5,6], C. Posth[5,6], P. Stockhammer[7], S. Chohadzhiev[8], S. Venelinova[9], N. Elenski[10], V. Petrova[11], G. Ganecovski[12], F. Veljanovska[13], S. Iliev[14], Y. Boyadzhiev[15], W. Haak[6], S. Vai[16], D. Caramelli[16], B. Gaydarska[17], S. Alpaslan Roodenberg[18], N. Patterson[3], J. Krause[5,6,19], R. Pinhasi[4], D. Reich[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; 3) Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA; 4) School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland; 5) Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, 72070 Tübingen, Germany; 6) Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; 7) Institute for Prehistory and Early History and Near Eastern Archaeology, Heidelberg University, Heidelberg, Germany; 8) St. Cyril and Methodius University, Veliko Turnovo, Bulgaria; 9) Regional museum of history, Shumen, Bulgaria; 10) Regional Museum of History, Veliko Tarnovo, 5000. Bulgaria; 11) Dr. Faculty of History, Sofia University, ST. Kliment Ohridski, 1504, Bulgaria; 12) Regional Historical Museum Vratsa, 3000. Bulgaria; 13) The Archaeological Museum of Macedonia, Skopje. 1000, Republic of Macedonia; 14) The Regional Museum of History. 6300,Haskovo. Bulgaria; 15) National Archaeological Institute with Museum, Bulgarian Academy of Sciences, 1000 Sofia, Bulgaria; 16) Dipartimento di Biologia, Università di Firenze, 50122 Florence, Italy; 17) Department of Archaeology, Durham University, UK; 18) Independent researcher, Santpoort-Noord, The Netherlands; 19) Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany.

The area of southeastern Europe known as the Balkans has always been a crossroads between Europe and Asia: a conduit for people, culture and language. Beginning around 6,500 BCE, the Balkans was the first place in Europe to become transformed by farming, brought by a new wave of migrants from Anatolia. From this staging point, farming and people spread to all corners of Europe. However, the dynamics of the interaction between farmers and indigenous European hunter-gatherers in the first place that they encountered each other remains poorly understood because of the near complete absence of genetic data from prehistoric specimens from this region.   We generated new genome-wide ancient DNA data from 65 farmers from the Balkans and adjacent regions dating as far back as 6,400 BCE. We document how the dynamics of admixture between the regions first farmers and its indigenous hunter-gatherers was complex, with evidence of local admixture from hunter-gatherers related to those from Scandinavia and Eastern Europe. The population admixture was patchy across both space and time, varying in magnitude between 0% and 30% for different early Balkan farming populations. The hunter-gatherer admixture in the early farmers of the Balkans is not closely related to the hunter-gatherer admixture that is predominant in present-day Europeans. This suggests that the waves of farmers that contributed most of the migrants to northern and western Europe were not ones that mixed substantially with local Balkan hunter-gatherers.   We also analyze the data to generate new insights about natural selection. The first farmers of the Balkans were in the initial stages of adaptation to environments that were dramatically different from those that their ancestors had encountered. We show that many of the adaptations related to diet and immunity that later become common in Europe were already present in early Balkan farmer populations, but not at high frequency. Thus, the adaptation of the first European farmers to their local environment was driven to a substantial extent by pre-existing variantion.

1166T

Detecting mutagenic recombination using genome-wide divergence data. *T.N. Phung, C.D. Huber, K.E. Lohmueller.* Bioinformatics IDP, University of California, Los Angeles, Los Angeles, CA.

Mutation is the ultimate source of genetic variation that is acted on by evolutionary processes and gives rise to disease. Currently, the reasons why the mutation rate varies across the genome are not well understood. One possibility is that recombination could be mutagenic and as such, variation in the recombination rate contributes to variation in the mutation rate. A positive correlation between human-chimp neutral divergence and recombination rate has lent support to this hypothesis. However, this correlation could also be driven by the effects of natural selection. To date, the degree to which recombination is mutagenic in the human genome remains unknown. To address this question, here we develop a likelihood-based framework to test the extent to which natural selection and/or the coupling between recombination and mutation contributes to the variation in human-chimp divergence. Our method is built upon the idea that natural selection primarily affects divergence in regions of low recombination while mutagenic recombination affects regions of high recombination. We model the effects of natural selection on linked neutral divergence following Hudson and Kaplan 1995. We also estimate the degree of coupling between mutation and recombination. Application of this method to human-chimp divergence reveals that a model including background selection and mutation-recombination coupling fits the data substantially better than a model only including background selection (P-value < 0.01). Further, we use various filters to evaluate the role of biased gene conversion in generating these patterns. We will discuss implications of our results for understanding genetic variation across the human genome. Our inference framework can be applied to study mutation rate variation in non-model species where direct sequencing of family trios is not feasible.

1167F

Population structure and admixture of Xinjiang's Uyghurs. *F. Qidi[1,2], L. Yan[1], Z. Ying[1,2], L. Dongsheng[1,2], Y. Yajun[3], G. Yaqun[4], L. Haiyi[1], Y. Kai[1,2], W. Yuchen[1,2], Y. Xiong[1,2], Z. Chao[1,2], S. Meng[1,2], T. Lei[1,2], W. Xiaoji[1,2,5], Z. Xi[1,2,5], L. Jing[1], K. Asifullah[1], T. Kun[1], W. Sijia[1,6], X. Shuhua[1,2,5,6].* 1) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai; 2) University of Chinese Academy of Sciences, Beijing 100049, China; 3) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 4) Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, P. R China;; 5) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China;; 6) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

With a population size more than 10 million, the Uyghur people residing in Xinjiang are believed to be descendants of the most ancient of Turkish tribes with mixed Caucasian and East Asian ancestries. However, their genetic origins, population structure and admixture history remain poorly understood and debatable. Here we systematically assessed genetic diversity and individual ancestry composition of the Xinjiang's Uyghurs (XJU) by genotyping ~1,000 Uyghur individuals, roughly proportional to population size of 13 geographical regions, with high-density single nucleotide polymorphism arrays. We observed a southwest-northeast differentiation within the XJU, which is different from the expected north-south differentiation as separated by Tianshan Mountain. In the context of comparative analyses of 2,477 individuals representing 206 worldwide populations, four major ancestries were identified in XJU without very much variation among individuals, i.e., East Asia, Siberia, West Eurasia, and South Asia. However, XJU showed an overall unique genetic make-up and divergent history from surrounding neighbors including the other modern Turkic speaking populations. The results suggest a long history of population admixture and isolation. Facilitated by new methods including one developed in this study, our analyses shed exciting new light on genetic origins and admixture history of Uyghurs.

1168W

Singapore Integrative Omics Cohort: Establishing multiple omics baselines for three Southeast Asian populations. *W. Saw[1,2], E. Tantoso[2], H. Begum[1], L. Zhou[3], C. He[3], R. Zou[3], S. Chan[4], L. Tan[2], L. Wong[2], W. Xu[2], D. Kyin[2], Y. Lim[2], B. Li[2], N. Pillai[1], T. Peterson[5,6], T. Bielawny[5,6], P. Meikle[7,8], P. Mundra[7], W. Lim[2], M. Luo[5,6], R. Ong[2], L. Brunham[4], C. Khor[9], H. Too[10,11,12], R. Soong[13], M. Wenk[10,14,15,16], K. Chia[2], P. Little[1], Y. Teo[1,2,9,14,17].* 1) Life Science Institute, National University of Singapore, Centre for Life Sciences, Singapore, Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549; 3) MiRXES, Agency for Science, Technology and Research Singapore 138670; 4) Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research Singapore 138648; 5) Deparment of Medical Microbology, University of Manitoba, 730 William Avenue, Winnipeg, Manitoba R3E 0Z2, Canada; 6) National Microbology Laboratory, Winniepeg, Manitoba, Canada; 7) Baker IDI Heart and Diabetes Institute, Melbourne, Victoria, Australia; 8) Department of Biochemistry and Molecular Biology, The University of Melbourne, Parkville, Australia; 9) Genome Institute of Singapore, Agency for Science, Technology and Research Singapore 138672; 10) Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 11) Molecular Engineering of Biological and Chemical System/Chemical Pharmaceutical Engineering, Singapore-Massachusetts Institute of Technology Alliance, Singapore 117576; 12) Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research, Singapore 138668; 13) Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore; 14) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456; 15) State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China; 16) Department of Biological Sciences, National University of Singapore, Singapore 117543; 17) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

The Singapore Integrative Omics Cohort (IOMICs) establishes population reference measurement across multiple omics technologies in 364 samples from three major population groups (Chinese, Malay and Indian) in Southeast Asia. Using resources of(i) more than 2.5 millions single nucleotide polymorphisms (SNPs) that cover whole genome, exon regions, and pharmacogenomic variants, 198 HLA typing alleles across eight HLA loci, (ii) 21,649 transcript variants, (iii) 297 lipid species, (iv) 274 circulating miRNAs and (v) 296 non-omic variables that were related to clinical, lifestyle and diet of each participant. About half of the Malays and a third of the Indians have also been deeply whole-genome sequenced. Omics data on the other hand potentially provides information on homogeneity of a population. For intra-omics analysis, we conducted population structure analyses with principal component analysis for intra-omics analysis and identified inter-ethnic variation using ANOVA for non-genetic data and Wright's Fst index for genetic data. From the population structure analyses, our study observed genetic information distinctly determined the membership of the population as compared to other omics information. Based on intra-omics analyses, we observed Chinese population in Singapore is biologically more homogenous whereas the Singapore Indians are biologically more heterogenous. The high-throughput omic datasets allows careful characterization of intra- and inter-omic correlation within and across all three populations, providing valuable insights towards bridging the functional relevance of existing genetic findings.

1169T

Robust and scalable inference of population history from hundreds of unphased whole-genomes. *J. Terhorst[1], J.A. Kamm[1,2], Y.S. Song[1,2,3,4].* 1) Department of Statistics, University of California, Berkeley, CA 94720, USA; 2) Computer Science Division, University of California, Berkeley, CA 94720, USA; 3) Department of Integrative Biology, University of California, Berkeley, CA 94720, USA; 4) Departments of Biology and Mathematics, University of Pennsylvania, PA 19104, USA.

The genomes of present-day individuals contain a wealth of information about their ancestors, and there has been a surge of interest in utilizing this information to learn about the history of humans and many other organisms. It has been demonstrated recently that inference methods based on genealogical processes with recombination can reveal past population history in unprecedented detail. However, these methods scale poorly with the sample size, which limits resolution in the recent past (within the last 1000 generations), and they require phased genomes, which contain switch errors that can catastrophically distort the inferred history.   In response to these challenges, we have developed SMC++, a new statistical tool capable of analyzing orders of magnitude larger samples than existing methods while requiring only unphased genomes as input (its results are independent of phasing). SMC++ also requires far less memory and computation time than these methods, and can analyze genome-scale data on a laptop computer. We have implemented SMC++ as an open-source software package which features an easy-to-use graphical user interface.   We apply SMC++ to analyze sequence data from over a thousand genomes in eight human populations in Africa and Eurasia, hundreds of genomes from a Drosophila population in Africa, and tens of genomes from zebra finch and long-tailed finch populations in Australia. Our results suggest that, despite the wide range of biological differences between these species (e.g., in the generation time, mutation rate, and recombination rate), their long-term effective population size changes in the last 200,000 years follow a remarkably similar pattern, with major growth, decline and bottleneck events lining up in time very closely.

1170F

Mutation rate estimation from population data. *X. Tian[1], B.L. Browning[2], S.R. Browning[1].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA.

The two primary methods for mutation rate estimation have been counting de novo mutations in parent-offspring trios and comparing sequence data between species. With trio analysis it is difficult to control for genotype error and resolution is limited because each trio provides information from only two meioses. Inter-species comparison is difficult to calibrate and can be affected by selection. Genome-wide mutation rate estimates from inter-species comparisons have been twice as high as those from family-based studies [1]. A new class of methods for mutation rate estimation is based on segments of identity by descent (IBD) between pairs of individuals. The length of an IBD segment provides information on the number of meioses linking the two haplotypes through their common ancestor, and differences in the haplotype sequences provide information on the total number of mutations from those meioses. Recently, Palamara et al. proposed an IBD-based method for estimating mutation rates from perfectly phased data, such as that obtained from parent-offspring trios [2]. Whereas ordinary trio-based analysis uses only meioses within trios, Palamara et al.'s approach uses meioses from IBD between pairs of trio offspring, and is thus more powerful than counting de novo mutations. We develop IBD-based mutation rate estimation further by using multi-individual IBD rather than pairwise IBD. Multi-individual IBD provides more information that can be used to address haplotype phase uncertainty and genotype error. Our method is applicable to sequence data with or without close relatives. As a first stage, we have implemented our method for perfectly phased data. Analysis of simulated data indicates that our method is unbiased and has high precision when using actual or inferred IBD segments, actual or inferred effective population size, and genotypes with or without genotype error. With simulated genome-wide data having a mutation rate of 1.25×10^{-8}, a constant population size of 10,000 and a sample size of 400 diploid individuals, our method has standard errors of approximately 0.01×10^{-8} for the mutation rate estimate.   [1] Scally and Durbin, 2012, Nat Rev Genet [2] Palamara et al. 2015, Am J Hum Genet.

1171W

Background selection in the human genome explains F_{ST} better than recombination rate alone. *R. Torres[1,3], R.D. Hernandez[2,3].* 1) Biomedical Sciences Graduate Program; 2) Department of Bioengineering and Therapeutic Sciences; 3) UCSF, San Francisco, CA.

The consequences of linked selection in the human genome have been well characterized and studied. As a proxy for understanding the consequences of linked selection, much previous work has focused on the correlation of genetic diversity and differentiation with recombination rate, since the effects of linked selection are expected to be stronger in regions of low recombination. Using this proxy, decreased genetic diversity and divergence and increased population differentiation in regions of the genome under low recombination have been observed. However, recombination rate alone is not sufficient to describe the total effects of linked selection since the density of functional elements and magnitude of natural selection along the genome are also of considerable importance. In order to better measure the total effect of linked selection on population differentiation in humans, we utilize phase 3 of the 1000 Genomes Project and measure F_{ST} as a function of background selection, which takes into account the density and magnitude of natural selection along the genome in addition to recombination rates. When modeling the effects of both background selection and recombination rate on F_{ST} using multiple linear regression, we observe that the effect of recombination rate on F_{ST} is no longer significant. We also observe substantial differences in estimates of F_{ST} between the strongest and weakest background selection regions of the human genome. We find that F_{ST} estimates at putatively neutral sites vary by as much as a factor of 2 when comparing populations from different continents. By comparison, when binning the genome by recombination rate alone, F_{ST} varies by less than 10%. Our results demonstrate the marked consequences of linked selection for generating accurate estimates of population differentiation in humans and may have implications for other measures of population structure utilizing genome-wide data.

1172T

Inference of evolutionary relationships among human populations based on the estimates of effective population size. *A. Urnikytė, A. Molytė, E. Pranckevičienė, V. Kučinskas.* Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariskiu St. 2, LT-08661 Vilnius, Lithuania.

The effective population size (*Ne*) quantifies the rate at which genetic diversity is eroded by genetic drift, a fundamental process of evolutionary change, as well as provides an insight into the demographic history and dynamics of modern human populations. The main interest of this study was to estimate the effective population size of the Lithuanian population and to elucidate the evolutionary relationships on a global scale. The data set consisted of the Lithuanian and HGDP-CEPH population samples. SNP genotyping of 294 Lithuanian population samples was performed with Illumina 770K HumanOmniExpress-12v1.0 array. The effective population size of the Lithuanian population was estimated using a new R package NeON based on linkage disequilibrium (LD) patterns. The estimates of *Ne* for the populations contained in the CEPH panel were obtained from M. Mezzavilla and S. Ghirrotto, 2015 (University of Trieste, Italy) estimated with the NeON R package. We also estimated the time of divergence between the Lithuanian and HGDP-CEPH (with a sample size > 20 individuals) populations with the NeON R package as well as. The pairwise Weir and Cockherman *Fst* between pairs of populations was estimated using the software 4P (Benazzo A et al., 2014). To visualize the evolutionary relationships among studied populations an unrooted UPGMA was calculated from the divergence time matrix using *phangorn* R package. The estimated *Ne* for Lithuanian population is 5404, confidence interval (CI) [4910; 5643]. The Ne estimates and the matrix of interpopulation *Fst* values of 23 populations studied are summarized in a neighbor-joining (NJ) phylogenetic tree. The tree provides a clear picture that recent separations of populations are related with geographical area. The separations happened more recently between Europe, Middle East and Central South Asia populations. We observe three major groups: Africans, East Asians and Europeans. The largest divergence times are observed between Africans and non-African populations. From the divergence analysis we can observe the recent separations between studied populations and reconstructed past dynamics of Ne and split times. This study is part of the LITGEN project, which was approved by the Vilnius Regional Research Ethics Committee No. 158200-05-329-79, date: 2011-05-03.

1173F

Assembling an ancestry reference panel of diverse Asian populations through genotype imputation. *C. Wang, S. Chothani.* Computational and Systems Biology, Genome Institute of Singapore, Singapore, Singapore.

Knowledge of ancestry background of the sample is important for many genetic studies, such as to control for population structure in association studies. Estimating individual ancestry often requires a reference dataset of diverse populations genotyped at a large number of markers overlapping with the study sample. In this study, we aimed to construct a publicly available Asian reference panel by combining genome-wide SNP data of 43 Asian populations from five studies. Based on 154,865 shared SNPs, we found a striking similarity between genes and geography in Asia through principal components analysis and projection Procrustes analysis, enabling us to construct a geographic reference ancestry map. We then expanded the data to ~5.5 million SNPs by genotype imputation. We evaluated imputation accuracy across diverse Asian populations, and the impact of imputation error on the estimation of ancestry and the Fst statistic based on three populations in Singapore: Chinese, Malay and Indian. We showed that imputation error tends to reduce population differentiation, as indicated by smaller Fst values based on imputed genotypes. Such impact can be greatly mitigated by filtering low-quality imputed SNPs. In contrast, ancestry estimates derived from the LASER software are more robust to imputation error than the estimation of Fst. Therefore, our Asian ancestry panel integrated through imputation will be a useful resource for ancestry estimation and other population genetic analyses on Asian populations.

1174W

Chromosome painting for arbitrary sample collections. *G. Weyenberg, D. Lawson.* Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.

Haplotype-based methods have been demonstrated to be capable of detecting fine scale structure within human populations—to the point of distinguishing genetic variation at the sub-county level in the South West of England [1]. However, the aforementioned method implements an all-against-all analysis of sampled individuals, which is not suited to all applications, including personal genomics where samples are obtained individually or in small batches. Here, we describe an extension of the FineSTRUCTURE [2] method to allow for painting of individual samples against a panel of pre-calculated reference haplotype clusters, making the method computationally feasible for on-demand analysis of individuals. The choice of the reference panel also allows the user to tailor the analysis to emphasise targeted features of the data. For example, in the context of a personal ancestry imputation, panels may be constructed to focus on global-, continental-, or national-scale genetic features, and the low computational cost of painting an individual against a pre-computed panel makes sample-level exploratory analysis feasible. Another application of the panel-based painting is to use high-quality reference data to impute unknown geographical labels to samples where such information is either unavailable, or was collected at an undesirable resolution. To demonstrate the latter application, we analysed several populations with suspected Northern-European ancestry—including the Hapmap CEU and ASW populations, and the UK10K dataset—with respect to panels of Europeans and the high-resolution People of the British Isles (POBI) samples. These individuals are characterised in terms of an admixture of inferred clusters in the reference populations. Whilst many individuals were best described as an complex admixture that likely occurred over many generations, many others had a clear signal of geographically distinct ancestry. [1] Leslie et al 2015, Nature 519:309–314 [2] Lawson et al 2012, PLoS Genet. 8:e1002453.

1175T

Inference of local pedigrees using Markov Chain Monte Carlo. *A. Ko, R. Nielsen.* University of California, Berkeley, Berkeley, CA.

As genomic datasets used for GWAS and other genetic analyses become larger, the fraction of individuals who have at least one relative in the sample increases. Family relationships are typically either ignored, by removing related individuals, or is modeled using pairwise relatedness matrices. However, the pedigree structure in itself can provide valuable information and can help increase mapping power and provide improved estimates of heritability. Despite the importance of pedigree inference, existing methods are limited to inferring only close relationships or analyzing a small number of individuals or loci. We present the first full Markov chain Monte Carlo (MCMC) for estimating pedigrees in large samples of otherwise seemingly unrelated individuals. The method supports complex pedigree structures such as polygamous families, multi-generational families, and pedigrees in which many of the member individuals are missing. Computational speed is greatly enhanced by the use of a composite likelihood function which approximates the full likelihood. Using simulations, we show that the new method leads to improved estimates of relatedness and improved mapping power in GWAS studies.

1176F

Large-scale whole genome sequencing of the Estonian population reveals novel loss-of-function variants and new insights into the population history. *M. Metspalu[1], G. Hudjashov[1,2], M. Mitt[3], L. Pagani[1,4], L. Saag[1], M. Kals[3], K. Pärn[3], L. Milani[3], D. Lawson[5], T. Esko[3], A. Metspalu[3], R. Mägi[3].* 1) Estonian Biocentre, Tartu, Estonia; 2) Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom; 5) Integrative Epidemiology Unit, School of Social and Community Medicine, 7 University of Bristol, Bristol BS8 2BN, UK.

Altogether 2244 whole genomes of geographically diverse individuals from Estonia were sequenced to a median depth of 30x using Illumina HiSeq with TruSeq PCR-free library preparation method. We found 19M SNVs and 6.6M indel variants with allele count larger than two and of which 8.4M were novel. Within this study we have analysed both loss-of-function variants revealed as well as the population structure of Estonia. We found a total of 14,531 autosomal loss-of-function (LOF) SNVs and indels in 6,991 genes. Out of these genes, 3.3% contained homozygous or compound heterozygotes LOF variants with minor allele frequency less than 2%. By combining the data of complete gene knockouts of individuals with their disease history and variety of available endophenotypes (proteomics, NMR, biochemistry) will help us to study the function of these genes and will lead to better understanding the phenotypic consequences of the variation within these genes. To study the fine-scale genetic structure of the Estonian population, we concentrate on a subset of the genomes (N=436), which comprehensively cover rural Estonia to minimize the mixing effect of historical urbanization. We further combine these genomes with a pan Eurasian panel of high coverage genomes from hundreds of populations. Using haplotype and allele frequency based methods we show that the genetic structure within Estonia is largely in line with the division of *inland* vs. *maritime* Estonia what has been proposed based on archaeological findings. Furthermore, we identify and quantify the relative contributions of the three major genetic domains of the European gene pool in Estonians and estimate split times from linguistically and geographically adjacent populations. We use Finestructure and inter-population doubleton distribution to reveal patterns of genetic sharing between Estonians and other European populations and infer population history in a series of population splits and admixture events in pre-historic and historic times.

1177W

Fine-scale population structure in southern Africa reflects ecoegeographic boundaries. *C.P. Uren[1], M. Kim[2], A.R. Martin[3,4], D. Bobo[2], C.R. Gignoux[5], P.D. van Helden[1], M. Möller[1], E.G. Hoal[1], B.M. Henn[2].* 1) SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; 2) Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY; 3) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 5) Department of Genetics, Stanford University, Stanford, CA.

Recent genetic studies have established that the KhoeSan populations of southern Africa are distinct from all other African populations and have remained largely isolated during human prehistory until about 2,000 years ago. Dozens of different KhoeSan groups exist, belonging to three different language families, but very little is known about population history within southern Africa. We examine new genome-wide polymorphism data and whole mitochondrial genomes for more than one hundred South Africans from the ≠Khomani San and Nama populations of the Northern Cape, analyzed in conjunction with 19 additional southern African populations. Our analyses reveal fine-scale population structure in and around the Kalahari Desert. Surprisingly, this structure does not always correspond to linguistic or subsistence categories as previously suggested, but rather reflects the role of geographic barriers and the ecology of the greater Kalahari Basin. Regardless of subsistence strategy, the indigenous Khoe-speaking Nama pastoralists and the N|u-speaking ≠Khomani (formerly hunter-gatherers) share recent ancestry with other Khoe-speaking forager populations that form a rim around the Kalahari Desert. We reconstruct earlier migration patterns and estimate that the southern Kalahari populations were among the last to experience gene flow from Bantu-speakers, approximately 14 generations ago. We conclude that local adoption of pastoralism, at least by the Nama, appears to have been primarily a cultural process with limited impact from eastern African genetic diffusion.

1178T

Local ancestry patterns inferred from one million genomes recapitulate fine-scale population history. *Y. Wang[1], K. Noto[1], J. Byrnes[1], R.E. Curtis[2], E. Han[1], E. Eyal[1], G. Harendra[1], P. Carbonetto[1], A.R Kermany[1], J.M. Granka[1], K.A. Rand[1], N.M. Natalie[2], E.L. Hong[1], C.A. Ball[1], K.G. Chahine[2].* 1) Ancestry.com DNA, LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT.

In a country of immigrants, population structure is shaped by a long, ongoing history of immigration, followed by subsequent admixture and migration. All these events have left their footprints in the genomic landscape of current residents and make it possible for geneticists to reconstruct population history from genomic data. However, deciphering the signature of these forces requires accurate inference of genomic tracts that one individual inherits from ancestors of different origins. Previously, several methods have been developed for inferring local ancestry with varying levels of success. Unfortunately, none of these methods can be feasibly applied to a data set of one million genomes. Recently, our team presented *Polly*, a novel algorithm for estimating genome-wide ancestry proportions in admixed samples. *Polly*, built on a modified version of the BEAGLE haplotype model, relies on this model to achieve two things: First, to account for phasing uncertainly, and second, to provide a measure of distance between a query haplotype and a reference haplotype. Using haplotype models learned from hundreds of thousands of haplotypes and subsequently annotated with over eight thousand single-origin reference individuals, *Polly* performs ultra-fast inference of both global and local ancestry. In this study, we evaluate *Polly*'s accuracy in predicting local ancestry using simulated admixed samples with known genomic composition. We assess the assignment accuracy, the switching pattern and the tract length distribution. Using cross-validation experiment, we confirm that *Polly* makes highly accurate local ancestry estimates even at the subcontinental level. We further use *Polly* to analyze one million genomes from the United States and discover distinct local ancestry patterns among different ethnic groups and communities, especially among African Americans and Latino Americans. We map local ancestry estimates to individuals' geographic locations. Our results illustrate clear population structure arising from immigration routes, assortative mating and isolation by distance. We also find evidence that supports large-scale domestic migration events, as exemplified by the Great Migration of African Americans following the abolition of slavery. Finally, we attempt to date known historical events from ancestry tract length distributions. Overall, our analysis demonstrates the power of combining local ancestry analysis with big data in studying fine-scale population history.

1179F

Identifying type 1 diabetes risk variants in a low prevalence population.
S.S. Rich[1,2], S. Onengut-Gumuscu[1,2], W-M. Chen[1,2], E. Farber[1], J. Bonnie[1], P. Concannon[3,4], Type 1 Diabetes Genetics Consortium (T1DGC). 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA; 3) Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL; 4) Genetics Institute, University of Florida, Gainesville,.

Type 1 diabetes (T1D) is one of a family of disorders characterized by dysregulation of glucose homeostasis. In T1D, an autoimmune process of unknown origin results in the destruction of the insulin-secreting pancreatic β cells and a life-long dependence on exogenously-administered insulin for survival. The reported incidence of T1D varies from 0.1 to 40.9 per 100,000 individuals, with the highest incidence rates in European ancestry populations and the lowest rates in Asian and African ancestry populations. To date, studies of the genetics of T1D have focused almost exclusively on European ancestry populations with ascertainment of large numbers of individual cases or multiple case families. GWAS and subsequent fine-mapping efforts, carried out by the T1DGC, identified more than 40 loci with significant evidence of association with T1D risk. Here, we extend these studies to subjects of African ancestry. We determine which of the T1D-associated SNPs in risk loci identified in European ancestry populations influence risk in African ancestry individuals, fine-map known T1D risk regions using linkage disequilibrium (LD) patterns in African ancestry samples, and identify new T1D risk loci. We densely genotyped 1036 T1D case and 3220 control samples of African ancestry using a custom genotyping panel (ImmunoChip), consisting of ~200K SNPs in 186 loci associated with autoimmunity (T1D and 11 other immune-mediated diseases). A total of 14 of the previously reported T1D risk loci exhibited evidence of association ($P < 1 \times 10^{-4}$) in the African ancestry T1D collection, and 4 other loci had nominal evidence of association ($1 \times 10^{-4} < P < 1 \times 10^{-3}$). We identified 2 novel regions (1q25.1, $P = 4.98 \times 10^{-7}$, OR=1.32, 95% CI=1.19–1.48, and 16q23.3, $P = 4.02 \times 10^{-7}$ OR=1.34, 95% CI=1.19–1.50) associated with T1D. The LD pattern differences between African and European ancestry samples have allowed us to refine associations in T1D risk loci, reducing the number of credible SNPs by 50% in many loci. The biological influence and detailed functional role of each risk variant will need to be further elucidated in future functional studies.

1180W

Emerging pattern of possible genetic and lifestyle risk factors associated with diabetic retinopathy: An Indian cohort study.
A. Bhanwer[1], R. Sikka[1], V. Vig[2]. 1) Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Dr. Sohan Singh Eye Hospital, Amritsar Punjab, India.

Diabetic Retinopathy (DR) is a complex polygenic complication of T2D in which both genetic and environmental factors play cumulative role for the disease development and progression. The previous candidate gene studies mainly focused on *VEGF, RAGE, eNOS, TCF7L2*; *TNFα*; *ALR* and *ACE* polymorphisms. The major limitation of these studies was combining of heterogeneous phenotype in DR cases and smaller sample size. Therefore, the present study is designed to investigate the various lifestyle risk factors responsible for the predisposition/progression of DR and to undertake comprehensive analysis of 30 SNPs predisposing to DR in a large cohort from North-West India (Punjab). A total of 414 DR cases and 434 gender and ethnicity matched above the age of 50 years healthy controls were screened with the help of Sequenom MassARRAY technology. The clinical and anthropological data analysis revealed that the mean values of Waist, Hip circumference, Waist-Hip Ratio and Waist-Height Ratio were higher among cases which represented role of central obesity towards DR. History of hypertension [p<0.001; OR=13.64 (9.15=20.34)], male gender [p=0.003; OR=1.86 (1.22-2.81)], BMI [p<0.001; OR=1.31 (1.17-1.46)] and increasing age were the main risk factor for the DR. Single locus analysis revealed that *TCF7L2* rs12255372 and rs11196205 were significantly associated with DR (p<0.05). Haplotype analysis revealed that *TCF7L2* (rs12255372-rs11196205-rs6585205) C-T-T and C-T-G haplotype provided risk however, G-G-T and G-G-G haplotypes provided protection towards DR. Strong linkage disequilibrium was observed for *TLR-4*, *TCF7L2* and *AKR1B1* among controls. The present study provides some insights for the various lifestyle and genetic risk factors in the studied population. .

1181T

HLA-DPB1 and type 1 diabetes in Han Chinese. *Y. Lee[1,3,4], C. Huang[2], W. Ting[2,6], F. Lo[5], C. Lin[1], T. Chang[1], H. Yang[1], W. Lin[1].* 1) Dept Pediatrics & Med Res, MacKay Memorial Hosp, New Taipei City, Taiwan; 2) Dept Pediatrics, MacKay Children's Hospital; 3) Dept Pediatrics, Taipei Medical University; 4) Institute of Biomedical Sciences, MacKay Medical College; 5) Dept Pediatrics, Chang-Gung Memorial Hospital, Taoyuan, Taiwan; 6) Dept Nursing, MacKay Medicine, Nursing and Management College.

Type 1 diabetes (T1D) is an autoimmune disease characterized by T cell-mediated destruction of pancreatic β cells and the presence of autoantibodies against several β cell autoantigens. Both genetic susceptibility and environmental factors contribute to the pathogenesis.Because the destruction of insulin-producing β cells in type 1 diabetes is likely to involve the activity of CD4+ T lymphocytes and because T-lymphocyte activating is dependent on recognition of MHC class II molecules. We reported association between DPB1 gene and type 1 diabetes in Han Chinese living in Taiwan.**Patients**The subjects were 404 T1D patients (181 boys and 223 girls) aged 8.0 (SD 4.1) years. Patients had a HbA1c level of ≥6.5%, a fasting plasma glucose level ≥ 7 mmol/l (126 mg/dl) at least 2 times, or a random glucose level ≥11.1 mmol/l (200 mg/dl) with diabetic symptoms, and at least one of autoantibodies to islet cell antigens, glutamic acid decarboxylasee (GAD) and Islet antigen-2 (IA-2).**Controls**The 381 control subjects consisted of 135 males and 246 females. They included hospital personnel and individuals who underwent routine health examinations or minor surgery. None had a history of autoimmune disease.All patients and controls were Han Chinese in Taiwan. Our institutional review board approved this study and all subjects or their guardians gave informed consent.**Genotyping of the DPB1 gene**We used sequencing-based typing to genotype the DPB1 gene.**Statistical analysis**Statistical difference in genotype, allele, and carrier distributions between patients and controls were assessed by the chi-square test. Odds ratios (OR) and 95% confidence intervals were also calculated.The Bonferroni correction, Pc = 1 –(1-P)n, was used for multiple comparisons where Pc is the corrected P value, P the uncorrected value, and n the number of comparisons. In this study, n is 13 for each allele, or carrier. A Pc value of less than 0.05 was considered statistically significant.**Results**The DPB1*05:01 allele was the most frequent in patients (30.2%) and controls (46.7%). It was also significantly less frequent in patients than in controls, OR (95%CI)=0.49 (0.40-0.61), Pc = 2.16E-10. The DPB1*04:01 allele was the second most frequent in patients (25.9%) and in controls (6.7%) and conferred susceptibility of T1D, OR= 4.86 (3.59-6.59), Pc = 2.22E-23.**Conclusion**DPB1 was associated with T1D in Han Chinese in Taiwan. DPB1*04:01conferred susceptibility to T1D while *05:01 rendered protection against T1D.

1182F

Childhood predictors of adult fatty liver - The Cardiovascular Risk in Young Finns Study. *A. Ahola-Olli[1], M. Oikonen[1], N. Pitkänen[1], E. Suomela[1], J. Virtanen[2], R. Parkkola[2], E. Jokinen[3], T. Laitinen[4], N. Hutri-Kähönen[5], M. Kähönen[6], T. Lehtimäki[7], L. Taittonen[8], P. Tossavainen[9], A. Jula[10], B.M. Loo[10], V. Mikkilä[11], R. Telama[12], J.S. Viikari[13], M. Juonala[13], O.T. Raitakari[14].* 1) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 2) Department of Radiology, Turku University Central Hospital, Medical Imaging Centre of Southwest Finland, Turku, Finland; 3) Department of Pediatric Cardiology, Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland; 4) Department of Clinical Physiology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Department of Pediatrics, University of Tampere and Tampere University Hospital, Tampere, Finland; 6) Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland; 7) Fimlab Laboratories and Department of Clinical Chemistry, School of Medicine, University of Tampere, Tampere, Finland; 8) Vaasa Central Hospital, Vaasa, Finland and Department of Pediatrics, University of Oulu, Oulu, Finland; 9) Department of Children and Adolescents, Oulu University Hospital, PEDEGO Research Unit and Medical Research Center Oulu, University of Oulu, Oulu, Finland; 10) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland; 11) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; Division of Nutrition, Department of Applied Chemistry and Microbiology, University of Helsinki, Helsinki, Finland; 12) LIKES-Research Center for Sport and Health Sciences, Jyväskylä; 13) Department of Medicine, University of Turku, and Division of Medicine, Turku University Hospital, Turku, Finland; 14) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.

Fatty liver without excessive alcohol consumption is the most common form of chronic liver disease in obese and diabetic people living in the Western countries. Since there is no cure for advanced fatty liver, the prevention of this disease is vital. The main modifiable risk factors assocaited with fatty liver include obesity, hyperlipidemia and type 2 diabetes. These risk factors tend to track from childhood to adulthood and therefore their early prevention is important. The current study aims to identify childhood risk factors that are associated with fatty liver later in adulthood. The study includes 3,596 Finnish children at baseline who have been followed for over 30 years. The subjects were randomly chosen from national population registers from cities of Tampere, Turku, Helsinki, Oulu, Kuopio and Turku in addition to their rural surroundings. The study began in 1980 and the latest follow-up study was organized in 2011. This follow-up included liver ultrasound examination along with extensive laboratory measures. Genes previously associated with liver fat accumulation include *PNPLA3* and *TM6SF2*. Variants from these genes (rs738409 and rs58542926) predict the development of fatty liver (OR (95 %CI) 1.63 (1.29-2.07); 1.57 (1.08-2.30), respectively) after adjusting for age, sex, body-mass index, insulin concentration, and birth weight. When the genetic variants from *PNPLA3* and *TM6SF2* were added to prediction model containing standard clinical risk factors the net reclassification and integrated discrimination index were significantly improved. Our results demonstrate that prediction of adulthood fatty liver can be significantly improved after taking genetic factors in account. This kind of genetic risk stratification can be used to target treatments to high risk individuals. This could reduce the number of subjects needed to treat in order to prevent one clinical end point.

1183W

Thrifty functional characteristics of the major variant in *CREBRF* associated with body mass index in Samoans. *C.T. Su[1], R.L. Minster[1], N.L. Hawley[2], E.E. Kershaw[3], R. Deka[4], D.E. Weeks[1,5], S.T. McGarvey[6,7], Z. Urban[1]. 1)* Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Epidemiology (Chronic Disease), Yale University School of Public Health, New Haven, CT; 3) Department of Medicine, Division of Endocrinology, University of Pittsburgh, Pittsburgh, PA; 4) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH; 5) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 6) International Health Institute, Department of Epidemiology, Brown University School of Public Health, Providence, RI; 7) Department of Anthropology, Brown University, Providence, RI.

Pacific islanders, and among them Samoans, have the highest rates of obesity worldwide. Our genome-wide association study of body mass index in Samoans and subsequent targeted sequencing, fine mapping and selection studies, identified a nonsynonymous coding variant (rs373863828, c.1370G>A, p.Arg457Gln) in *CREBRF* as the most significantly associated and likely causative variant. To evaluate the functional consequences of this variant, we performed overexpression, knockdown, adipocyte differentiation, neutral lipid accumulation and metabolic profiling studies using a mouse 3T3-L1 pre-adipocyte cell model. Overexpression of *CREBRF* significantly induced adipocyte differentiation, lipid accumulation, glycolysis, oxidative phosphorylation, and protected cells from starvation-induced death. Knockdown of *CREBRF* produced the opposite effect for each of these variables. Overexpression of the p.Arg457Gln variant *CREBRF* increased lipid accumulation more than wild type, but decreased both glycolysis and oxidative phosphorylation. However, it protected against starvation-induced cell death the same as the wild type *CREBRF*. Thus, p.Arg457Gln variant could cause obesity-related cellular phenotypes by promoting the accumulation of neutral lipids while using less energy than wild type *CREBRF*. Prioritization of nutritional energy storage over energy use by p.Arg457Gln *CREBRF* supports a "thrifty variant" hypothesis for obesity and suggests that this variant could be selected for in populations subject to nutritional insecurity, as archeological and historical evidence indicates Samoans may have been.

1184T

Identification of interactions in the *MADD-PTPRN-RAB3* complex with type 2 diabetes in multiple populations. *G. Chen[1], T. Cai[2], F. Tekola-Ayele[1], A. Doumatey[1], J. Zhou[1], A. Adeyemo[1], A. Notkins[2], C. Rotimi[1]. 1)* NHGRI/NIH, National Institutes of Health, Bethesda, MD; 2) NIDCR/NIH, National Institutes of Health, Bethesda, MD.

The missing heritability for complex traits including type 2 diabetes (T2D) remains a major unresolved challenge; It has been hypothesized that gene-gene interactions may explain some of this missing heritability. Genes encoding proteins in insulin pathway of islet beta cells are known to be associated with T2D. We investigated gene-gene interactions among *MADD, PTPRN/PTPRN2*, and *RAB3A/RAB3C*, three genes that encode proteins known to form a complex that has a molecular link with insulin-containing dense core secretory vesicles (DCVs). Using the model based multifactor dimensionality reduction (MB-MDR) method , we identified statistically significant gene–gene interactions among these genes with respect to T2D status in eight independent studies, including exome SNP array analyses in three studies [i.e., the Africa America Diabetes Mellitus (AADM, n = 2,189, p value 3.1×10^{-5}), Howard University Family Study (HUFS, n = 1,841, p value 4.7×10^{-5}), Chinese American Diabetes Mellitus (CADM, n = 1,908, p value 5.7×10^{-6})]. The other studies were GWAS SNP array analyses in five additional studies [i.e., the Atherosclerosis Risk in Communities study (ARIC, n = 3,137, p value 3.9×10^{-8}) , Cleveland Family Study (CFS, n = 653, p value 6.2×10^{-8}), HUFS (n = 1,976, p value 4.0×10^{-9}), Jackson Heart study (JHS, n = 2,187, p value 4.7×10^{-8}), and Multi Ethnic Study of Atherosclerosis (MESA, n = 1,611, p value 1.9×10^{-5})]. Notably, single gene analyses did not find any significant association between these genes and T2D in any of these studies. These interactions are consistent with previous studies that showed a molecular link between the MADD, PTPRN/PTPRN2, and RAB3A/RAB3C complex and insulin-containing DCVs. *PTPRN/PTPRN2* is a specific linker for tethering *MADD-RAB3A/RAB3C* complex (in *MADD* domain) to DCV. MADD is implicated in the GDP/GTP exchange of RAB3A/RAB3C to provide energy to move cargo to membrane and release insulin. In summary, gene–gene interactions between *MADD, PTPRN/PTPRN2*, and *RAB3A/RAB3C* were identified in association with T2D across eight independent multi-ethnic studies that included a total of 15,502 participants. The underlying biological mechanism is likely to involve how the complex of *MADD, PTPRN/PTPRN2*, and *RAB3A/RAB3C* is linked to DCVs thereby influencing the process of insulin release.

1185F

An exome-wide association study for type 2 diabetes-attributed end-stage kidney disease in African Americans. *M. Guan[1,2], J.M. Keaton[2], L. Dimitrov[2], P. Mudgal[2], P.J. Hicks[2], J.G. Wilson[3], B.I. Freedman[4], D.W. Bowden[2,5,6], M.C.Y. Ng[2,5].* 1) Integrative Physiology and Pharmacology Program, Wake Forest School of Medicine, Winston-Salem, NC, USA; 2) Center for Human Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 3) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 4) Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA; 5) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 6) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA.

African Americans (AAs) are at higher risk for developing end-stage kidney disease (ESKD) compared to European Americans. Genome-wide association studies (GWAS) have identified variants associated with diabetic and non-diabetic kidney disease. However, common SNPs identified in GWAS are largely non-coding variants, have modest effects, and cumulatively explain only a small proportion of disease variance. Therefore, we evaluated the contribution of coding variants in type 2 diabetic (T2D) patients on dialysis or with high risk for T2D-ESKD (eGFR<30 ml/min/1.73 m^2). Exome-sequencing data in 1,391 AA subjects including 456 T2D-ESKD cases and 935 non-diabetic, non-nephropathy controls from the T2D-GENES Consortium were examined. Single-variant association was performed for an additive effect with a mixed logistic regression model that corrects for sample structure and hidden relatedness using the analysis package GMMAT. A total of 618 nominal associations ($P<0.01$) from exome-sequencing were replicated in 2,041 T2D-ESKD cases and 1,140 non-diabetic, non-nephropathy controls genotyped on a custom array. Discrimination analysis in 667 T2D cases lacking nephropathy was performed to exclude T2D-associated SNPs. Meta-analysis of 4,572 discovery and replication samples revealed five modest associations ($P<1E-4$) with T2D-ESKD at *DLGAP5* (rs10144326), *PRX* (rs268671), *APOL1* (rs73885319), *RAD51AP2* (rs834514) and *RREB1* (rs41302867). Removal of *APOL1* renal-risk genotype carriers identified an additional association at *IGSF22* (rs61886892) and confirmed association in *RREB1* (rs41302867). Coding variants at *PRX* (rs268671, V882A, $P=2.54E-5$, OR=1.33), *APOL1* (rs73885319, S338G, $P=1.21E-5$, OR=1.32), and *RAD51AP2* (rs834514, G1037D, $P=3.95E-5$, OR=1.32) are predicted to disrupt protein function. These findings may partially explain the genetic predisposition to T2D-ESKD in AAs. Future directions include replication in additional AA populations and evaluation of the contribution of T2D-ESKD associated genetic variants to the susceptibility of non-diabetic ESKD. Functional studies are also warranted to illustrate the potential mechanisms underlying the effect of these newly identified genetic associations.

1186W

Exome sequencing of >20,000 Finnish individuals identifies rare variants influencing cardiovascular and metabolic phenotypes. *A.E. Locke[1,2], K.M. Steinberg[1], S. Service[3], H. Abel[1], V. Ramensky[3], M. Pirinen[4], L. Stell[5], C. Chiang[3], H.M. Stringham[2], A.U. Jackson[2], P. Yajnik[2], D. Ray[2], D.E. Larson[1], D.C. Koboldt[1], L.J. Scott[2], R.S. Fulton[1], J. Nelson[1], T.J. Nicholas[1], N.O. Stitziel[1], I.M. Hall[1], C. Sabatti[5,6], S. Ripatti[4,7,8], V. Salomaa[9], A. Palotie[4,7,10], M. Laakso[11], M. Boehnke[2], R.K. Wilson[1], N. Freimer[3].* 1) McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Department of Biomedical Data Science, Stanford University, Stanford, CA; 6) Department of Statistics, Stanford University, Stanford, CA; 7) Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland; 8) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 9) National Institute for Health and Welfare, Helsinki, Finland; 10) Broad Institute of MIT and Harvard, Cambridge, MA; 11) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland.

Low frequency, deleterious variants are enriched in the Finnish population due to recent bottlenecks, increasing power to detect genetic associations for complex traits. We sequenced the exomes of 20,029 individuals from two large population-based longitudinal studies of metabolic and cardiovascular disease in Finland: METSIM and FINRISK. We identified >1.5M SNPs and >138,000 short indels, of which nearly 90% have MAF<1%; 44,126 are loss-of-function, and 420,940 are protein-altering. PolyPhen2 predicts 168,969 of these non-synonymous SNPs to be damaging to the resulting protein, and 54,933 are predicted to be damaging by five different prediction algorithms used in Purcell et al. (Nature 2014). In addition to this robust genetic dataset, we have data for 255 quantitative measures of cardiovascular and metabolic health, covering lifestyle factors, anthropometrics and obesity, blood pressure, kidney function, inflammation, glycemic traits, and >200 measures from proton NMR. The population history of Finland, the breadth of phenotypes, and the large scale of this exome sequencing study allow us a unique opportunity to identify rare coding variation influencing cardiovascular and metabolic traits. We tested each phenotype for genetic associations using EMMAX and EMMAX-SKAT-O. Using an exome-wide significance level of $5x10^{-7}$, we identified 248 associations at 79 variants with MAF<1% in 112 unique phenotypes. Sixty associations, at 13 variants, remain significant after Bonferroni correction based on an assumption of 255 independent traits ($1.9x10^{-9}$). We continue to work with multivariate approaches to account for the correlation between phenotypes. Among the associated rare variants are two missense variants in *GLDC* (MAFs=.0098 and .0025) and a premature stop codon in *ALDH1L1* (MAF=.0026) associated with glycine levels, a frameshift variant (MAF=.0015) in *SLC16A10* associated with increased tyrosine levels, a splice variant in *ADIPOQ* (MAF=.0026) associated with decreased adiponectin, a splice donor variant in *CETP* associated with increased serum ApoA1, and a premature stop variant and splice variant (MAFs=.0001) in *PKD1* both associated with increased serum creatinine levels. We also identified a splice donor variant in *APOC3*; missense variants in *LIPC*, *LIPG, and CD300LG;* frameshift variants in *LDLR* and *ANGPTL4*; and a premature stop variant in *c19orf80* associated with altered levels of multiple lipoprotein particles, primarily subclasses of LDL and HDL.

1187T

Metabolomics of insulin resistance and impaired insulin secretion: A multi-cohort Mendelian randomization study identifies causal effects on monounsaturated fatty acid and tyrosine levels. *C. Nowak[1], S. Salihovic[1], A. Ganna[2], S. Brandmeier[3,4], T. Tukiainen[5], C.D. Broeckling[6], P.K. Magnusson[7], J. Prenni[6], R. Wang-Sattler[3,4,8], A. Peters[4,8,9], K. Strauch[10,11], T. Meitinger[12,13], V. Giedraitis[14], J. Ärnlöv[15,16], C. Berne[17], C. Gieger[3,4,8], S. Ripatti[5,18,19], L. Lind[16], N.L. Pedersen[7], J. Sundström[16], E. Ingelsson[1,20], T. Fall[1].* 1) Department of Medical Sciences and Science for Life Laboratory, Molecular; 2) Massachusetts General Hospital, Harvard Medical School and Broad Institute, Boston, Massachusetts, USA; 3) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, München-Neuherberg, Germany; 4) Institute of Epidemiology II, Helmholtz Zentrum München, München-Neuherberg, Germany; 5) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 6) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, USA; 7) Department of Medical Epidemiology and Biostatistics (MEB), Karolinska Institutet, Stockholm, Sweden; 8) German Center for Diabetes Research (DZD), München-Neuherberg, Germany; 9) Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA; 10) Institute of Genetic Epidemiology, Helmholtz Zentrum München, München-Neuherberg, Germany; 11) Institute of Medical Bioinformatics, Biometry and Epidemiology, Ludwig-Maximillians-University, München, Germany; 12) Institute of Human Genetics, Helmholtz Zentrum München, München-Neuherberg, Germany; 13) Institute of Human Genetics, München, Germany; 14) Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden; 15) School of Health and Social Studies, Dalarna University, Falun, Sweden; 16) Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden; 17) Department of Medical Sciences, Clinical Diabetology and Metabolism, Uppsala University, Uppsala, Sweden; 18) Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 19) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 20) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.

   Insulin resistance (IR) and impaired insulin secretion contribute to type 2 diabetes and cardiovascular disease. Both are associated with changes in the circulating metabolome, but causal directions have been difficult to disentangle. We combined untargeted plasma metabolomics by liquid chromatography/mass spectrometry in three non-diabetic cohorts with Mendelian Randomization (MR) analysis to obtain new insights into early metabolic alterations in IR and impaired insulin secretion. In up to 910 elderly men we found associations of 52 metabolites with hyperinsulinemic-euglycemic clamp-measured IR and/or β-cell responsiveness (disposition index). These implicated bile acid, glycerophospholipid and caffeine metabolism for IR and fatty acid biosynthesis for impaired insulin secretion. In MR analysis including two additional cohorts (n = 3,342) followed by replication in three independent studies profiled on different metabolomics platforms (n = 7,824 / 8,961 / 8,330), we discovered and replicated causal effects of IR on lower levels of palmitoleic acid and oleic acid, as well as on higher levels of tyrosine. In summary, in one of the largest studies combining "gold standard" measures for insulin responsiveness with non-targeted metabolomics, we found distinct metabolic profiles related to IR or impaired insulin secretion. We hypothesize that the causal effects on monounsaturated fatty acid levels explain parts of the raised cardiovascular disease risk in IR that is independent of diabetes development.

1188F

A coding 27bp deletion in *SLC4A1* is associated with 0.53% lower HbA1c in Malays. *X. Sim[1], C. Wang[2], J.F. Chai[1], S.L. Kao[3], A.I. Podgornaia[4], S. Chothani[2], X. Liu[5], C.Y. Cheng[6,7,8], C. Sabanayagam[6], Y.Y. Teo[1,2,9,10,11], T.Y. Wong[6,7,8], R.M. van Dam[1], J. Liu[2], D.F. Reilly[4], A. Paterson[12,13], E.S. Tai[1,3,7].* 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; 2) Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore; 3) Division of Endocrinology, Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore, Singapore; 4) Merck Research Laboratories, Kenilworth, NJ, USA; 5) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 6) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 7) Duke-NUS Medical School Singapore, Singapore, Singapore; 8) Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore; 9) Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore; 10) Life Sciences Institute, National University of Singapore, Singapore, Singapore; 11) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore; 12) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 13) Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

   Glycated hemoglobin (HbA1c;%) is produced by non-enzymatic glycation of hemoglobin by glucose, and widely used as a long-term biomarker of glycemic status. We previously identified a genome-wide significant signal for HbA1c in 1,721 Singapore Malays without diabetes from the Singapore Malay Eye Study (SiMES) ($P = 3.89\times10^{-22}$) on chromosome 17q21.31. The lead SNP rs12603404 (MAF = 3%) was an intronic variant in *C17orf53*, and the minor allele was associated with lower HbA1c by 0.33%. However, this association was not replicated in other populations where allele frequencies were similar. To test the hypothesis that this represents an ancestry-specific association, we extended our study to an independent set of 1,003 Singapore Malays and 1,206 Singapore Chinese with both genome-wide array and whole exome sequences. We replicated the association at this locus with meta-analysis combined P-value of 2.15×10^{-34} in 2,724 Malays (lead variant chr17: 4227290; intronic on *ATXN7L3*). No association was observed in Chinese. We further tested the exome sequence data for association with HbA1c in Malays, and the lead variant was a 27bp deletion chr17:42335410 on *SLC4A1* (MAF = 1.5%; $P = 9.33\times10^{-11}$). This deletion was not observed in Chinese. Conditional analysis on the 27bp del eliminated the association at both rs12603404 and chr17: 4227290. Adjustment for fasting glucose did not attenuate the association with HbA1c. We then combined GWAS array and exomes to create a Malay-specific reference panel, and imputed the 27bp del into SiMES. The strongest association with HbA1c in SiMES was observed at the well-imputed 27bp del (imputation $R^2 = 0.94$; $P = 1.38\times10^{-32}$). We confirmed a population-specific association between a 27bp deletion at *SLC4A1* and HbA1c in Malays (meta-analysis beta = -0.53%; $P = 1.38\times10^{-39}$). This deletion results in the deletion of nine amino acids from *SLC4A1*, and is the cause of Southeast Asian ovalocytosis, an asymptomatic hereditary red blood cell condition (OMIM #166900), suggesting that this association is mediated via a non-glycemic pathway. The presence of this variant may have implications for the diagnosis of diabetes using HbA1c in this population.

1189W

Aggregate effects of type 2 diabetes genetic loci on diabetic retinopathy in a multi-ethnic Asian population. *C.Y. Cheng[1,2,3], Y.H. Chong[2], Y.C. Tham[1], A. Gan[1], S.P. Tan[1], G. Tan[1,2], J.J. Wang[4], P. Mitchell[4], T.Y. Wong[1,2,3], The Singapore Epidemiology of Eye Diseases Study.* 1) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; 2) Duke-NUS Medical School, Singapore; 3) Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 4) Centre for Vision Research, Department of Ophthalmology and Westmead Millennium Institute, University of Sydney, Sydney, Australia.

   Purpose: Genetic association studies to date have not identified any widely reproducible risk loci for diabetic retinopathy (DR). We hypothesized that individuals having more diabetes genetic risk alleles had a higher risk of developing DR. We aimed to evaluate the aggregate effects of multiple type 2 diabetes-associated genetic variants on DR.   Research Design and Methods: A total of 1528 diabetic participants (480 Malays, 750 Indians and 298 Chinese) were recruited from the Singapore Epidemiology of Eye Diseases Study, of which 547 (35.8%) were diagnosed to have DR. Participants underwent a comprehensive ocular examination, including dilated fundus photography. Retinal photos were graded using the modified Airlie House classification system to assess the presence and severity of DR following a standardized protocol. We identified 76 previously discovered type 2 diabetes-associated single nucleotide polymorphisms (SNPs) and constructed multi-locus genetic risk scores (GRSs) for each individual by summing the number of risk alleles for each SNP weighted by the respective effect estimates on DR. Two GRSs were generated: an overall GRS that includes all 76 discovered type 2 diabetes-associated SNPs, and an Asian-specific GRS that includes a subset of 55 SNPs which were previously found to be associated with type 2 diabetes in East and/ or South Asian ancestry populations. Associations between the GRSs with DR were determined using linear and logistic regression analyses. Discriminating ability of the GRSs was determined by the area under the receiver operating characteristic curve (AUC).   Results: Participants in the top tertile of the overall GRS were 2.56 times (95% CI: 1.92–3.40, P = 1.5 x 10[-10]) likely to have DR compared to participants in the bottom tertile. Participants in the top tertile of the Asian-specific GRS were 2.00 times (95% CI: 1.51–2.65, P = 1.3 x 10[-6]) likely to have DR compared to participants in the bottom tertile. Both GRSs were associated with higher DR severity levels (P < 0.001). We did not observe any inter-ethnic differences. Addition of the GRSs to traditional risk factors improved the AUC modestly (AUC difference = 0.02–0.03, P < 0.05). Conclusions: GRSs constructed from type 2 diabetes-associated SNPs were significantly associated with greater risks of DR. Our findings may provide new insights to further our understanding of the genetic pathogenesis of DR.

1190T

Relative contribution of type 1 and type 2 genetic risk loci to latent autoimmune diabetes in adults. *D.L. Cousminer[1,2], R. Mishra[1], A. Chesi[1], M.I. Hawa[3], J.P. Bradfield[1], K.M. Hodge[1], V.C. Guy[1], H. Hakonarson[1], D. Mauricio[4], N. Schloot[5], K.B. Yderstræde[6], B.F. Voight[2], S. Schwartz[7], R.D. Leslie[3], B.O. Boehm[8,9], S.F.A. Grant[1,2].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA; 3) Queen Mary University of London, London, UK; 4) Hospital Universitari Germans Trias i Pujol, Spain; 5) German Diabetes Center, Düsseldorf, Germany; 6) Odense University Hospital, Odense, Denmark; 7) Main Line Health System, Wynnewood, PA; 8) University of Ulm , Ulm, Germany; 9) Nanyang Technological University, Singapore.

   Latent autoimmune diabetes in adults (LADA) exhibits features of both type 1 (T1D) and type 2 diabetes (T2D), and its genetic etiology remains largely unresolved. In this study, we investigated 67 T1D and 71 T2D established genetic risk loci in a large cohort of LADA cases to assess where LADA is situated relative to these two well-characterized, classic forms of diabetes. We tested the association of these loci in 978 LADA cases recruited from the ActionLADA and German multicenter studies versus 1,057 healthy North American controls with European ancestry, and tested the difference between the LADA effect size versus that in T1D or T2D from the Wellcome Trust Case Control Consortium. We found that LADA is genetically more similar to T1D than to T2D, with the exception of an association at the T2D *HNF1A* locus (*P* = 3.4 x 10[-4]). Several T1D loci were associated with LADA (P < 7 x 10[-4]), including the major histocompatibility complex (MHC), as well as the *PTPN22*, *NAA25*, and *INS* loci. Contrary to previous studies, the key T2D risk allele at *TCF7L2* (rs7903146-T) was not associated with LADA and had a significantly lower frequency in LADA than in T2D cases (*P*difference = 5.21 x 10[-6]). A genetic risk score (GRS) weighted across T1D loci also better distinguished between LADA cases and controls than one constructed from T2D loci (AUC = 0.63 and 0.57, respectively). Furthermore, in LADA cases positive for two diabetes autoantibodies (GAD and IA2), the similarity between LADA and T1D became more apparent, both for individual variants and the performance of the T1D GRS (AUC = 0.71). Our results show that the genetic spectrum of T1D extends into adulthood and cases who do not initially require insulin; however, the genetic load of T1D risk alleles is less severe than in childhood-onset T1D. Additionally, T2D genetic risk also plays a small role in LADA, highlighting the potential for GRS in reclassifying and understanding diabetes subtypes.

1191F

Effect of genetic polymorphism of OCT and MATE transporters on metformin response in Korean type 2 diabetes mellitus patients. *N. Gu[1], J.E. Park[2], K.H. Shin[2].* 1) Department of Clinical Pharmacology and Therapeutics, Dongguk University College of Medicine and Ilsan Hospital, Goyang, Republic of Korea; 2) Pharmacotherapy & Translational Research Lab. College of Pharmacy & Research Institute of Pharmaceutical Sciences, Kyungpook National University, Daegu, Republic of Korea.

Background: Metformin is an oral antihyperglycemic agent, the most widely used to treat type 2 diabetes. It improves insulin sensitivity and decreases the insulin resistance. Metformin is transported into the hepatocytes and renal epithelium by organic cation transporter (OCTs) and then excreted into the bile and urine by multidrug and toxin extrusion (MATE). The current study was aimed to evaluate influence of genetic polymorphism of several transporters on metformin response in T2DM patients. **Methods:** Twenty Korean type 2 diabetic patients prescribed metformin were enrolled in the study. Their hemoglobin A1c (HbA1c), fasting plasma glucose (FPG) and insulin levels were measured at baseline, 3 months and 4 months after the first administration. We classified responders as follows; whose HbA1c levels at the end of 3 months decreased by more than 0.5% from the baseline. To evaluate the influence of genetic polymorphisms of OCT1, OCT2, OCTN1, MATE1 and MATE2-K, genomic DNA was extracted and analyzed using Taqman genotyping kit by real-time PCR. **Results:** The analysis was carried out for 20 patients with 7 responders and 13 non-responders. HbA1c %change (mean standard deviation) in responders and non-responders were -1.00 0.48, 0.003 0.34., respectively. For genotype frequencies of OCTN1 (rs272893), both CC and CT types were 43% in responders, otherwise CT type (84.62%) is the highest in non-responders. For MATE1 (rs2289669) genotype, the highest type was AG gene as 71% in responders and the next was AA type (50%) in non-responder group. **Conclusion:** Among the five selected genes, genetic polymorphisms of OCTN1 and MATE1 showed different pattern between responder and non-responder groups.

1192W

Prevalence of vitamin D receptor (*VDR*) gene polymorphisms in Kuwaiti Arab children with type-1 diabetes mellitus. *M.Z. Haider[1], M.A. Rasoul[1,4], M. Al-Mahdi[2], H. Al-Kandari[3], G.S. Dhaunsi[1,4].* 1) Dept Pediatrics, Fac Med, Kuwait Univ, Safat, Kuwait; 2) Dept Pediatrics, Al-Adan Hospital, Kuwait; 3) Dept Pediatrics, Farwania Hospital, Kuwait; 4) Dept Pediatrics, Mubarak Al-Kabeer Hospital, Kuwait.

Introduction & Objectives: Type-1 Diabetes Mellitus (T1DM) is a multifactorial disease in which a complex interplay of genetic and environmental factors modulate the immune system and result in destruction of insulin producing beta cells in pancreas. Recent evidence indicate that production and degradation of vitamin D is a major signaling component in both innate and adaptive immunity. We have determined the prevalence of four *VDR* gene polymorphisms and measured serum vitamin D levels in Kuwaiti children with T1DM and controls to evaluate their relationship with the disease onset. Methods: This study included 210 Kuwaiti children with T1DM and 136 ethnically matched controls. The ISPAD criteria was used for diagnosis of T1DM. The control subjects were healthy and were evaluated by a Diabetes specialist. The genotypes of four *VDR* gene polymorphisms (*Bsm*I, *Fok*I, *Apa*I and *Taq*I) were identified by PCR-RFLP. Serum vitamin D levels were determined by enzyme immunoassay method. The study was approved by KIMS-HSC Ethics Committee. Results: The genotype frequency of all four *VDR* gene polymorphisms was found to be significantly different between T1DM patients and controls. The genotype frequencies of all four polymorphisms were correlated with serum vitamin D levels in T1DM patients and controls. Serum vitamin D levels were significantly lower in T1DM patients compared to controls in the case of all genotypes except for bb genotype (minor allele) of the *VDR* gene *Bsm*I polymorphism. Conclusions: Our data demonstrate that four *VDR* gene polymorphisms (*Bsm*I, *Fok*I, *Apa*I and *Taq*I) show strong association with T1DM onset and with lower serum vitamin D levels in T1DM patients. Therefore *VDR* gene polymorphisms can be considered as important determinants of genetic predisposition of T1DM in Kuwaiti children.

1193T

Overall and central obesity with insulin sensitivity and secretion in a Han Chinese population: A Mendelian randomization analysis. *C. Hu.* Shanghai Diabetes Institute, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, 600 Yishan Road, Shanghai, 200233, China.

Objective: We aimed to characterize the causal role of obesity in insulin sensitivity and secretion by using Mendelian randomization (MR) approach with genetic variants for measuring overall and central obesity. Research design and methods: We constructed two genetic risk scores (GRSs) based on 38 established loci for body mass index (BMI; a surrogate of overall obesity) and 13 waist-to-hip ratio (WHR; a surrogate of central obesity) to assess the causal effects of BMI and WHR on several glycemic-related traits in 2884 community-based Han Chinese individuals. Results: Both of BMI and WHR were observationally correlated with the glycemic-related traits, including insulin sensitivity and secretion indices. We demonstrated these associations were partly causal in MR analysis. A genetically determined one standard deviation (SD, 3.35 kg/m^2) higher BMI caused a unit of 178.18 pmol/L higher Stumvoll 1st phase and 35.52 pmol/L higher Stumvoll 2nd phase insulin secretion (P range from 0.001 to 0.002), which were even independent of central obesity (P range from 0.019 to 0.039). In contrast, a genetically determined 1 SD higher WHR (the change of 0.002 in WHR) caused a unit of 1.21 higher HOMA-IR and 18.40 lower Gutt index (representing the insulin sensitivity) (P=0.048 and 0.028, respectively). Conclusions: We provided the causal evidence that overall obesity affects to some extent compensatory insulin secretion but central obesity inversely links to insulin sensitivity in Han Chinese population.

1194F

Clinical whole exome sequencing in early onset diabetes patients. *S. Kwak[1], C. Jung[1], C. Ahn[1,2], J. Park[3], J. Chae[4], H. Jung[1], Y. Cho[1,2], D. Lee[5], J. Kim[4], K. Park[1,2,6].* 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea; 3) New Business Division, SK Telecom, Seongnam, Gyeonggi-Do, Korea; 4) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; 5) Department of Internal Medicine, Wonkwang University School of Medicine & Hospital, Iksan, Korea; 6) Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea.

AIMS There could be an overlap of monogenic diabetes and early onset type 2 diabetes in those who are diagnosed before age of 30 years. Genetic diagnosis in these patients might improve quality of care. There are limited experience of using whole exome sequencing (WES) in Asian early onset diabetes patients and the clinical utility of WES is largely unknown. METHODS We performed whole exome capture and massive parallel sequencing in 28 early onset diabetes patients. Those who had strong family history of diabetes were preferentially enrolled. Rare and non-silent variants in 29 genes known to cause monogenic diabetes including 12 maturity-onset diabetes of the young (MODY) genes were investigated for its pathogenicity. RESULTS Average depth of on-target reads of WES was 97X. For the 29 genes of interest, 90% of sequences were covered for more than 20X. A total of four pathogenic or likely pathogenic rare missense variants (p.Leu319Pro of *HNF4A*, p.His103Tyr and p.Arg74Gln of *ABCC8*, and p.Leu139Val of *HNF1A*) in MODY genes were identified in three patients. Although four rare non-silent variants in MODY genes (p.Arg183Cys in *PAX4*, p.Val139Ile and p.Pro740fs in *CEL*, and p.Val147Ile in *HNF4A*) and two in monogenic diabetes genes (p.Glu169Lys in *WFS1*, and p.Pro407Gln in *GATA4*) were identified, their pathogenicity was uncertain. CONCLUSIONS WES could be an initial option for genetic testing in early onset diabetes patients. However, sufficient and universal coverage of genes of interest is required. In addition, there could be difficulties in interpreting variant pathogenicity and might require further validation.

1195W

Differential effects of genetic variants from across the allelic frequency spectrum at the 2q36.3-IRS1 locus on fasting insulin. *A. Leong[1], A.K. Manning[2], N. Nikpoor[3], CHARGE, T2D-GENES, GoT2D, and UK10K.* 1) Massachusetts General Hospital, Division of General Internal Medicine, 9th floor, Boston, MA, 02114, USA; 2) Massachusetts General Hospital, Richard B. Simches Research Center-7, 185 Cambridge St, Boston, MA, 02114, USA; 3) McGill University and Genome Quebec Innovation Centre, 740 Penfield, Room 6213, Montreal, Quebec, H3A 0G1, Canada.

Genome-wide association studies (GWAS) have identified a cluster of common single nucleotide variants (SNV) within a gene desert at 2q36.3 to be associated with fasting insulin (FI), type 2 diabetes, insulin resistance and cardiovascular disease. *IRS1*, residing ~500 kb away, has been implicated as the putative causal gene through long-range enhancer-promoter interactions. We hypothesize that rare and low-frequency SNVs at the 2q36.3-*IRS1* locus may have relatively large effects on FI. We sought to identify novel and potentially functional SNVs associated with FI after body mass index adjustment from across the allelic frequency spectrum at this locus. We examined sequencing data (5x-50x depth) at 2q36.3-*IRS1* from 8493 non-diabetic individuals in 7 studies, FHS, ARIC, CHS, RS, UK10K, GoT2D and T2D-GENES, and combined results in an inverse-variance weighted fixed effects meta-analysis. We then performed stepwise conditional analyses to identify additional distinct signals of SNVs with MAF≥1%. To examine the combined effects of SNVs, we piloted a haplotype analysis in FHS. We annotated SNVs using ENCyclopedia Of DNA Elements data to indicate regulatory sites. We used Madsen-Browning Burden tests and Sequence Kernel Association Tests to determine whether aggregated rare variants (MAF<1%) at regulatory sites were associated with FI. Of 331 SNVs with MAF≥1%, 30 reached $P<1.5 \times 10^{-4}$ (Bonferroni correction: 0.05/331); one was a low-frequency SNV (rs62188785; MAF=4.1%; $P=8 \times 10^{-5}$), 15kb upstream of the long-non coding RNA, LOC646736, near a putative enhancer, and 537kb away from *IRS1*. Conditioning on the most significant SNV (rs952227, MAF=32%, $P=4 \times 10^{-6}$), a known GWAS signal for FI, revealed that rs62188785 was a distinct signal (conditional $P=9 \times 10^{-4}$) whose minor allele had an effect in the opposite direction to the minor allele of rs952227. A generalized linear model regression analysis of FI on haplotypes showed that the haplotype containing the major allele at rs952227 and the minor allele at rs62188785 was associated with higher FI (HF=3.3%; $P=7 \times 10^{-3}$) compared to the most common haplotype (global P=0.03; N=842). Results from tests that aggregated rare variants at transcription factor binding sites, LOC646736, enhancers and *IRS1* were not statistically significant. In sum, through sequencing of 2q36.3-*IRS1*, we discovered a low-frequency SNV independent of the GWAS index SNP. Investigating the functional consequence of this SNV may identify a role for *LOC646736* on FI.

1196T

Genome-wide association study identifies new type 2 diabetes risk loci in Jordan subpopulations. *J. Li[1], R. Dajani[2], Z. Wei[3], Y. Khader[4], N. Hakooz[5], R. Fatahallah[6], M. El-Khateeb[6], A. Arafat[6], T. Saleh[2], A. Dajani[2], Z. Al-Abbadi[2], M. Abdul Qader[2], A. Shiyab[7], H. Hakonarson[1,8,9].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Biology and Biotechnology, Hashemite University, Zarqa, Jordan; 3) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 4) Department of Community Medicine, Public Health and Family Medicine, Faculty of Medicine, Jordan University for Science and Technology, Irbid, Jordan; 5) Department of Biopharmaceutics and Clinical Pharmacy Faculty of Pharmacy-University of Jordan, Amman, Jordan; 6) National Center for Diabetes, Endocrinology and Genetics, Amman, Jordan; 7) Department of Anthropology, Yarmouk University, Irbid, Jordan; 8) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 9) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Diabetes is among the most common chronic diseases globally and the prevalence is increasing. In Jordan, T2D affects 16% of the adults and another 23.8% with pre-diabetes. Type 2 diabetes (T2D) accounts for 90% of the diabetes cases in adults, causing a significant public health burden. Previous genome-wide association studies (GWAS) have identified over 90 susceptibility loci, but none of these studies have specifically investigated the Jodan populations. We genotyped genomic DNA of T2D cases and controls from Jordan, using the InfiniumII OMNI-Express BeadChip. After quality control filtering, 33 cases and 105 controls in the Circassian subgroup and 34 cases and 109 controls in the Chechen subgroup were left for analysis. We conducted GWAS on these two subpopulations in Jordan separately and further performed meta-analysis. We identified a novel T2D locus at chr20p12.2 at genome-wide significance (rs6134031, $P= 1.12 \times 10^{-8}$). It is located at an intergenic region, near gene *JAG1*. Another locus at chr12q24.31 (rs4758690, $P= 4.20 \times 10^{-5}$) was associated with T2D at suggestive significance level. Interestingly, the locus has a significant eQTL for gene *MLXIP* ($P=1.1 \times 10^{-14}$ for MLXIP expression in transverse colon; 4.2×10^{-7} for MLXIP expression in small intestine-terminal Ileum) and this SNP is also significantly associated with methylation level in MLXIP ($P=3.07 \times 10^{-5}$). MLXIP shows highest expression in colon and functions in transcriptional regulation of cellular glucose response, which makes it an interesting candidate target gene. Taken together, in the first GWAS of T2D cases constituting Jordan subpopulations, we identified novel susceptibility loci which may be unique risk factors in populations of the middle-east.

1197F

Integration of multi-tissue transcriptomic and genetic data in type 2 diabetes. *A. Ndungu[1], M. van de Bunt[1,2], A. Viñuela[3,4,5], S.C.J. Parker[6,7], N. Oskolkov[8], P.E MacDonald[9], H.K Im[10], M.L. Stitzel[11], InsPIRE Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ UK; 3) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Geneva, Switzerland; 5) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 6) Department of Human Genetics, Medical School, University of Michigan, Ann Arbor, MI, USA; 7) Department of Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, USA; 8) Department of Clinical Sciences, CRC, Lund University, Malmö, Sweden; 9) Alberta Diabetes Institute, University of Alberta, Edmonton, Canada; 10) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA; 11) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

C*is*-expression data are increasingly used to map non-coding GWAS signals to effector transcripts, usually through single variants (SV-eQTLs). Recently developed multi-variant (MV-eQTL) methods, such as PrediXcan integrate multiple *cis* variant effects and could prove more powerful genetic instruments e.g. for 2-sample Mendelian Randomisation experiments. We compared SV- and MV-eQTL methods in islet RNA-Seq data from 426 cadaveric donors. Initial results, comparing exon-level SV-eQTL and MV-eQTLs (PrediXcan), indicated MV-eQTLs generally accounted for more expression variation than the best SV-eQTL. This varied with gene; SV-eQTLs explained 0.13 of the variance in *DGKB* whilst MV-eQTLs accounted for 0.16. For *IGF2BP2,* the SV-eQTL accounted for 0.12 of phenotypic variance whilst MV-eQTLs accounted for 0.9. These differences could be a combination of genuine secondary signals and model over-fitting. Next, we compared SV-eQTL and MV-eQTLs in defining effector transcripts for T2D loci using DIAGRAM GWAS data. Of 6039 genes with a significant SV-eQTL (FDR <1%), 11 were highly correlated with T2D GWAS loci. Predicted expression from islet-derived MV-eQTL models identified 11 genes (10 at GWAS loci) significantly associated with T2D, 8 genes were implicated in T2D by both methods. MV-eQTLs identified putative effector transcripts at a further 3 genes with no SV-eQTL-GWAS overlap. We then compared T2D association results using predicted expression from islets to those from 36 tissues of broadly comparable sample sizes in GTEx v6. Compared to 11 significant signals (FDR<1%) from islets, muscle had 4 signals whilst other T2D-relevant tissues (liver, adipose, whole pancreas) had none: the most informative GTEx tissue was tibial artery (7 significant genes). Tissue-specificity varied; two genes in GWAS loci (*DGKB*, *UBE2E2*), had significant MV-eQTL signals evident from islet-generated models. In contrast, significant signals with *RCCD1* were evident for MV-eQTL analyses from 16 GTEx tissues and islets. *RCCD1* was significant in MV-eQTL analyses but not SV-eQTL. Whilst in principle, MV-eQTL based methods might be extensions of SV-eQTLs, complexities of these regions (multiple, overlapping signals) indicate that both have value in identifying candidate effector transcripts for further validation. The T2D MV-eQTL results reinforce the notion (consistent with physiological and genomic enrichment analyses) that events in the pancreatic islet are central to the pathogenesis of T2D. .

1198W

Identification of common drug targets for type 2 diabetes mellitus and cardiovascular diseases in Indian population using whole exome sequencing. *T. Prakash[1], A.M. Veerappa[1], N.B. Ramachandra[1], M.S. Basavaraj[2].* 1) Department of Studies in Genetics and Genomics, University of Mysore, Mysuru, Karnataka, India; 2) University Health Centre, University of Mysore, Mysuru, Karnataka, India.

The pathophysiology of the link between Type 2 Diabetes Mellitus (T2DM) and cardiovascular disease (CVD) is multifactorial. Diabetes is a prime risk factor for CVD and is also the most prevalent cause of morbidity and mortality among patients with T2DM. Understanding the profound mechanisms of these diseases can identify novel and effective drug targets. Current study was aimed at understanding the biological basis of the link between T2DM and CVD. We analyzed whole exome sequences of three individuals hailing from India and used custom variant identification pipeline, to identify and annotate pathological variants. We report the consequences of damaging novel and rare mutations in several genes for the shared pathophysiology that exists between T2DM and CVD. The findings indicate *PPARG*, *AKT2*, *APOB*, and *RELA* to be the prime drug targets in individuals with both T2DM and CVD pathology. The proposed mechanisms link hyperglycemia and intracellular metabolic changes of T2DM with oxidative stress, low-grade inflammation and endothelial dysfunction of CVD. The implications of the same will be discussed.

1199T

Genetics of diabetes in U.S. Hispanic/Latino individuals: Results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL).

Q. Qi[1], A.M. Stilp[2], T. Sofer[2], T. Wang[1], C.C. Laurie[2], M.C. Ng[3], I. Chen[4], K. Taylor[4], J. Pankow[5], M.L. Avilés-Santa[6], G. Papanicolaou[5], N. Schneiderman[7], J. Rotter[4], R.C. Kaplan[1], MEDIA consortium; HCHS/SOL. 1) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 5) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 6) Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD; 7) Department of Psychology, University of Miami, Miami, FL.

A number of type 2 diabetes (T2D) genetic loci have been identified through recent genome-wide association studies (GWAS) in populations of European, Asian, and African ancestry, while few GWAS have been conducted in US Hispanic/Latinos who are disproportionately affected by diabetes. We conducted a GWAS (>25M SNPs) in 2499 diabetes cases (defined by medication use, fasting glucose≥126 mg/dL, post-OGTT glucose≥200 mg/dL, or HbA1c≥6.5%) and 5247 normal controls (excluding pre-diabetes) from 6 Hispanic/Latino background groups (Cuban, Dominican, Puerto Rican, Mexican, Central American, and South American) in the HCHS/SOL. Goals were to 1) search for novel diabetes genetic loci; and 2) examine 76 known T2D loci in US Hispanic/Latinos. Our GWAS identified two known loci (*TCF7L2* and *KCNQ1)*, but no novel signals, that reached genome-wide significance ($P<5×10^{-8}$). Conditional analysis suggested an additional independent signal at *KCNQ1*, represented by the SNP rs1049549 (MAF=0.06; OR=1.49 [95% CI 1.08, 1.27]; $P= 2.9 ×10^{-7}$), which might be an African ancestry-specific allele (Africans: MAF=0.28; Europeans: MAF=0.01). This association was replicated in African Americans from the MEDIA consortium ($P=0.006$). To examine overall genetic predisposition to diabetes in US Hispanic/Latinos, we calculated a genetic risk score (GRS) based on 76 known index SNPs and found that the GRS had a highly significant association with diabetes (OR=1.07 [1.06, 1.09]; $P=1.2×10^{-39}$), with relatively stronger effect in non-obese compared with obese (OR=1.10 [1.08, 1.12] vs 1.06 [1.05-1.08]; P for interaction =0.005). We also studied the generalization of 76 known index SNPs to the Hispanic/Latinos. Among 76 index SNPs, 66 SNPs showed consistency with the previously reported direction of associations (binomial test $P=7.8×10^{-75}$); and 14 SNPs were significantly generalized to the HCHS/SOL using the directional FDR r-values framework. In addition, a GRS based on the SNPs that did not pass generalization testing also showed a highly significant association with diabetes, suggesting that some of the non-generalized SNPs are in fact associated with diabetes in US Hispanics/Latinos and non-generalization might be due to lack of power. In summary, our study identified a novel independent signal suggesting an African ancestry-specific allele at *KCNQ1* locus for diabetes. Associations between known genetic loci and diabetes were shown in a large cohort of US Hispanic/Latinos of diverse background.

1200F

A novel 'metabolic syndrome' genetic risk score characterizes high-risk individuals independent of body mass index in the Diabetes Prevention Program.

S. Srinivasan[1,8], K. Jablonski[2], W. Knowler[3], S. Dagogo-Jack[4], S. Kahn[5], E. Boyko[5], G. Bray[6], E. Horton[12], M.F. Hivert[1,7,8], R. Goldberg[9], L. Chen[1], J. Mercarder[10], M. Harden[11], J.C. Florez[1,8,11], The Diabetes Prevention Program. 1) Massachusetts General Hospital , Boston, MA; 2) George Washington University, Washington, DC; 3) National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ; 4) The University of Tennessee Health Science Center, Memphis, TN; 5) University of Washington, Seattle, WA; 6) Pennington Biomedical Research Center, Baton Rouge, LA; 7) Harvard Pilgrim Healthcare Institute, Boston, MA; 8) Harvard Medical School, Boston, MA; 9) University of Miami Leonard M.Miller School of Medicine, Miami, FL; 10) Barcelona Supercomputing Center, Barcelona, Spain; 11) Broad Institute, Cambridge, MA; 12) Joslin Diabetes Center, Boston, MA.

There is considerable heterogeneity in insulin resistance (IR), and phenotypic features of monogenic diseases can be used to understand potential mechanisms by which common variants act to induce IR. In the Diabetes Prevention Program (DPP), we tested a validated genetic risk score (GRS) consisting of 11 common genetic variants associated with a metabolic profile consistent with lipodystrophy including high triglycerides (TG), lower HDL cholesterol, and hepatic steatosis despite lower body mass index (BMI). We tested this GRS in 2,713 DPP participants who were at high risk for diabetes and had been randomized to intensive lifestyle intervention, metformin, or placebo. All measures were adjusted for age, sex, self reported ethnicity and baseline variable when applicable. At baseline and consistent with prior results, a higher GRS was associated with higher ln fasting insulin (µU/ml) (per allele β=0.012 [SE=0.005], $P=0.01$), higher ln TG (mg/dl) (β=0.02 [SE=0.005], $P<0.001$) and higher ln alanine transaminase levels (U/l) (β=0.009 [SE=0.008], $P<0.001$). In contrast and as expected, a higher GRS was associated with lower weight (kg) (β=-1.077 [SE=0.176], $P<0.001$), lower BMI (kg/m²) (β=-0.333 [SE=0.059], $P< 0.001$) and lower waist circumference (cm) (β=-0.659 [SE=0.128], $P<0.001$). Despite starting with lower weight and BMI at baseline, a higher GRS was associated with less reduction in weight (β=0.125 [SE=0.053], $P=0.018$) and BMI (β=0.047 [SE=0.019], $P=0.01$) at 1 year without significant interactions with treatment. Lifestyle intervention and metformin appeared to raise the insulin sensitivity index independent of genetic risk. In secondary analyses, we tested the association of the GRS with cardiovascular disease (CVD) endophenotypes in 2,708 participants. A higher GRS was associated with lower ln fibrinogen (mg/dl) (β=-0.006 [SE=0.002], $P=0.003$), lower ln C-reactive-protein (mg/dl) (β=-0.037 [SE=0.01], $P<0.001$), lower ln LDL peak-particle-density (Rf) (β=-0.005 [SE=0.001], $P<0.001$) and a lower ln ACC/AHA 10-year CVD risk score (β=0.04[SE=0.012], $P<0.001$) at baseline, reflecting lower adiposity. We found no association with coronary artery calcification scores. Among individuals with pre-diabetes, a higher GRS for IR with lower BMI was associated with lower CVD risk, in contrast to previous work showing increased coronary artery disease risk. Our results indicate that this GRS may have a differential effect on CVD in people with high metabolic risk at baseline.

1201W

Associations between *SLC16A11* variants and diabetes in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *A.M. Stilp[1], B. Hidalgo[2], T. Sofer[1], Q. Qi[3], N. Schneiderman[4], Y.-D.I. Chen[5], R. Kaplan[3], L. Aviles-Santa[6], K.E. North[7], D. Arnett[8], A. Szpiro[1], J. Cai[9], K. Taylor[5], G. Papanicolaou[6], C.C. Laurie[1], J.I. Rotter[5].* 1) University of Washington, Seattle, Department of Biostatistics; 2) University of Alabama at Birmingham; 3) Albert Einstein College of Medicine, Department of Epidemiology and Population Health; 4) University of Miami, Department of Psychological and Behavioral Medicine Research Center; 5) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics at Harbor-UCLA Medical Center; 6) National Institutes of Health, National Heart, Lung, and Blood Institute; 7) University of Chapel Hill, Department of Epidemiology; 8) University of Kentucky, College of Public Health; 9) University of North Carolina, Chapel Hill, Department of Biostatistics.

Five sequence variants in *SLC16A11* (rs117767867, rs13342692, rs13342232, rs75418188, and rs75493593), which occur in two non-reference haplotypes, were recently shown to be associated with type-2 diabetes in Mexicans and Mexican Americans analyzed by the SIGMA consortium (Williams et al. 2014). We analyzed these variants in 2492 diabetes cases and 5236 controls from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), which includes U.S. participants from six diverse background groups: Mexican, Central American, and South American (Mainland) and Puerto Rican, Cuban, and Dominican (Caribbean). We aimed to determine whether previous findings in Mexicans would replicate in the HCHS/SOL Mexican group and whether genotypic effects were similar in the other HCHS/SOL groups. The risk alleles occur in two non-reference haplotypes in HCHS/SOL, as in the SIGMA Mexicans. The haplotypes frequencies are very similar between SIGMA Mexicans and the HCHS/SOL Mainland groups, but are different in the Caribbean background groups. We found that *SLC16A11* sequence variants were significantly associated with risk of diabetes in the Mexican group (p<0.025), replicating the SIGMA findings. However, these variants were not significantly associated with diabetes in a combined analysis of all groups or in any of the five non-Mexican groups analyzed separately even though the risk variants were present at some frequency in all groups. Thus, the lack of replication in the combined analysis was likely not due to low power (85 % power to detect the previously-reported odds ratio). A sensitivity analysis in which younger controls were excluded (to better match the SIGMA studies) did not show qualitatively different results from the primary analysis. Recent results by Traurig et al. (2016) indicated that one of the SNPs, rs75493593, had a significant interaction with obesity in a sample of Native North Americans. We sought to explain the lack of replication by testing for a lower mean BMI in the Mexican group and for a similar interaction with obesity, but neither of these conditions was observed in the HCHS/SOL data. Therefore, the lack of replication is likely due to other factors, such as interaction with other Mexican-specific genetic variants or differing LD patterns in Mexicans compared to other Hispanic/Latino groups. These findings suggest the importance of recognizing the diversity of the Hispanic/Latino population in genetic analyses.

1202T

Genetic variants in *KCNB1* are modestly associated with type 2 diabetes in American Indians from the Gila River Indian Community. *M. Traurig, V. Ossowski, P. Chen, P. Piaggi, S. Kobes, R. Hanson, C. Bogardus, L. Baier.* NIH, Phoenix, AZ.

Pima Indians from the Gila River Indian Community (GRIC) have a high prevalence of type 2 diabetes (T2D) and we recently completed a genome-wide association study to identify determinants of this disease in 7701 community members using a custom designed Pima Indian specific Axiom Array. Based on whole-genome sequencing data on 266 full-heritage Pima Indians, we calculate that this array successfully tags (r^2 >0.85) 92% of all variation with a minor allele frequency (mAF) ≥0.05. Imputation was subsequently performed to capture remaining variations, and genotypes of novel variants in biological candidate genes with associations with T2D were prioritized for validation. One of these imputed variants was an intronic 16 bp deletion (mAF = 0.02) in the *KCNB1* gene (*KCNB1[16bpΔ]*). *KCNB1* encodes the potassium voltage-gated channel, Shab-related subfamily, member 1 which is involved in regulating insulin secretion. Genotyping of the deletion via fragment analysis in 7,351 Pima Indians confirmed the nominal association with T2D, where individuals with the deletion were protected from diabetes (OR = 0.51, 95%CI [0.34–0.77], $P = 1.2 \times 10^{-3}$ after genomic control, adjusted for age, sex, birth-year, and principal components 1-5). Analysis of other variation in this gene identified a common variant, rs3827088, (mAF = 0.31) directly genotyped by the Axiom array which also had a nominal association with T2D ($N = 7444$, OR = 1.17, 95%CI [1.05–1.30], $P = 4.9 \times 10^{-3}$). Genotyping of rs3827088 in an independent group of 3004 American Indians replicated the association with T2D (combined analysis; $N = 10448$, OR = 1.18, 95%CI [1.08–1.29], $P = 2.5 \times 10^{-4}$). Conditional analysis for the GRIC sample suggests that *KCNB1[16bpΔ]* and rs3827088 have independent effects on T2D; after conditioning for the other variant, the associations with T2D were essentially unchanged (*KCNB1[16bpΔ]*; OR = 0.55, 95%CI [0.39–0.77], $P = 3.2 \times 10^{-3}$ and rs3827088; OR = 1.18, 95%CI [1.08–1.30], $P = 2.8 \times 10^{-3}$). As expected, combining these 2 genetic variants showed that carriers of *KCNB1[16bpΔ]* and the non-risk (G) allele for rs3827088 are the most protected from diabetes, while those without the deletion and carrying the rs3827088 risk (A) allele are at the highest risk for diabetes. Studies of pathways potentially affected by *KCNB1* are ongoing.

1203F

Allelic variants associated with type 2 diabetes mellitus and related disorders in a Romanian cohort of elder male individuals. *G.F. Ursu[5], L.C. Bohiltea[1], P. Iordache[3], N. Cucu[4], V.E. Radoi[1,2], I. Craciunesc[5], C. Stiru[5], G. Chelu[2], C. Arsene[2], A. Voinoiu[3], V. Calota[3], C. Staicu[3], R.E. Bohiltea[3], C. Strugaru[1], D. Mates[3], A. Manolescu[6], V. Jinga[3], R.I. Ursu[1,2].* 1) Medical Genetics Department, "Carol Davila" University of Medicine and Pharmacy, Faculty of General Medicine, Bucharest, Romania; 2) SYNEVO Romania, Central Laboratory, Medical Genetics Department, Bucharest, Romania; 3) "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania; 4) Genetics Departmenr, Faculty of Biology, University of Bucharest, Romania; 5) "Prof. Dr. Agrippa Ionescu" Emergency Military Hospital, Bucharest, Romania; 6) University of Reykjavik, School of Science and Engineering, Reykjavik, Iceland.

   Background. Diabetes mellitus type 2 (t2D) is one of the world`s pandemics. This research wishes to determin high risk polymorphisms associated with t2D for the Romanian population. It is the first large diabetes-related genome-wide association study in Romania. It is a follow-up study of a previous PhD research involving the correlation between the FTO rs9939609 and ADRB3 rs4994 variants and obesity and diabetes in a cohort of 70 obese individuals. **Materials. Methods**. A GWAS was performed on a Romanian cohort of 750 males aged over 50 y.o. which have been divided into 2 study groups: 1. T2D group (250 individuals) 2. non-diabetic group (controls – 500 individuals) The research included 700000 SNPs. Genotyping results were analyzed in correlation with the presence of t2D and t2D-associated clinical parameters and disorders (weight, BMI, obesity level, flebitis, hypertension, prostate cancer). Lifestyle risk-factors were considered (environment, smoking, alcohol, coffee). Allele frequencies were calculated and compared with literature. **Results.** The results revealed a great number of allelic variants statistically corelated with t2D in the studied cohort. A cluster of SNVs (single nucleotide variants) on chromosome 6 (6 SNVs, 6q21, BVES gene) showed the highest association (p-values of 10^{-6}). Several variants on chromosomes 12, 13, 16, 18, 2, 4, 11, 19, 1, including SNVs in the previously reported FTO, MC4R, POMC, TMEM18, LEPR, BDNF, ADIPOQ, INSR, GHRH genes, also showed important correlations with t2D. Among the analyzed t2D parameters and comorbidities, obesity (including weight and BMI) and hypertension revealed the strongest association, SNVs on chromosomes 16, 6, 12, 13, 11, 4, 1 being correlated with these conditions. Over 50 SNPs were correlated with prostate cancer, 28 reavealing the strongest results (chromosomes 4,8,13,16,2,6). **Conclusions.** The study identified several gene clusters correlated with t2D and comorbidities. The implications of these variants for the Romanian population need to be further analyzed, replication studies having to be undertaken for confirmation (follow-up study under way - EU FP7 ROMCAN project, 5000 Romanian individuals). **Acknowledgemets.** The research is part of EU FP7 ProMark project.

1204W

Gene enrichment evaluation of glucocorticoid pathways for a role in type 2 diabetes pathogenesis. *L.N. Brenner, L. Chen, J.C. Florez.* Massachusetts General Hospital, Boston, MA.

   Glucocorticoids are used to treat many types of diseases. Notwithstanding their many therapeutic uses, they often cause hyperglycemia or worsening of pre-existing diabetes: up to 50% of patients who receive glucocorticoids have glucocorticoid-induced hyperglycemia. The mechanism of this intolerance is thought to be due to decreased insulin sensitivity, as well as decreased insulin secretion. However, the factors that predispose certain individuals to glucocorticoid-induced hyperglycemia are not known. To further evaluate whether genes in glucocorticoid pathways contribute to type 2 diabetes pathogenesis, we conducted a gene set enrichment analysis of different cohort results using the program Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA) deployed in publicly available genome-wide association meta-analyses for type 2 diabetes (DIAGRAM) and quantitative glycemic traits (MAGIC). We constructed three gene sets: one from an Online Mendelian Inheritance in Man (OMIM) literature search for "glucocorticoid," one from a published set of glucocorticoid-related expression changes and one including all genes in the glucocorticoid biosynthesis pathway. The different glucocorticoid-related gene sets show enrichment in both the type 2 diabetes and glycemic trait analyses. Using a stringent 95% cutoff, the gene sets derived from the OMIM literature-search and the glucocorticoid biosynthesis pathway show an enrichment for type 2 diabetes genetic associations (p=0.02 for both). The glucocorticoid synthesis gene set is also enriched for genetic associations with fasting insulin adjusted for BMI (p=0.02). The glucocorticoid-related expression gene set demonstrates enrichment for genetic associations of fasting glucose (p=0.03). A less stringent cutoff of 75% reveals additional significant enrichment for associations for the OMIM gene set with fasting insulin, with (p=0.04) and without (p=0.01) BMI adjustment, and with the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR; p=0.009). At this lower 75% threshold, the published expression gene set also shows enrichment for the corrected insulin response (p=0.03). Overall, this suggests that genetic variation in the glucocorticoid pathway contributes to hyperglycemia and type 2 diabetes. Further investigation is needed to identify the specific genes involved and to develop better prediction algorithms to distinguish patients who will and will not develop glucocorticoid-induced hyperglycemia.

1205T

Exome sequencing in DiscovEHR identifies rare variants in anion transporter genes that exert large effects on uric acid levels and gout. *J. Freudenberg, S. Mukherjee, C. O' Dushlaine, A. Li, GHS RGC DiscovEHR, J. Overton, J. Reid, L. Mitnaul, I. Borecki, A. Baras, O. Gottesman, F. Dewey, A. Shuldiner.* Regeneron Genetics Center, Tarrytown, NY.

Genome-wide association studies of uric acid level and gout have found associations with common variants kidney expressed anion transporter genes (*SLC22A12, SLC2A9, ABCG2*). *SLC22A12* and *SLC2A9* function in uric acid re-absorption, whereas *ABCG2* functions in uric acid excretion. We comprehensively evaluated the role of rare protein altering variants in these genes by exome sequencing of 3357 cases of gout and 35,721 controls without a history of gout, including 8677 individuals with measurements of serum uric acid. All individuals were of European ancestry and from the Geisinger Health System DiscovEHR cohort. Common variant associations were replicated at the anion transporter loci. Nonsynonymous variants were aggregated for gene-based analyses based on the prediction to be damaging by five bioinformatics methods (nsStrict), or alternatively one out of the five methods combined with minor allele frequency less than 1% (nsBroad). A strong association of rare variant burden was observed for *SLC22A12* with lower uric acid level (nsStrict: Beta=-0.26, P=1e-10; nsBroad: Beta=-0.21, P=1.7e-21) and decreased risk of gout (nsStrict: OR=0.27, P=6e-4; nsBroad: OR=0.39, P=8.2e-7). Similarly, rare variant burden in *SLC2A9* was associated with lower with uric acid levels (nsStrict: Beta=-0.19, P=5.5e-4; nsBroad: Beta=-0.12, P=1.6e-4) and decreased gout risk (nsStrict: OR=0.32, P=4.7e-2; nsBroad: OR=0.65, P=1.1e-1). In contrast, rare variant burden at *ABCG2* was associated with higher uric acid levels (nsStrict: Beta=0.075, P=4.5e-3; nsBroad: Beta=0.042, P=7.8e-3) and increased risk of gout (nsStrict:OR=1.58, P=3.3e-3; nsBroad: OR=1.32, P=4.6e-3). In summary, in addition to known common variant associations, we find strong rare variant effects at uric acid transporter genes. We hypothesize that drugs that block uric acid re-absorption might work best among individuals where uric acid excretion is impaired due to damaging variants in *ABCG2*, whereas such drugs may work poorly among individuals where re-absorption is already impaired due to damaging variants in *SLC22A12* or *SLC2A9*. Further, these findings suggest that given the strong effects of the multiple rare variants in these genes, drugs interfering with uric acid excretion, such diuretics or salicylates, may be used more carefully in those individuals.

1206F

Family-based linkage and association analysis of whole exome sequencing data with cardiometabolic phenotypes in Hispanic Americans: The Insulin Resistance Atherosclerosis Family Study. *N.D. Palmer[1,2,3,4,5], K.L. Tabb[1,2,3], J.N. Hellwege[6,7], L. Dimitrov[2], S. Sajuthi[4,5,8], K.D. Taylor[9], M.C.Y. Ng[2,3], G.A. Hawkins[2,4], Y.I. Chen[9], W.M. Brown[4,8], D. McWilliams[4,8], A. Williams[4,8], C. Lorenzo[10], J.M. Norris[11], J. Long[6,7], J.I. Rotter[9], J.E. Curran[12], J. Blangero[12], L.E. Wagenknecht[4,13], C.D. Langefeld[4,8], D.W. Bowden[1,2,3].* 1) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 5) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 6) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 7) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 8) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 9) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 10) Department of Medicine, University of Texas Health Science Center, San Antonio, TX; 11) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO; 12) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX; 13) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC.

Linkage methods can be a powerful tool to elucidate low frequency/high impact variants in extended families, particularly when used to compliment conventional association analysis. We utilized two-point linkage and single variant association analysis using whole exome sequencing (WES) data to distinguish variants which are strongly linked and/or associated with cardiometabolic traits. WES of 1205 Hispanic Americans (78 families) from the Insulin Resistance Atherosclerosis Family Study identified 555,651 variants. Linkage among variants with a minor allele frequency greater than 0.5% (n=211,612) with 50 metabolic traits yielded 10,580,600 LOD scores with 1148 LOD scores ≥3, 183 LOD scores ≥4, and 29 LOD scores ≥5. The highest LOD score was 17.65 at rs200573126 with serum adiponectin levels. This missense variant (G45R) in *ADIPOQ* was also significantly associated with adiponectin levels (*p*=4.53×10[-41]) as previously reported. The maximal novel LOD score detected was 5.50 for rs2289043 in *UNC5C* with subcutaneous adipose tissue measurements. *UNC5C* encodes a netrin-1 receptor; expression of netrin-1 has been reported to be higher in obese, but not lean, adipose tissue. Additional strong linkage signals included: *GAS2L3* with gamma-glutamyl transpeptidase levels (LOD=5.33), *AQP12B* with adiponectin levels (LOD=5.30), and *GPX6* and *FAM109A* with waist/hip ratio adjusted for BMI (LOD=5.13 and 5.02, respectively). Thirteen variants attained genome-wide significance, with the strongest association between rs651821, a 5′-UTR variant in *APOA5*, and triglyceride levels (*p*=3.67×10[-10]). APO AV is known to be involved in the regulation of plasma triglycerides. Other genes containing variants significantly associated with cardiometabolic phenotypes included *PNPLA3* with liver density (*p*=2.17×10[-09]), *CCDC115* with glucose effectiveness (*p*=4.25×10[-09]), and *IDH1* with waist circumference (*p*=2.45×10[-08]). Two genes contained variants showing coincident linkage and association: rs189547099 in *FNIP2* and a novel locus in *GLRB* with triglyceride levels (*p*=6.31×10[-08], LOD=3.13; both variants). *FNIP2* suppresses *PPARGC1a* RNA and is expressed in liver, pancreas, and adipose tissue; *GLRB* encodes the b unit of the glycine receptor. Overall, WES data provided insights that were not observed in exome chip data analysis and thus enabled identification of novel signals contributing to cardiometabolic traits when integrating results from two-point linkage and single-variant association analyses.

1207W

Genome-wide association studies identify two novel susceptibility loci to diabetic retinopathy in Japanese patients with type 2 diabetes. *M. Imamura[1,2,3], M. Taira[3], A. Takahashi[4,5], Y. Kamatani[4], M. Kubo[6], S. Maeda[1,2,3].* 1) Department of Advanced Genomic and Laboratory Medicine Graduate School of Medicine, University of the Ryukyus; 2) Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital; 3) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences; 4) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences; 5) National Cerebral and Cardiovascular Center, Omics Research Center; 6) RIKEN Center for Integrative Medical Sciences.

Several reports have shown familial aggregations of diabetic retinopathy (DR) or advanced DR among patients with type 1 or type 2 diabetes, suggesting genetic susceptibility contributes to the development and/or progression of DR. However, the genes involved in the susceptibility to DR are still unknown. To identify novel genetic loci associated with the susceptibility to DR, we performed a meta-analysis of genome-wide association studies (GWAS) for DR in Japanese patients with type 2 diabetes registered in BioBank Japan (Study-1; 4,839 DR cases and 4,041 controls, Study-2; 693 DR cases and 1,524 controls). DR cases were patients having any stages of DR and controls were type 2 diabetes patients who did not have any sign of DR with long duration of diabetes (≥5 years) or with diabetic nephropathy. We analyzed the association with DR of ~7.5 million single nucleotide polymorphisms (SNPs) from directly genotyped data (Study-1; Omni-express exome, Study-2; Illumina 610K) and genotype imputation using MACH and minimac. We have identified 85 loci showing suggestive association with DR ($p < 1 \times 10^{-4}$) through the GWAS meta-analysis. These loci were further evaluated in an independent case-control study (Study-3; 2,260 DR cases and 723 controls, BioBank Japan). After combining all the association data (Study-1, 2 and 3) by a meta-analysis, the association of two loci reached genome-wide significance level; rs12630354 on Ch3: $p = 1.62 \times 10^{-9}$, Odds ratio [OR] = 1.17, 95% Confidence Interval [CI] 1.11-1.23, rs140508424 on Ch9: $p = 4.19 \times 10^{-8}$, OR = 1.61, 95% CI 1.36-1.91. None of the association with advanced DR reached genome-wide significance ($p > 5.0 \times 10^{-8}$) in a sub-group analysis: excluding patients with simple DR (DR cases n = 2,003, 368 and 1,260 for Study-1, Study-2 and Study-3, respectively). In conclusion, we have identified two novel loci contributing to DR susceptibility, although further replication studies are required to validate the association of these loci with DR.

1208T

Genome-wide associations of objective activity-monitor derived measures of sleep patterns. *S.E. Jones[1], A.R. Wood[1], H. Yaghootkar[1], J. Tyrrell[1], M.A. Tuke[1], K.S. Ruth[1], R.N. Beaumont[1], R.M. Freathy[1], A. Murray[1], V.H. Stiles[2], M. Hillsdon[2], T.M. Frayling[1], M.N. Weedon[1].* 1) Genetics of Complex Traits, Medical School, University of Exeter, Exeter, Devon, United Kingdom; 2) Sport and Health Sciences, University of Exeter, Exeter, Devon, United Kingdom.

Sleep is a fundamental component of our daily lives. Poor quality sleep, lack of or excessive sleep and disrupted sleep timings are important risk factors for diseases such as obesity, type 2 diabetes and cardiovascular disease. Until recently, little was known about the genetic factors that influence sleep patterns. Using data from 23andMe and UK Biobank, we and others recently identified 22 variants influencing chronotype (whether you are a morning or evening person) and three loci influencing sleep duration. The identification of additional variants associated with sleep will provide further insight into the biology of this essential biological requirement. The use of objective measures of sleep may provide more power to identify variants than self-report measures of sleep, which may be subject to reporting bias.   We aimed to derive objective measures of sleep using data from activity monitors worn by ~103,000 of the 500,000 UK Biobank individuals for ten days. We aimed to identify genetic variants associated with sleep and validate variants recently identified by self-report measures. We present preliminary GWAS results of sleep duration and sleep midpoint using the ~22,000 white British UK Biobank individuals that have both actigraphy data and genetic data available.   Objective, actigraphy derived sleep measures were correlated with self-report measures: sleep duration r=0.18, chronotype (derived from midpoint of sleep) r=0.24.   Of the 22 variants associated with self-report chronotype, 21 had directionally consistent effects (binomial P=2.4x10-7) in objective actigraphy-based measures, and six replicate at P<0.05, including variants in the *PER2* and *AK5*/*PIGK* loci. Using a GRS (genetic risk score) of the 22 lead chronotype variants, we saw strong associations between sleep midpoint and chronotype (P=3.6x10-10, P=5.9x10-8 and P=8.6x10-9 for average over all nights, weekday nights and weekend nights respectively). The strongest self-report sleep duration signal in *PAX8* was reproduced in the actigraphy sleep duration GWAS (P=1.9x10-4) with almost identical effect size (2.3mins per allele for both).   In conclusion, we have performed the first large-scale GWAS study of objective measures of sleep patterns. We validate known self-report sleep duration and chronotype signals. With the upcoming release of the full UK Biobank cohort, we aim to identify further signals associated with these objective sleep pattern phenotypes.

1209F

Genome-wide association study of 1,5-anhydroglucitol in adults without diabetes: Results from the Atherosclerosis Risk in Communities Study.
M. Li[1,2], N. Maruthur[1,3,4], S. Loomis[1], K. North[5], H. Mei[6], A. Morrison[7], J. Pankow[8], E. Boerwinkle[7,9], E. Selvin[1,3,4], A. Kottgen[1,10]. 1) Department of Internal Medicine, University of Utah, Salt Lake City, UT; 2) Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 3) Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 4) The Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD; 5) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 6) Center of Bio-statistics & Bioinformatics, University of Mississippi Medical Center, Jackson, MS; 7) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 8) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 10) Division of Genetic Epidemiology, Institute for Medical Biometry and Statistics, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany.

1,5-anhydroglucitol (1,5-AG) is a marker of glucose peaks over 1-2 weeks and has been shown to predict diabetes-related outcomes. Lower levels of 1,5-AG are indicative of hyperglycemia. To gain additional insights into the genetic underpinnings of hyperglycemia, we aimed to identify genetic variants associated with 1,5-AG and to examine their relationship with the most commonly used clinical glycemic markers, HbA1c and fasting glucose. We conducted genome-wide association studies of 1,5-AG in 7,550 European ancestry and 2,030 African American participants without diagnosed diabetes from the Atherosclerosis Risk in Communities (ARIC) Study. Variants that were significantly associated with serum 1,5-AG concentrations were evaluated for replication in a study of blood metabolite concentrations among 7,824 European individuals. In the ARIC study, mean 1,5-AG levels were lower in African American than European ancestry participants (17.3 vs 18.9 ug/mL; $P<0.0001$). Genome-wide significant associations ($p<5 \times 10^{-8}$) among European ancestry participants were identified for single nucleotide polymorphisms (SNPs) in or near *EFNA1*, *MCM6*, *SI*, *MGAM/MGAM2*, *SLC5A10*, and *SLC5A1*. Among the smaller African American sample, no genome-wide significant associations were observed. The index SNPs at each of the six loci together explained 4.2% of the variation in 1,5-AG levels. A common theme among the genes mapping into the identified loci is their role in intestinal glucose digestion as well as glucose and 1,5-AG reabsorption in gut and kidney. None of the SNPs were significantly associated with traditional glycemic markers (fasting glucose and HbA1c) or other nontraditional glycemic markers (fructosamine and glycated albumin). Four of the six loci were successfully replicated in an external European population study. We have identified, for the first time, genes that influence 1,5-AG in non-diabetic adults suggesting genetic variation related to glucose reabsorption and metabolism as a strong influence on 1,5-AG levels and highlight mechanisms that influence the body's ability to metabolize glucose.

1210W

Novel locus discovery through trans-ethnic association analyses of glycemic traits using densely imputed genetic data. *G. Marenne on behalf of the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigator.* Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Previous large-scale glycemic trait genetic association analyses, by us (MAGIC) and others, have identified more than 120 loci associated with fasting glucose (FG), fasting insulin (FI), glycated hemoglobin (HbA1c) and 2-hour glucose (2hG), focusing on European descent populations. Here, we extended meta-analyses to five ancestries (71% Europeans, 13% East Asian, 7% Hispanic, 6% African-American and 3% South Asian), and used genetic data imputed to the 1000 Genomes Project (phase 1 v3, March 2012, or later) to aid additional locus discovery and fine-mapping. We report preliminary association results for FG, FI, HbA1c, and 2hG levels in up to 281,416 individuals without diabetes from 144 cohorts. Genetic association analyses were performed using an additive model, and inverse normal transformation for each trait. FG, FI and 2hG were further adjusted for body mass index. Individual cohort results were combined by fixed-effect ancestry-specific meta-analyses. Trans-ethnic meta-analyses were then performed using MANTRA which allows for heterogeneity in allelic effects between diverse ethnic groups, with a $\log_{10}$ Bayes factor ($\log_{10}BF$) threshold of 6 used to identify genome-wide significant signals. Lastly, GARFIELD was used to perform functional enrichment analysis. Preliminary results identified 132 regions associated with FG, of which 93 did not overlap previously established FG-associated regions. New FG loci include those at *CDK14* and *GAD2* genes ($\log_{10}BF=19.75$ and 16.44). We also identified 78 FI-associated regions (61 novel), 189 HbA1c-associated regions (138 novel) and 31 2hG-associated regions (23 novel). The most significant FI, HbA1c and 2hG novel signals fall within *BCL2*, *HBB* and *DVL2* genes ($\log_{10}BF$ = 11.56, 25.28 and 11.10, respectively). Enrichment analysis showed that HbA1c signals were mostly enriched in open chromatin seen in blood cell type, whereas FG and 2hG signals were mostly enriched in liver cell type, and FI signals were mostly enriched in fetal lung cell type. In conclusion, this large international effort has identified over a hundred novel loci, and suggested new hypotheses about the biology and the genetic architecture underlying glycemic traits.

1211T

Genome-wide association study of iron traits in relation to diabetes mellitus in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL): Potential genomic intersection of iron and glucose regulation. *L.M. Raffield[1], T. Louie[2], T. Sofer[2], D. Jain[2], E. Ipp[3], K.D. Taylor[4], G.J. Papanicolaou[5], L. Avilés-Santa[5], L.A. Lange[1], C.C. Laurie[2], M.P. Conomos[2], T.A. Thornton[2], Y.I. Chen[4], B. Thyagaran[6], A.P. Reiner[7], N. Schneiderman[8], H.J. Lin[4], J.I. Rotter[4].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biostatistics, University of Washington, Seattle, WA 98195; 3) Department of Medicine and Division of Endocrinology, Harbor-UCLA Medical Center, Torrance, CA 90502, and the David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA 90502, and the David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Division of Cardiovascular Sciences, NHLBI, NIH, Bethesda, MD 20892; 6) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455; 7) Department of Epidemiology, University of Washington, Seattle, WA 98195; 8) Department of Psychology and Behavioral Medicine, University of Miami, FL 33124.

Genetic variants are known to affect iron stores in the body, and both iron overload and deficiency can adversely affect human health. For example, elevated iron storage has been associated with increased diabetes risk, although exact mechanisms are still being investigated. We conducted the first genome-wide association study (GWAS) of ferritin, serum iron, total iron binding capacity (TIBC), and transferrin saturation in a Hispanic/Latino population, the Hispanic Community Health Study/Study of Latinos (HCHS/SOL; >12,000 participants with iron measures; mean age 46.1; 58.6% female). We identified variants associated with iron traits and then evaluated associations with type 2 (apparently >95% in HCHS/SOL) diabetes and related glycemic traits. Genotyping was done using an Illumina array (>2.5 million variants from the HumanOmni2.5-8v1-1 array and custom content), with imputation to the 1000 Genomes phase 1 version 1 reference panel. Associations were assessed using a mixed model approach (the GENESIS R package) adjusting for sex, age, cigarette use, heritage group, sampling design, kinship, household, census block, recruitment center, and five ancestry principal components. A novel association between TIBC and an intergenic variant near the gene for protein phosphatase 1, regulatory subunit 3B (*PPP1R3B*), was observed (rs4841132, β= -0.116, P = 7.44 x 10^{-8}, effect allele frequency=18.2%), which strengthened when iron deficient individuals were excluded (β= -0.121 P = 4.78 x 10^{-9}). Genome-wide significant associations (P<5 x 10^{-8}) were also identified between iron traits and four previously known loci -- transferrin (*TF*), hemochromatosis (*HFE*), fatty acid desaturase 2 (*FADS2*)/myelin regulatory factor (*MYRF*), and transmembrane protease, serine 6 (*TMPRSS6*). Eight of twelve European GWAS loci and two of four African American GWAS loci generalized to HCHS/SOL (P<0.05, after correction for false discovery rate). Regarding glucose homeostasis, the iron-raising variant of *TMPRSS6* (rs855791, Val736Ala) was associated with lower HbA1c (P = 8.66 x 10^{-10}), and the iron-raising allele of rs4841132 (near *PPP1R3B*) was associated with elevated fasting glucose (P = 7.70 x 10^{-7}) and elevated fasting insulin (P = 4.79 x 10^{-6}). Adjustment for TIBC did not attenuate these *PPP1R3B* associations, so iron is not likely the mediator. These results provide new genetic information on iron traits and on the connection between iron and glucose homeostasis.

1212F

Genome-wide association analysis identifies novel loci associated with circulating visfatin levels. *S.N. Adebamowo[1], A.A. Adeyemo[1], A.P. Doumatey[1], F. Tekola-Ayele[1], A.R. Bentley[1], G. Chen[1], H. Huang[1], J. Zhou[1], D. Shriner[1], O. Fasanmade[2], G. Okafor[3], B. Eghan[4], K. Agyenim-Boateng[4], A. Owusu[5], A. Amoah[5], J. Acheampong[4], T. Johnson[2], J. Oli[3], C.A. Adebamowo[6], F. Collins[7], G. Dunston[8], C.N. Rotimi[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) University of Lagos, Lagos, Nigeria; 3) University of Nigeria Teaching Hospital, Enugu, Nigeria; 4) University of Science and Technology, Kumasi, Ghana; 5) University of Ghana Medical School, Accra, Ghana; 6) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 7) National Institutes of Health, Bethesda, MD; 8) National Human Genome Center at Howard University.

Background: Visfatin is an adipokine produced by multiple tissues but primarily by brown adipose tissue. Animal experiments and epidemiological studies have suggested a role for visfatin in insulin secretion; type 2 diabetes (T2D) and obesity. However, the genetic factors influencing circulating visfatin levels have not yet been identified. We conducted a genome wide association study (GWAS) for circulating visfatin in individuals of African ancestry. **Methods:** The study participants were drawn from two independent studies, an African cohort (n=1382) enrolled in Nigeria, Ghana and Kenya and an African-American cohort (n=2109) enrolled from the Washington DC metropolitan area. Serum visfatin was measured on fasting samples using the BioRad Bio-Plex Pro™ Human Diabetes Immunoassay. Samples were genotyped on the Affymetrix Axiom® PanAFR SNP array and the Affymetrix Axiom® Exome 319 Array, followed by imputation into the 1000 Genomes Phase 1 v.3 cosmopolitan reference. Association tests were done for visfatin levels under an additive genetic model with adjustment for age, sex, body mass index, T2D status and the first 3 principal components of the genotypes. **Results:** In the African cohort, the GWAS yielded a significant cluster of 4 SNPs ~30kb upstream of the *CNTNAP2* gene on chromosome 7: rs112623225 (p = 2.6 x 10^{-9}), rs17480133 (p = 2.5 x 10^{-9}), rs4726781 (p = 5.1 x 10^{-9}) and rs4726782 (p = 4.2 x 10^{-8}). These variants alter several regulatory motifs in key transcriptional factors including *ARID5A*, which is overexpressed in adipocytes and whole blood, the two primary sites of visfatin abundance. In the African-American cohort as well as in a meta-analyses of both cohorts, there was a genome-wide significant variant (rs2238776, p = 1.9 x 10^{-12}) in the *TBX1* gene. This intronic variant is in enhancer histone marks in several tissues and alters the motifs Ets*, HNF1, TCF12*, most of which are expressed in adipocytes. However, this variant is less common in African populations (YRI MAF=0.004), compared to African-Americans (ASW MAF 0.04). This may explain, in part, why it was not observed among continental Africans included in this study. **Conclusions:** We identified novel genetic variants near *TBX1* and *CNTNAP2* that influence circulating visfatin levels. These findings should be investigated further in other global populations. Mechanistic studies examining how these loci regulate circulating visfatin levels should also be conducted.

1213W

From genetic informatics to a biological model: Analysis of genetic variants of *SLC5A. J.C. Fox[1], B.A. Hoch[2], J. Mayer[2], H. Ye[2], E.W. McPherson[1], S.J. Hebbring[1].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, MARSHFIELD, WI; 2) Bioinformatics Research Center, Marshfield Clinic Research Foundation, MARSHFIELD, WI.

SLC5A2 encodes the sodium/glucose co-transporter 2 (SGLT2). Primarily found in the kidneys, this protein plays a vital role in glucose transport and regulation of blood glucose levels. SGLT2 variants are implicated in a disorder known as glycosuria, which is characterized by an increased secretion of glucose in the urine. SGLT2 has become a popular target for type 2 diabetes (T2D) therapeutics. However, many individuals using these drugs develop side effects that include ketoacidosis, a condition that is not normally found in individuals with glycosuria. Using Phenome Wide Association Studies, we have identified 2 novel and likely clinically relevant variants within the gene *SLC5A2*. Herein, we present a study that begins with the identification of two biologically-relevant genetic variants and describe the experimental design to analyze their functional consequences. Additionally, we will test the impact of functional variants on commercially-available SGLT2 inhibitors. From these studies, we can begin to develop a pharmacogenetic model that may link SGLT2 inhibitors to variant-specific outcomes that may affect their use in patients with T2D. As precision medicine takes precedence in health care it is hoped that studies like these, which begin with a genetic informatics approach, will have a clinical impact on patient care.

1214T

Opposite effects of maternal and paternal alleles on body mass index in the Hutterites. *S. Mozaffari[1], D. Nicolae[2], C. Ober[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Departments of Medicine and Statistics, University of Chicago, Chicago, IL.

Genome Wide Association Studies (GWAS) typically treat alleles inherited from the mother and from the father as equivalent, although it is possible that variants can affect traits differently depending on whether they are maternal or paternal in origin. Many studies have utilized pedigree information to test for parent of origin effects (POEs) on phenotypes and gene expression to uncover new imprinted loci. However, none have investigated whether variants can have opposite effects depending on the parental origin. We developed a novel statistical model that compares the effects of the maternal and paternal variants on trait values, including effects that are opposite. We tested for POEs with 10 common quantitative traits in >1000 Hutterites, who are members of a single pedigree, using ~5M variants (MAF>5%) imputed and phased from Hutterite whole genome sequences. Although a standard GWAS of body mass index (BMI) in the Hutterites did not yield any significant associations, our new method revealed SNPs in 12 independent regions with POEs that surpassed genome wide significance ($p<10^{-9}$); all were due to opposite effects of paternally and maternally inherited alleles. Four regions include known or predicted imprinted genes: paternally expressed *FAM50B* and *GATA3* and maternally expressed *CTNNA3* and *ABCC9*. The two most significant associations were not in known imprinted regions. The lead SNP in one region (rs114757733 A/G, $P=7.17\times10^{-15}$) is in the intron of the *DIET1* gene, which modulates bile acid and lipid levels. At this SNP, the paternally inherited A allele is associated with increased BMI whereas the maternally inherited A allele is associated with decreased BMI. The lead SNP in the second region (rs113661800 G/del, $P=1.75\times10^{-14}$) is intergenic between a long noncoding RNA gene, *LINC0449*, and a gene involved in metabolic function, *CCRN4L*. LINC genes are often found in imprinted regions. The G allele at this SNP is associated with increased BMI when inherited from the father but decreased BMI when inherited from the mother. Five of the other nine phenotypes showed significant POEs, but none with opposite effect alleles. Thus, BMI is unique in this respect and fits with the theory that imprinted loci are primarily involved in metabolic functions related to growth. In summary, our new method revealed extensive opposite POE on BMI, potentially identifying new imprinted regions and potential risk alleles for metabolic disorders. Supported by HL085197. .

1215F

Joint effects of common genetic variants for type 2 diabetes on the risk of gestational diabetes. *V.K. Kawai[1], R.T. Levinson[2], A. Adefurin[1], D. Kurnik[1,3,4], S. Collier[5], D. Conway[5], C.M. Stein[1].* 1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 2) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, USA; 3) Clinical Pharmacology Unit, Rambam Medical Center, Haifa, Israel; 4) Rappaport Faculty of Medicine, Technion - Israel Institute of Technology, Haifa, Israel; 5) Vanderbilt Institute of Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA.

Gestational diabetes (GDM) is characterized by maternal glucose intolerance that occurs only during pregnancy. Because GDM resembles closely type 2 diabetes (T2DM) in its pathophysiology and risk factors, shared genetic predisposition is likely but has not been established. Here, we test the hypothesis that a genetic risk score (GRS) that included variants known to be associated with T2DM is associated with the risk of GDM. **Methods**: We conducted a case-control study using BioVU, a DNA biobank linked to de-identified electronic medical records at Vanderbilt Medical Center. To calculate a simple-count (0, 1, 2 for each risk allele) GRS, we selected 34 SNPs previously associated with T2DM or fasting glucose in two or more genome wide association studies ($P < 5 \times 10^{-8}$) in the general population. Using bioinformatic algorithms and manual review in women without diabetes, we identified 458 cases of GDM and 1537 pregnant controls of European ancestry. We defined a case of GDM as having a positive 100g oral glucose tolerance test during pregnancy using Carpenter and Coustan criteria, or a physician's diagnosis, or an ICD9 billing code for glucose intolerance in pregnancy with an intervention for GDM (nutritional or antidiabetic drug); controls were pregnant women with a normal 50g glucose challenge test. We performed a logistic regression analysis adjusting for maternal age, parity, and maternal body mass index (BMI) to assess the effect of the GRS on the risk of GDM and we calculated the area under the curve (AUC) for the receiver-operating characteristic curve (ROC) for the logistic regression model. **Results**: GDM cases had a higher number of risk alleles compared to controls (38.9 ± 4.0 risk vs. 37.4 ± 4.0 risk alleles, P<0.0001). The odds ratio (OR) for GDM associated with an increase of 1 risk allele in the GRS was 1.10 (95%CI 1.07 - 1.13) after adjustment for maternal age, parity, and BMI. Women in the top quintile of the GRS (≥ 42 risk alleles) had an OR=3.19 (95%CI 2.17 - 4.71) for GDM when compared with women in the bottom quintile (≤ 34 risk alleles). The AUC for the ROC curve for predicting GDM was 0.67 (95% CI 0.64 - 0.70) for clinical covariates alone and 0.70 (95%CI 0.67 - 0.73) for clinical covariates and the GRS. **Conclusions**: The main finding was that a GRS, which incorporated the effects of common genetic variants for T2DM, was associated with the risk of GDM.

1216W

Circulating aromatic amino acids modify the *GCKR* genotype on changes of glucose metabolism in response to weight-loss diets: The POUNDS Lost trial. *Y. Heianza[1], W. Ma[2], Y. Zheng[3], G.A. Bray[4], F.M. Sacks[3], L. Qi[1,2,3,5].* 1) Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA; 5) Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

The type 2 diabetes (T2D) associated glucokinase regulatory protein gene (*GCKR*) variant was recently found to be related to circulating levels of amino acids, including branched-chain amino acids (BCAAs) and aromatic amino acids (AAAs) that are predictors of T2D. We investigated whether blood BCAAs and AAAs levels modified associations of the *GCKR* variant on changes in glucose metabolism among overweight or obese individuals who participated in the POUNDS Lost trial, a 2-year weight-loss diet intervention study. This study included 691 participants with genotyping data on *GCKR* variant rs780094 and blood BCAAs (valine, leucine and isoleucine) and AAAs (tyrosine and phenylalanine). We found that plasma phenylalanine significantly modified association of the *GCKR* genotype with 6-month changes in fasting glucose ($P_{interaction}$=0.019) regardless of body mass index at baseline. Among individuals with high phenylalanine levels at baseline, increasing number of the diabetes-risk-reducing A allele was associated with increase in glucose concentration (β [SE], 1.56 [0.69]; *p*=0.02). Conversely, among those with low phenylalanine levels, carrying the A allele showed a significant reduction (β [SE], −1.46 [0.65]; *p*=0.03). Similar but less significant interaction patterns were also observed at 2 years. Blood BCAAs did not interact with *GCKR* for the changes in glucose concentrations. In conclusion, blood AAA modified associations of the *GCKR* variant on changes in glucose after the diet intervention. Our study suggests that low phenylalanine levels improve glucose metabolism in response to weight-loss diets for overweight or obese individuals with the diabetes-risk-reducing *GCKR* A allele.

1217T

Adiponectin isoform distribution in carriers of ethnic-specific *ADIPOQ* **mutations: The Insulin Resistance Atherosclerosis Family Study (IRASFS).** *K.L. Tabb[1,2,3], C. Gao[2,3,4,5], G.A. Hawkins[2,5], J.I. Rotter[6], Y.-D.I. Chen[6], J.M. Norris[7], C. Lorenzo[8], B.I. Freedman[2,3,5,9], D.W. Bowden[1,2,3], N.D. Palmer[1,2,3,4,5].* 1) Dept of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics & Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 5) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 6) Inst for Translational Genomics and Population Sciences/ Dept of Pediatrics, LA BioMed-UCLA Med Center, Torrance, CA; 7) Dept of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO; 8) Dept of Medicine, University of Texas Health Science Center, San Antonio, TX; 9) Dept of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC.

   Adiponectin, an adipokine synthesized and secreted specifically by adipocytes, is found in human serum in three distinct multimers (high, medium, and low molecular weight). The high molecular weight (HMW) multimer has been reported to be the most biologically active isoform. Our laboratory has previously reported two ethnic-specific variants in *ADIPOQ*, G45R in Hispanic Americans and R55C in African Americans. In our original report, carriers of both variants had mean adiponectin levels, measured by radioimmunoassay (RIA), that are approximately 19% of those of non-carriers. While adiponectin levels are observed to be inversely correlated with cardiometabolic disease, heterozygous mutation carriers were not clinically different from non-carriers. In this study, we used western blotting to explore the multimeric distribution of adiponectin in carriers of each variant to elucidate the lack of a clinical phenotype associated with cardiometabolic disease. Using a native PAGE protocol, serum proteins of archived human serum samples from the Insulin Resistance Atherosclerosis Family Study (Hispanic American, n=125; African American, n=47) were separated and transferred to a nitrocellulose membrane for antibody-based detection and densitometry analysis. Fluorescent imaging of the membranes showed an absence of the HMW isoform and a marked (>90%) reduction in the medium molecular weight isoform in serum samples from G45R and R55C carriers compared to those of non-carriers. There was a corresponding 20% increase (p<0.01) in the amount of the low molecular weight isoform (LMW) compared to non-carriers. This is in contrast to our original findings and is likely due to the preferential detection of the HMW by the antibody used in the RIA. These results challenge the current understanding of the importance of the HMW multimer, suggesting that the increase in the LMW isoform, and thus total serum adiponectin, may functionally compensate for the loss/reduction of the higher order multimers in carriers of the G45R and R55C *ADIPOQ* variants and could therefore explain the clinically normal phenotype of these individuals. These results have implications for efforts to use adiponectin therapeutically.

1218F

New insights into the control of obesigenic pathways - A comprehensive proteomics and genetics data integration approach. *J. Carayol[1], A. Valsesia[1], S. Métairon[2], P. Descombes[2], J. Hager[1].* 1) Nutrition and Metabolic Health, Nestlé Institute of Health Sciences SA, Lausanne, Vaud, Switzerland; 2) Functional Genomics, Nestlé Institute of Health Sciences SA, Lausanne, Vaud, Switzerland.

   Many genes and proteins influencing bodyweight, like leptin or adiponectin have been identified and characterized. While they have considerably enriched our understanding of the pathways involved in obesity, many aspects of how these genes/proteins are controlled on the molecular level remain unknown. Many of these proteins show large inter-individual variability and/or changes during weight loss interventions. It is not clear if these changes mainly stem from environmental influences or systemic biological differences. Availability of high-throughput omics technologies including genomics and proteomics should allow the characterization of the molecular mechanisms underlying obesity and their regulation. We performed the first large-scale proteomics screen combined with analysis of variation in the genome in one of the largest multi-centric weight loss intervention trials, the Diogenes study. Information from 1,129 proteins and 4 million SNPs was available for 498 participants before and after a low calorie diet (LCD) intervention. Association was tested first between BMI and proteins levels, both for baseline and fold changes during LCD intervention. For proteins associated with BMI variations before and during LCD, protein Quantitative Trait Locus (pQTL) analysis was performed using a linear mixed model. At baseline out of 146 proteins associated with BMI, 50 had pQTL signals (29 cis-acting and 21 trans-acting loci). In contrast among 127 proteins associated with BMI change during LCD, only trans-acting pQTLs (n=28) were identified, even at a relaxed suggestive association pvalue cutoff set to 1e-06. One of the strongest trans-acting pQTL signals for weight loss associated proteins was for leptin plasma levels. The SNPs for this pQTL are in a region on chromosome 6 (SNP rs1336257, p = 1.1e-7) that is enriched in transcription factor binding sites and close to the BCKDHB gene involved in branched chain aminoacid catabolism. We hypothesize that this or other genes in this region may be new regulators of leptin levels during weight loss. To our knowledge, we performed the first pQTL analysis in obesity and provide novel insight into the regulation of gene's protein product before and after weight-loss in subjects with overweight and obesity.

1219W

Association of *ENPP1* K121Q polymorphism with common obesity in North Indian Punjabi population. *H. Kaur[1], B. Doza[1,2], M. Singh[1], J. Kaur[1].* 1) Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Director, Aligarh Muslim University off-campus, Murshidabad, West Bengal.

Obesity is a complex phenotype resulting from the combined effects of genes, behavioral factors, and their interactions. A large number of genetic association studies suggest that polymorphic variants in genes are associated with BMI. Genetic variants of the *ENPP1* are reported to be associated with obesity in different ethnic populations. The aim of the present study was to examine the association of *ENPP1*(K121Q) polymorphism with obesity and obesity related phenotypes in Punjabi population. A total of 200 subjects, 100 obese (50 male/ 50 female) and 100 gender matched non obese were included from Amritsar region of Punjab. Various anthropometric and physiometric measurements were taken for all the subjects. Genotyping of *ENPP1*gene K121Q polymorphism was performed by the PCR-RFLP method, PCR products digested with restriction enzyme Ava Π, followed by Agarose Gel Electrophoresis. The genetic effects were analyzed according to gender (male and female). Deviation from Hardy- Weinberg Equilibrium (HWE) was observed for this gene variant within both the male cases (p= 0.05) and control (p=0.001) groups , whereas in female group controls showed significant deviation (p= 0.001). In males the association in dominant model was the highest but not significant (OR: 1.44; p=0.673), no significant associations were observed in the recessive, co-dominant and dominant mode of inheritance whereas in females the association in dominant model was very negligible and significant (OR:6.47; p=0.014) and the association in co-dominant model (QQ vs. KQ) = (KQ vs. KK) was found comparatively higher (OR: 3.26; p= 0.038) as compared to recessive model (QQ vs. KK/KQ) (OR: 2.09; p=0.548).The Q allele did not differ between obese cases and control subjects in both male(OR: 0.2263; 95%CI 0.0589-0.8689 ; p= 0.0304) and females (OR: 0.4681; 95%CI: 0.11-1.99; p= 0.2358). Further in females the Q allele was found to have significant association with BMI (p=0.0038). The estimates of relative risk (odds ratio) for the *ENPP1* K121Q polymorphism in obese subjects and normal control subjects for both genotypic and allelic frequencies were observed to be statistically significant (genotypic frequencies: OR: 0.1546; 95%CI: 0.03-0.81; p=0.0275; allelic frequencies: OR: 0.2263, 9s5%CI: 0.06-0.87; p=0.0304) in females. In conclusion *ENPP1/PC-1* K121Q polymorphism influences the risk of obesity in north Indian Punjabi female population whereas no association was found in Male population.

1220T

A study on genetic variation associated with visceral adipose tissue and interaction of life style on the expression of genetic variation. *H. Kwon, J. Park, B. Cho, J. Yoon.* Family Medicine, Seoul National University Hospital, Seoul, South Korea.

Introduction : Obesity, especially abdominal obesity is known as a risk factor for various diseases, and also to increase mortality. Among components of abdominal obesity, visceral adipose tissue (VAT) is a risk for various metabolic diseases and mortality. Genetic variations affecting abdominal obesity have been identified, and interactions between genetic variations and lifestyle factors also have been studied. But there has been no study on genetic variations altering VAT in Koreans, nor study on interaction between genetic variations and life style factors affecting VAT in Koreans. In this study, I have tried to find the association of VAT and SNPs which were previously known to be associated with obesity, and also to test the interaction of life style such as smoking, alcohol drinking and physical activity on the association of VAT and SNPs. Methods : I have selected 17 SNPs previously known to be associated with obesity, and analyzed selected SNPs through realtime PCR from the blood of Korean men aged 20 to 65 who took comprehensive health checkup programs including abdominal fat analysis with CT scan. I have tested the association between selected SNPs and obesity traits including VAT with linear regression analysis, and also checked interaction of life style factors on the association. I also tried stratified analysis with the criteria of body mass index 25kg/m2. Results : Rs9939609 was associated with VAT in dominant model, with AT/AA allele type tends to 6.8cm2 higher in VAT, 0.6kg/m2 higher in body mass index, 1.4cm thicker in waist circumference, 0.7% more in body fat, and 11.2cm2 higher in abdominal subcutaneous fat. On interaction analysis, rs1514175, rs2112347, rs10767664 were affected by alcohol drinking, and rs1718537 and rs3817334 were affected by physical activity on their association with VAT, but on stratified analysis by life style factors there was no significant difference in VAT according to allele type. On stratified analysis by body mass index, rs713586 and rs3810291 were associated with VAT only in subjects with body mass index higher than 24kg/m2 Conclusions : Rs9939609, rs713586 and rs3810291 were associated with VAT in Korean male, and this finding could be implemented in practice and study in the field of obesity.

1221F

Variants in fat mass and obesity gene are associated with variability of body mass index from childhood to adulthood. *S. Li, T. Zhang, C. Fernandez.* Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, LA.

Variants in the fat mass and obesity gene (*FTO*) have been associated with phenotypic variability of body mass index (BMI) in large-scale genome-wide association studies. However, no studies have examined associations between the *FTO* variants and BMI variability across the life course. Taking advantage of a longitudinal black-white cohort beginning in early childhood (n=1,201with 10,313 observations; age: 4.0-51.0 years with an average follow-up period of 31.4 years), we examined the variant rs994134 in the *FTO* gene with BMI variability from childhood to adulthood. BMI residual standard deviation (RSD) derived from mixed models was used as a measure of BMI variability. The association of rs994134 with BMI and RSD was similar in blacks and whites (P>0.49 for interactions with race). On average, the T allele was associated with 0.60 kg/m^2 increase in BMI (P=0.005) and 0.09 kg/m^2 (P=0.008) increase in RSD, relative to the C allele after adjusting for race and sex. The association of rs994134 with BMI RSD persisted (P=0.04) when RSD was adjusted for absolute BMI values. In conclusion, the BMI-increasing allele of the *FTO* variant rs994134 is also associated with BMI variability from childhood to adult life.

1222W

Assessment of copy number variation in two protocadherin genes and its effect on childhood obesity predisposition in Koreans. *S. Moon, M.Y. Hwang, B.J. Kim.* Korea National Intsitute of Health, Cheongwon-gun, Chungcheongbuk-, South Korea.

Recently, the prevalence of childhood obesity has significantly increased in industrialized countries, including Korea, and now controlling obesity is becoming an economic burden. However, knowledge of the risk factors associated with obesity is still limited. In this study, we aimed to discover additional obesity-associated loci in children. To achieve this, we conducted a genome-wide association study of copy number variation (CNV) using whole-exome sequencing (WES) data from 102 extremely obese and 86 normal weight children.Using a logistic regression analysis, we identified a CNV locus that overlapped two protocadherin genes, PCDHB7 and PCDHB8, which are brain function–related genes (P-value = 6.40 × 10^{-4}, odds ratio = 2.2189). A subsequent replication analysis using WES data from 203 obese and 291 normal weight children showed that this CNV region satisfied the genome-wide significance standard (Fisher's combined P-value = 7.29 × 10^{-5}). Moreover, correlation test using 199 additional samples supported significant association between CNV and increased body mass index. This region also showed a meaningful, although not enough significant, association with obesity in 4,694 community-based adult samples. Our findings suggest that differences in the common CNV region at 5q31.3 may have an impact on the pathophysiology of obesity.

1223T

Assessing variation across 98 established loci for body mass index in American Indians. *Y.L. Muller, P. Piaggi, P. Chen, G. Wiessner, C. Okani, G. Skelton, S. Kobes, M. Hohenadel, W.C. Knowler, C. Bogardus, R.L. Hanson, L.J. Baier. NIDDK, NIH, Phoenix, AZ.*

Meta-analyses of genome-wide association studies (GWAS) have identified ~98 established ($p < 5 \times 10^{-8}$) loci for Body Mass Index (BMI). The GWAS data included in these meta-analyses were predominately derived from individuals of European ancestry. To assess the role of these established BMI loci in American Indians who have a high prevalence of obesity, we analyzed the lead SNPs at these loci in a longitudinally studied population of full heritage Pima Indians who were informative for BMI in adulthood (maximum BMI recorded at age ≥15 years, n=3412), BMI z-score in childhood (maximum age-and-sex specific z-score at age 5-20 years, n=2296) and percent body fat in adulthood (n=555). Among the 98 lead SNPs, 16 were monomorphic or had a minor allele frequency (mAF) <0.01 in Pima Indians. Of the remaining 82 SNPs (mAF≥0.01) analyzed, 11(*TMEM18, MRPS33P4, TCF7L2, ZFP64, LMX1B, PRKD1, FTO, GRP, NLRC3, TAL1, RPTOR*) nominally associated with BMI in Pima Indian adults (p=1.6 x 10^{-4} to 0.05) in a direction consistent with that reported in the European meta-analyses. A multi-allelic genetic risk score (GRS), which summed the risk alleles for all 82 SNPs, showed a significant trend for the relationship between increasing number of risk alleles in Pima Indians and BMI in adulthood (β=0.16 kg/m^2 per risk allele, p=2.8 x 10^{-8}), BMI z-score in childhood (β=0.024 SD, p=6.8 x 10^{-9}) and % body fat (β=0.17 %, p=3.0 x10^{-4}). To determine whether different variants at these loci also contribute to BMI in American Indians, ~6000 SNPs which tag (r^2≥0.85) common variation (mAF ≥0.05) detected in whole genome sequence data from 295 Pima Indians across all 98 loci were evaluated in an expanded population-based sample of 7710 Pima Indians (including mixed heritage individuals). Tag SNPs in 26 of 98 loci had additional variants that associated with BMI (p=8.1x10^{-6} to 0.002) after accounting for the number of tag SNPs at each locus using the Sidak correction. The strongest BMI signal was in *GPR* (rs1563713, p=8.1 x 10^{-5}). In conclusion, nominal but directionally consistent replication of established BMI variants was observed in Pima Indians. In aggregate these SNPs had a notable effect on BMI in both adulthood and childhood, suggesting that these variants, or nearby functional variants, influence BMI in American Indians. Analysis of additional variation in these loci provided evidence for additional variants in some of the loci.

1224F

Genome-wide gene-based SNP analyses identify mir-486 as a predictive marker of weight loss in obese, non-diabetic patients upon low-caloric diet intervention. *A. Valsesia[1], J. Carayol[1], N. Gheldof[1], G. Lefebvre[1], S. Metairon[1], C. Chabert[1], N. Viguerie[2], D. Langin[2], P. Descombes[1], M.E. Harper[3], R. Mc Pherson[4], R. Dent[5], A. Astrup[6], W. Saris[7], J. Hager[1]. 1) Nestlé Institute of Health Sciences, Lausanne, Switzerland; 2) INSERM UMR1048, Obesity Research Laboratory, Institute of Metabolic and Cardiovascular Diseases, University of Toulouse, France; 3) Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa; 4) Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada; 5) Ottawa Hospital Weight Management Clinic, The Ottawa Hospital, Ottawa, Canada; 6) University of Copenhagen, Department of Nutrition, Exercise and Sports, Faculty of Science, Copenhagen, Denmark; 7) Department of Human Biology, NUTRIM, School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre + (MUMC+), Maastricht, The Netherlands.*

Obesity is a major risk factor for a number of co-morbidities including cardiovascular disease, dyslipidemia, hypertension, insulin resistance and type 2 diabetes (T2D) as well as cancer. Multiple studies have shown that weight loss through energy-restricted dietary interventions improves the metabolic dysfunctions involved in metabolic syndrome. Nevertheless, large inter-individual variability is observed regarding the capacity to lose and maintain weight. In addition, initial BMI has been shown to have some predictive value but is not useful in obesity intervention studies where subjects have similar starting BMI's. Here, we present a gene-based genome-wide association analysis of two large dietary intervention studies: the Ottawa weight maintenance program (n=1166), used as discovery cohort; and the DiOGenes randomized clinical trial (n=798, NCT00390637), used as validation cohort. Our analyses aimed at identifying SNPs associated with weight loss upon a low-caloric diet (LCD) intervention. Using a two-stage approach, we found that SNPs nearby the mir-486 gene were associated with weight loss (FDR=5.5% in the validation cohort). Nominal gene-based p-values were as follows: p=1.9e-4 in the discovery cohort, p=5.2e-3 in the validation cohort and p=1.1e-5 from meta-analysis. Quantification of mir-486 levels in adipose tissue biopsies (n = 440) in the validation cohort showed that baseline levels were associated with clinical outcomes both for weight and glycemic control. Upon LCD, we observed significant association between baseline mir levels with weight loss (p=0.019) and marginal association with improvement in glycemic control (p=0.10). Six months after the LCD, the association was significant for both endpoints (p=0.034 and 0.021, respectively for weight and glycemic control). In conclusion, our analyses identified SNPs, encompassing the mir-486 locus, associated with weight loss upon low-calorie diet. Analyses of baseline mir-RNA levels from adipose tissue biopsies showed association with long-term clinical improvements, both for weight and glycemic outcomes. Our study identified a potential biomarker for stratification of obese, non-diabetic subjects prior to weight loss intervention.

1225W

A meta-analysis of genome-wide association studies for susceptibility loci to diabetic nephropathy in Japanese patients with type 2 diabetes. *M. Taira[1], M. Imamura[1,4,5], A. Takahashi[2,6], Y. Kamatani[2], M. Kubo[3], S. Maeda[1,4,5].* 1) Laboratory for Endocrinology, Metabolism and kidney Diseases, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 2) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 3) RIKEN Center for Integrative Medical Science, Kanagawa, Japan; 4) Department of Advanced Genomic and Laboratory Medicine Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan; 5) Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, Japan; 6) National Cerebral and Cardiovascular Center, Omics Research Center, Osaka, Japan.

   Background: Diabetic nephropathy (DN) is a leading cause of end-stage renal disease in Western countries as well as in Japan. Although several reports have shown that genetic factors play a pivotal role in the development and progression of DN, worldwide efforts to identify susceptibility loci for DN have not been successful so far. To explore novel genetic loci associated with the susceptibility to DN, we performed a meta-analysis of genome-wide association studies (GWAS) for DN in Japanese patients with type 2 diabetes (T2DM). **Methods**: We divided Japanese patients with T2DM registered in Bio-Bank Japan, who had been analyzed in previously reported GWAS for T2DM (Set-1: 9,343, Set-2: 3,607), into two groups; 1) DN cases, defined as patients having overt nephropathy or end-stage renal disease, 2) controls, who did not have any sign of DN and with diabetic retinopathy or with long duration of diabetes (> 5 years). As our discovery stage, we examined two independent case-control groups (Study-1: 2,380 DN cases and 5,238 controls, Study-2: 429 DN cases and 358 controls) obtained by genotype imputation using 1000 genomes reference data (phased JPT+CHB+CHS, March 2012). Results of the two GWAS were combined with an inverse variance method. **Results:** We did not observe any loci showing genome-wide significant association with DN in the initial GWAS meta-analysis, and we selected candidate SNP loci (p-value < 1.0×10^{-4}) for further analysis. Subsequently, we performed *de novo* genotyping for the candidate SNP loci by Multiplex-PCR Invader assay in an independent case-control study (Set-3: 1,213 DN cases and 1,298 controls) and all association data (Set-1, Set-2 and Set-3) were combined with a meta-analysis. After integrating all association data by a meta-analysis, we identified that the association of one SNP locus reached a genome-wide significance level (combined *P*-value =7.74×10^{-10}, odds ratio 1.23) and six SNP loci showed borderline association with DN (5×10^{-8} < combined *P*-value < 5×10^{-7}). There is no overlap between the above seven SNP loci and previously-reported loci for susceptibility to DN. **Conclusion**: We have identified a novel locus for conferring susceptibility to DN. Further studies using independent case-control groups are required to confirm the association of this locus with DN. .

1226T

Expression QTL analysis reveals *RHBDF2* as a candidate gene for Non-Alcoholic Steatohepatitis. *M. Olivier[1], E. Drigalenko[1], M.I. Goldblatt[2], J.R. Wallace[2], H.H.H. Göring[1,5], D.E. Kleiner[3], S. Gawrieh[4].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Surgery, Medical College of Wisconsin, Milwaukee, WI; 3) Laboratory of Pathology, National Cancer Institute, Bethesda, MD; 4) Department of Medicine, Indiana University, Indianapolis, IN; 5) current affiliation: South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.

   Nonalcoholic fatty liver disease (NAFLD) is characterized by hepatic steatosis in the absence of significant alcohol use. Some patients with NAFLD have nonalcoholic steatohepatitis (NASH), which may progress to cirrhosis and liver failure. Both genetic and environmental factors contribute to the development and progression of the disorder, but only few genes have been identified so far that contribute to NAFLD or NASH. We collected liver biopsy samples from 76 morbidly obese patients of Northern European descent undergoing bariatric surgery. All biopsies were evaluated according to the NASH Clinical Research Network system the NAFLD Activity Score (NAS) was recorded. 30 individuals had unremarkable histology and served as controls, 28 had NAFLD but not steatohepatitis, 9 as borderline steatohepatitis, and 9 individuals were diagnosed with definite NASH. Gene expression in the liver was analyzed using Illumina BeadChip arrays, HumanWG-6. In the 76 samples, expression levels of 23 genes were significantly associated with NAS scores, including *TGFBI*, *FABP4*, and several collagen genes. To identify sequence variants associated with these changes in gene expression (eQTL), SNPs were selected from Affymetrix SNP6 chop data within 100 kb of the expression probes for these 23 genes, and analyzed for association with the respective gene expression level. Two sequence variants upstream of the transcription start site of the gene *RHBDF2* on chromosome 17 were significantly associated after correction for multiple testing ($p<1 \times 10^{-3}$), and four additional variants were nominally associated ($p<0.05$). This gene (Rhomboid 5 Homolog 2) has not previously been reported to contribute to the development or progression of NAFLD. *RHBDF2* has been associated with esophageal cancer, and related pathways include immune system and signaling by GPCR. GO annotations related to this gene include serine-type endopeptidase activity and growth factor binding. Interestingly, the expression of its paralog gene, *RHBDF1*, is also significantly associated with NAS score. Our analysis suggests that *RHBDF2* (and possibly *RHBDF1*) may contribute to NAFLD and NASH, however, the specific role(s) of *RHBDF1* and *RHBDF2* remain to be elucidated.

1227F

SUGP1 is a novel regulator of cholesterol metabolism. *M.W. Medina[1], M. Kim[1], C.Y. Yu[1], E. Theusch[1], D. Naidoo[1], K. Stevens[1], Y.L. Kuang[1], E. Schuetz[2], A. Chaudhry[2]. 1) Children's Hospital Oakland Research Institute, Oakland, CA; 2) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN.*

A large haplotype on chromosome 19p13.11, tagged by rs10401969 in intron 8 of SURP and G patch domain containing 1 (*SUGP1*), has been associated with a number of cardiometabolic phenotypes including plasma total cholesterol, LDL-cholesterol, and triglycerides as well as incidence of coronary artery disease, hepatic steatosis and nonalcoholic fatty liver disease. Several recent studies have suggested that transmembrane 6 superfamily member 2, *TM6SF2*, is the causal gene within the locus. However, in individuals of European ancestry, the 19p13.11 haplotype spans ~360kb, contains 16 protein-coding genes, and harbors several SNPs in perfect or near perfect linkage disequilibrium ($r^2 > 0.95$, D'=1) with rs10401969. Thus, we postulated that this locus may harbor additional risk genes, including the putative splicing factor *SUGP1*. Using a mini-gene construct, we found that rs10401969 regulates *SUGP1* exon 8 skipping, causing nonsense-mediated mRNA decay. Through an RNA electrophoretic mobility shift assay (EMSA), we observed altered binding of HNRNPA1, a well-known splicing factor, depending on rs10401969 allele. Hepatic *Sugp1* overexpression in CD1 male mice increased plasma cholesterol levels 20-50%. In human hepatoma cell lines, *SUGP1* knockdown stimulated 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGCR*) alternative splicing and decreased *HMGCR* transcript stability. HMGCR encodes the rate-limiting step of cholesterol biosynthesis, thus *SUGP1* knock-down also caused reduced cholesterol synthesis and increased LDL uptake, consistent with the change in plasma cholesterol observed *in vivo*. Furthermore, the directionality of the effects from modulating SUGP1 both *in vivo* and *in vitro* was also consistent with the original GWAS finding that the rs10401969 minor allele is associated with reduced plasma LDL-cholesterol. Our results strongly support a role for SUGP1 as a novel regulator of cholesterol metabolism and suggest that it contributes to the relationship between rs10401969 and plasma cholesterol. Thus, our findings highlight the possibility that a GWAS locus may harbor more than one causal gene.

1228W

The impact of genetic risk on effectiveness of lifestyle intervention in a general population. *M. Conomos, J.C. Lovejoy, A.T. Magis. Arivale, Seattle, WA.*

With increased focus on personalized and precision medicine in the media, public interest in lifestyle intervention aimed at improving biomarker levels associated with disease risk that is tailored to personal genetics has grown considerably in the past few years. However, it is well known that an individual's phenotype for many complex traits is determined by both a heritable genetic component that cannot be altered, as well as an environmental component that may be modifiable. In this study, we investigate the impact of genetic risk on the effectiveness of lifestyle intervention aimed at improving clinical markers of cardiovascular risk. Blood lipid levels were regularly collected and used for lipid profiling, and whole genome sequencing was performed on a cohort of more than 1000 individuals who were enrolled in a lifestyle intervention program for at least six months. Using previously published GWAS data from the entire genome, we construct polygenic scores predictive of total cholesterol, LDL and HDL cholesterol, and triglyceride levels that are used to partition individuals into genetic risk strata for having unhealthy lipid levels. Taking advantage of the longitudinal blood measurements collected, we are able to quantify the improvement of lipid levels over time in the cohort as a whole, as well as within the genetic risk strata. This allows us to test the hypothesis that, among individuals with unhealthy baseline lipid levels, lifestyle intervention can more effectively improve lipid levels for individuals in low genetic risk strata than individuals in high genetic risk strata. This study provides insight into strategies for genetics-based personalization of lifestyle intervention to improve specific clinical markers and optimize wellness.

1229T

Identifying genetic variants for serum lipoprotein(a) independent of kringle repeat polymorphism. *S.M. Hosseini[1,2], J.K. Snell-Bergeon[3], A.P. Boright[4], A.J. Canty[5], L. Sun[6,7], S.B. Bull[6,8], S.M. Marcovina[9], J.D. Brunzell[9], A.D. Paterson[2,6], the DCCT/EDIC Research Group.* 1) Ted Rogers Cardiac Genome Clinic, Hosp Sick Children, Toronto, ON, Canada; 2) Genetics and Genome Biology Program, The Hospital for Sick Children Research Institute, Toronto, Canada; 3) Barbara Davis Centre for Diabetes, University of Colorado, Aurora, CO, USA; 4) Department of Medicine, University of Toronto and LMC Diabetes & Endocrinology, Toronto, Canada; 5) Department of Mathematics & Statistics, McMaster University, Hamilton, Canada; 6) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 7) Department of Statistical Sciences, University of Toronto, Toronto, Canada; 8) Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada; 9) Northwest Lipid Metabolism and Diabetes Research Laboratories, Division of Metabolism, Endocrinology and Nutrition, Department of Medicine, University of Washington, Seattle, WA, USA.

Lipoprotein(a) [Lp(a)] is a highly heritable, independent risk factor for cardiovascular diseases (CVD). Lp(a) level is mainly determined by variation in the *LPA* region with significant contribution from a VNTR (variable number tandem repeat) in *LPA* coding for the KIV-2 (kringle IV type 2) domain of apolipoprotein(a). Diabetes is an important CVD risk factor and may also influence Lp(a) metabolism. To investigate genetic determinants of Lp(a) in diabetes and identify novel quantitative trait loci (QTL) for Lp(a) independent of KIV-2 copy number (KCN) we studied two cohorts of white European ancestry with type 1 diabetes (T1D): The Diabetes Control and Complications Trial (DCCT) and Coronary Artery Calcification in T1D (CACTI). Both cohorts were genotyped using Illumina arrays, with imputation of untyped SNPs using the 1000 Genomes data for the conditional GWAS. Here, we report four major findings. First, in the univariate GWAS, 56 genotyped SNPs in the *LPA* region show genome-wide significant ($p<5\times10^{-8}$) association with Lp(a) in DCCT, all of which replicate in CACTI ($3.5\times10^{-7}<p<0.01$). Second, KCN alone, determined using multiplex qPCR, explains 19% of Lp(a) variation in DCCT. 38 SNPs in the *LPA* region show association with KCN ($p<5\times10^{-8}$), with evidence for 15 independent signals in multivariate linear regression, capturing >41% of KCN variation. A genetic risk score for KCN, calculated based on these 15 SNPs, explains 14% of variation in Lp(a) in CACTI. Third, after adjusting for the effect of KCN on Lp(a) in DCCT, 63 SNPs from *LPA* region are still associated with Lp(a) ($p<5\times10^{-8}$), of which 15 showed independent evidence in multivariate analysis (5 with $p<5\times10^{-8}$). Fourth, conditional GWAS accounting for the main *cis* effects, revealed no genome-wide significant association in DCCT, but discovered one in CACTI (rs111850011, $p=3\times10^{-8}$). This SNP is located 26 kb upstream of steroid-5-alpha-reductase, alpha polypeptide 2 (*SRD5A2*). 5α-reductase is an important regulator of lipoprotein metabolism. The SNP shows consistent direction of effect in DCCT without being significant ($p=0.6$). Attempts at replication in independent populations are underway. In conclusion, consistent with studies in non-diabetics, we identify several SNPs in the *LPA* region that contribute to serum Lp(a) in T1D. We show that these variants exert their effect either independent of KIV-2 VNTR or as its proxies. Finally, we identify a novel genome-wide significant *trans* QTL for Lp(a).

1230F

Multiplexed targeted resequencing identifies rare noncoding variants associated with high HDL cholesterol. *S.A. Khetarpal[1,2,4], P.L. Babb[2,3,4], W. Zhao[1,2], S. DerOrhannessiani[1,2], W.F. Hancock-Cerutti[1,2], S. Elwyn[1,2], T. Tran[1,2], C.D. Brown[2], B.F. Voight[2,3,4], D.J. Rader[1,2,4].* 1) Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Department of Pharmacology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 4) equal contribution.

Genome-wide studies in large populations have uncovered many loci influencing levels of high-density lipoprotein cholesterol (HDL-C), a biomarker for risk of heart disease. However, the contribution of the identified genes is unknown and difficult to ascertain from current genotyping studies of common variants with relatively small effects, prompting the need for efficient, cost-effective methods to identify rare variants with putatively larger impact. Here, we utilized a multiplexed inversion probe (MIP) target capture approach to resequence seven HDL-C associated loci in 797 individuals with extremely high HDL-C (>95th percentile for age and gender) vs. 735 normal HDL-C controls. Our targets included the coding regions of *APOC3*, *SCARB1*, *CETP*, *LIPG*, *CCDC92*, *ZNF664*, and *GALNT2* (>29 kb), along with regulatory features in the noncoding genome (>16 kb). Validation in 1,119 of the 1,532 participants on the Exome Array resulted in >95% genotyping concordance across SNPs. From our resequencing, we first identified known GWAS signatures in *GALNT2*, *CETP*, and *LIPG* and found consistent associations with high HDL-C as reported in the larger GWAS, thus supporting our study design involving HDL-C phenotypic extremes. Second, we discovered multiple novel, rare, noncoding variants individually associated with increased HDL-C ($P<1.74\times10^{-05}$, Score Test). Additionally, gene-burden testing of aggregated coding variants across each region with HDL-C identified a significant excess of variants in *SCARB1* associated with high HDL-C levels, consistent with recent discoveries in humans with complete *SCARB1* loss-of-function. Our targeted resequencing at the extremes of HDL-C thus revealed multiple novel noncoding variants for further study and supports the rationale for larger efforts to resequence the noncoding genome to find rare variants for complex traits.

1231W

Replication of a single nucleotide polymorphism variant in *CETP* gene associated with large-HDL particle in the ClinSeq® Study. *H. Sung[1], M. Sampson[2], K. Lewis[3], D. Ng[3], S.G. Gonsalves[3], J.C. Mullikin[4,5], A. Remaley[6], A.F. Wilson[1], National Institutes of Health Intramural Sequencing Center(NISC).* 1) Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD, USA; 2) Clinical Center, NIH, Bethesda, MD, USA; 3) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD, USA; 4) National Institutes of Health Intramural Sequencing Center(NISC), National Human Genome Research Institute, NIH, Bethesda, MD, USA; 5) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD, USA; 6) Lipoprotein Metabolism Section, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of rare sequence variants with traits related to Coronary Artery Disease (CAD). The study currently includes more than 1000 non-smoking participants between the ages of 45 to 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, MD. Whole-exome sequencing was performed with the Agilent SureSelect 38Mb and 50Mb capture kits for 387 and 325 individuals, respectively, at the NIH Intramural Sequencing Center. Single nucleotide variants (SNVs) common to both capture regions with a call rate > 98% and minor allele frequency (MAF) > 1% were used to check for cryptic relatedness and for mis-specified population stratification by multidimensional scaling analysis – 635 unrelated European Americans (EAs) remained. For each capture kit with EAs only, SNVs with at least one homozygote and a call rate greater or equal to 50% were included. The two capture regions with only SNVs in common were merged, yielding 439,807 SNVs. Of these SNVs, 74% and 46% had MAFs < 0.01 and < 0.001, respectively. The SNVs with MAF < 0.01 were collapsed into a single derived variant for each genomic region defined by hotspot blocks. Collapsed variants were coded as the proportion of minor allele occurring within each region; common variants were coded as the number of minor alleles (scaled from 0 to 1). Lipoprotein particle profiles including High Density Lipoprotein particle (HDLp) by nuclear magnetic resonance spectroscopy were measured at LipoScience/LabCorp Global Research Services. In this study, tests of association of each SNV with each classified HDLp were performed on untransformed, log-transformed and rank-inverse transformed HDLp with simple linear regression, adjusting for age, sex, BMI, use of medication. The SNVs rs1532625 and rs7205804 in the intron of *CETP* gene were associated (P = 1e-05) for all untransformed and transformed large-HDLp traits (combining the HDLp with estimated particle diameters of 9.7, 10.5 and 12 nm) after Bonferroni correction. This finding replicates the association by Reilly et al.(2013) between rs1532625 and HDL level.

1232T

Genome-wide association study identifies *TRAF3* as a novel susceptibility locus for Gallstone disease in Latin Chilean population. *B. Bustos[1], E. Pérez-Palma[1], S. Buch[2], L. Azócar[3], E. Riveras[3], C. Moraga[4], M. Toilat[5], J. Hampe[2], P. Nürnberg[5], R.A. Gutierrérrez[4,6], G.V. De Ferrari[1,6], J.F. Miquel[3,6].* 1) Centro de Investigaciones Biomédicas, Universidad Andres Bello, Santiago, Chile; 2) Medical Department I, University Hospital Dresden, TU Dresden, Dresden , Germany; 3) Depto. de Gastroenterología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chille; 4) Depto. Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile; 5) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 6) FONDAP Center for Genome Regulation (CGR).

Gallstone disease (GSD) is a gastrointestinal disorder with high prevalence in western countries. Genome-wide association studies and meta-analysis approach on GSD have reported a handful of lithogenic variants, being the most important the sterolin transporter ABCG8 (p.D19H). This signal has been replicated in different populations across the world, including Latin Chileans, however it only explains a small portion of the population attributable risk. The Chilean population has one of the highest prevalence of GSD in the world, but has been the target of only a few genetic studies. Here, we show the results from the first GWAS performed on the Chilean population. Subjects were recruited from community medical centers in Santiago de Chile. Status of GSD (cases or controls) was diagnosed by abdominal ultrasonography. Genotyping was done using Affymetrix Axiom LAT 1 World array Plates. Genome-wide Imputation was performed with IMPUTE2 using the 1000 Genomes Project Phase 3 reference panel. Association analyses were done using SNPTEST. Replication and gene expression analyses were performed using RT-qPCR with TaqMan probes and custom DNA primers, respectively. We selected 10 candidate loci surpassing suggestive genome-wide significance (p < 1×10^{-5}) in a discovery cohort of 1,095 individuals (529 cases, 566 controls). Replication was done in a cohort of 1,868 individuals (707 cases and 1,161 controls). We successfully replicated 2 signals: the previously known ABCG8 locus (discovery p=5.26×10^{-6}, OR=1.88; replication p=0.001, OR=1.58) and a novel signal on chromosome 14 (discovery p=4.9×10^{-6}, OR=1.50; replication p=0.003, OR=1.29) inside the TNF receptor-associated factor gene (TRAF3). Gene expression analysis in 22 RNA samples from human gallbladder tissues (11 cases, 11 controls) showed a significant decrease of TRAF3 levels in cases versus controls (fold change=0.51, p=0.015). Same results were observed in 41 RNA samples from duodenum tissue (19 cases, 22 controls; fold change=0.50, p=0.001). TRAF3 regulates pathways leading to the activation of NF-kappa-B and their lower expression levels could be linked to enhanced chronic and acute gallbladder inflammation observed in gallstone patients, which is implicated both in the pathogenesis as in development of symptoms in GSD. These results provide new insights into the genetic susceptibility of this prevalent disease in a high risk Latin population. **Funding:** FONDECYT #1130303-JFM, #1140353-GVD, FONDAP #1509000.

1233F

Genetic and environmental determinants of the susceptibility of hyperlipidemia in Taiwanese using Taiwan Biobank dataset. *C.N. Hsiung[1,2], C.Y. Shen[1,2].* 1) Taiwan Biobank, Academia Sinica, Taipei, Taipei, Taiwan; 2) Institute of Biomedical Sciences,Academia Sinica, Taipei, Taipei, Taiwan.

Hyperlipidemia is an important risk factor for cardiovascular disease and is affected by both genetic and environmental factors. Recent genome wide association studies (GWAS) have identified several loci associating with lipid level, but these have been identified mainly in European populations. We used Taiwan biobank database in order to identify not only environmental factors but also genetic markers predicting hyperlipidemia in a Taiwanese population The Taiwan biobank database has collected information from over 56,000 participants and whole-genome typing has been performed in 14,432 individuals. We used logistic regressions to build prediction models for hyperlipidemia. Whole-genome genotyping using TWBv1.0 chip (653,291 single-nucleotide polymorphisms, SNPs, specifically for the Han Chinese in Taiwan) is performed to generate polygenic scores. Our findings show that the model based on conventional risk factors resulted in good prediction with high efficiency area under the receiver operating characteristic (ROC-AUC=0.72). We then examined the risk associated with 55 SNPs for hyperlipidemia detected by GWAS, and resulted in a polygenic score, which increased the risk of hyperlipidemia (OR=2.64, 95% C.I. =2.477-2.828). The full prediction model, based on environmental risk factor and polygenic score together, led to better efficiency (ROC-AUC=0.82, P<0.0001). This study builds a prediction model, including environmental risk factors and polygenic score, for hyperlipidemia. On the basis of this finding, further measures aiming at public health intervention are designing to decrease the incidence of hyperlipidemia and cardiovascular disease in Taiwan population.

1234W

Meta-analysis of genome-wide association studies in Chinese and European populations identify novel loci associated with circulating monounsaturated fatty acids. *H. Li[1], H. Yao[1], J. Zhu[1], T. Tanaka[2], W. Guan[3], J.H.Y. Wu[4], B.M. Psaty[5,6], B. McKnight[7], I.B. King[8], Q. Sun[9,10], M. Richard[11], A. Manichaikul[12,13], A.C. Frazier-Wood[14], E.K. Kabagambe[15], P.N. Hopkins[16], J.M. Ordovas[17,18], L. Ferrucci[2], S. Bandinelli[19], D.K. Arnett[20], Y.D.I. Chen[21], S. Liang[22], D.S. Siscovick[5,23,24], M.Y. Tsa[25], S.S. Rich[12], M. Fornage[11], F.B. Hu[9,10], E.B. Rimm[9,10], L.M. Steffen[26], A.P. Morris[27], X. Lin[1].* 1) Chinese Academy of Sciences, Shanghai, China; 2) Translational Gerontology Branch, NIA, Baltimore, Maryland, United States of America; 3) Division of Biostatistics, School of Public Health, University of Minnesota; Minneapolis, Minnesota, United States of America; 4) George Institute for Global Health, Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia; 5) Cardiovascular Health Research Unit, Departments of Medicine, University of Washington, Seattle, Washington, United States of America; 6) Group Health Research Institute, Group Health Cooperative, Seattle, Washington, United States of America; 7) Department of Biostatistics, University of Washington, Seattle, Washington, United States of America; 8) Department of Internal Medicine, University of New Mexico, Albuquerque, New Mexico, United States of America; 9) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America; 10) Department of Nutrition, Harvard T.H. Chan School of Public Health, Harvard University, Cambridge, Massachusetts, United States of America; 11) Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas, United States of America; 12) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, United States of America; 13) Department of Public Health Sciences, Biostatistics Section, University of Virginia, Charlottesville, Virginia, United States of America; 14) USDA Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States of America; 15) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America; 16) Department of Internal Medicine, University of Utah, Salt Lake City, Utah, United States of America; 17) Nutrition and Genomics Laboratory, Jean Mayer-US Department of Agriculture, Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States of America; 18) Department of Epidemiology and Population Genetics, National Center for Cardiovascular Investigation, Madrid, Spain; 19) Geriatric Unit, Azienda Sanitaria Firenze, Florence, Italy; 20) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, United States of America; 21) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, United States of America; 22) Department of Laboratory Medicine & Pathology, University of Minnesota, Minneapolis, Minnesota, United States of America; 23) Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, United States of America; 24) New York Academy of Medicine, New York, New York, United States of America; 25) Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts, United States of America; 26) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States of America; 27) Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, United Kingdom.

Monounsaturated fatty acids (MUFAs) have been suggested to have protective effects against cardio-metabolic disorders including metabolic syndrome, type 2 diabetes (T2D), and cardiovascular disease (CVD). Previous genome-wide association studies (GWAS) have identified loci for plasma phospholipid and erythrocyte palmitoleic acid (16:1n-7) and oleic acid (18:1n-9) levels in populations of European origin. To identify loci for other MUFAs, including vaccenic aicd (18:1n-7), gondoic acid (20:1n-9), erucid acid (22:1n-9), and nervonic acid (24:1n-9), we performed ethnic-specific GWAS meta-analyses and trans-ethnic meta-analyses in over 15,000 participants from nine cohorts of Chinese- and European-ancestry. Ten loci were genome-wide significant ($\log_{10}$Bayes factor)>6) in the trans-ethnic meta-analysis, including five novel loci (*PCDH15*, *STARP1*, *GGCT*, *IQCA1* and *ST8SIA5*), and five previously reported ones (*FADS1/2*, *PKD2L1*, *GCKR*, *HIF1AN* and *LPCAT3*). Novel associations were added to previously reported loci (*FADS1/2*, *PKD2L1* and *GCKR*) in our study. The previously reported associations of *PKD2L1*, *FADS1/2*, *GCKR* and *HIF1AN* with 16:1n-7 and of *FADS1/2* and *LPCAT3* with 18:1n-9 were confirmed in the Chinese populations and in the trans-ethnic meta-analysis. Additional genes and novel pathways were identified in the

gene- and pathway-based analysis. Our findings provide novel insight into the genetic basis for MUFA metabolism and emphasize the benefits of trans-ethnic meta-analysis for complex traits by discovering additional novel variants and testing for the difference of the identified loci across ethnic groups.

1235T

Parent-of-origin effects analysis of NMR metabolites. *N. Pervjakova[1,2], T. Haller[1], K. Kristiansson[4,6], A. Joensuu[3,4,5], M. Perola[4,6], R. Mägi[1].* 1) Estonian Genome Center University of Tartu, Tartu, Tartumaa, Estonia; 2) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia;; 3) Institute for Molecular Medicine (FIMM), University of Helsinki, Biomedicum 2, Tukholmankatu 8, 00290 Helsinki, Finland; 4) Genomics and Biomarkers Unit, Department of Health, National Institute for Health and Welfare (THL), Biomedicum 1, Haartmaninkatu 8, 00290 Helsinki, Finland; 5) Diabetes and Obesity Research Program, Research Programs Unit, University of Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland; 6) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland.

   Parent-of-origin effect (POE) occurs when the allelic effect on phenotype depends on whether the allele is inherited from mother or father. POE can be mediated through several mechanisms, such as genomic imprinting, trans-generational effects, *in utero* effects and the maternal environment. The aim of this study was to identify POE of common variants in NMR metabolites. We considered 5,853 individuals from Estonian Biobank (EGCUT) for 82 NMR metabolites. Individuals were imputed to the 1000 Genomes Project reference panel (March 2012 release). POE analysis was conducted using POE method in QUICKTEST using markers with MAF ≥ 1%, and imputation quality ≥ 0.8. As the next step similar approach have been used for analysing two additional cohorts DILGOM (500 individuals) and FINRISK1997 (10,000 individuals) from Finland. All three cohorts were meta-analysed using fixed-effect inverse variance based meta-analysis in GWAMA. We observed evidence of novel POE variants for 7 metabolites - Identified variants are associated with lipids and lipoproteins. Two of the found variants (rs1330350 and rs189194743) located in *TNC* and *PTPRD* genes, respectively, which are associated with signaling molecules, regulating a variety of cellular processes, such as neuronal regeneration, mitotic cycle, cell growth, neurons and axons migration during development. These regulatory features provide normal embryonic growth and therefore it suggests that identified POE may be mediated through genomic imprinting. This idea motivated us to perform further steps, such as allele-specific expression analysis, epigenome-wide association analysis with further replication in an independent cohort and methylation QTL to investigate mechanisms underlying POE and its association with genomic imprinting. We believe that our results will highly contribute to uncovering complex relations between genetic variants and common traits and help to reveal some of the hidden heritability.

1236F

Gene-smoking interactions in 133,802 multi-ethnic subjects identifies novel loci for high-density lipoprotein cholesterol. *A.R. Bentley[1], C.N. Rotimi[1], L.A. Cupples[2], CHARGE Gene-Lifestyle Interactions Working Group.* 1) Center for Research on Genomics and Global Health, NHGRI, National Institutes of Health, Bethesda, MD; 2) Department of Biostatistics, Boston University, Boston, MA.

High-density lipoprotein cholesterol (HDLC) concentration is associated with smoking, but it is unknown to what extent smoking may modify genetic associations with HDLC. We investigated whether accounting for the interaction of current smoking and genetic variants may help identify HDLC loci in 133,802 individuals of European (n=90,266), African (n=23,745), Asian (n=13,171), and Hispanic ancestry (n=6,620), using 1000G imputed data. To maximize statistical power, in addition to testing the interaction with a 1 degree of freedom (df) test, we employed a joint 2 df test that simultaneously evaluates the SNP and interaction effects. Two models of the genetic main effect with and without smoking adjustment were considered so that we could determine what loci identified in the interaction could also be identified in a main effect analysis. Meta-analyses were conducted for each ancestry, and then all groups were meta-analyzed. In multi-ethnic meta-analysis, we identified 3 novel loci with a main effect on HDLC: gene-rich regions on chromosome 15 (e.g. *UBR1*, *TGM7*, *ADAL*, *TUBGCP4*, *TP53BP1*) and chromosome 1 (e.g. *TMED5, CCDC18, DR1)*, and non-coding variants in *PTPRJ*. Three additional novel loci were identified in the ancestry-stratified meta-analyses: rs9935936 (*GNAO1*) and rs143691761 (*CTNNA1*) in European and rs2160394 (*PRKD3*) in Asian ancestry individuals. Using the 2 df interaction test in multi-ethnic meta-analysis, a smoking interaction was observed for a novel locus near *STOML3*. These variants were African ancestry-specific and reached statistical significance in the meta-analysis limited to African ancestry studies. Two additional smoking interactions were observed using the 1 df test: rs10743071 (*STK33*) in African ancestry individuals, and rs17311510 (*TGM*) in European ancestry and multi-ethnic meta-analyses. In addition to these novel results, our meta-analyses confirmed associations for 50 previously-identified lipids loci. Further work is ongoing to replicate these loci and to understand the biological mechanisms underlying these associations which may lead to novel drug targets and/or therapeutic interventions. In summary, large multi-ethnic meta-analyses of a genome-wide interaction study of HDLC and current smoking reveals 9 novel loci across and within ancestry groups.

1237W

Novel rare variants in *GCKR* affect triglyceride concentrations and increase the risk for type 2 diabetes. *D. Sanghera[1], B. Sapkota[1], H. Mussa[2], D. Goins[4], C. Borga[3], P. Whitby[2], C. Sansam[4], K. Frazer[3].* 1) Department of Pediatrics, Section of Genetics, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; 2) Department of Pediatrics, Section of Infectious Diseases, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; 3) Department of Pediatrics, Section of Pediatric Hematology-Oncology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; 4) Cell Cycle & Cancer Biology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA.

Dyslipidemia is a well-established risk factor for cardiovascular disease and a principal cause of mortality in individuals with type 2 diabetes (T2D). Despite the high heritability (50-80%) of lipid traits, previous genome-wide association studies (GWAS) have only been able to account for a fraction of this heritability (<10%) in genes involved in lipid metabolism. In this study, we aim to identify the rare functional variants in known candidate genes for diabetic dyslipidemia. Here, we performed targeted sequencing of 14 confirmed candidate genes (~215 kb) for 940 diabetic dyslipidemia individuals [572 cases with high serum triglycerides (TG) (>150 mg/dl) and 368 controls with low TG (<100 mg/dl)] from the Asian Indians Diabetic Heart Study. Of a total of 2361 high-quality variants analyzed, 953 variants (40%) were unique to high TG cases and 321 variants (13.6%) were unique to controls. Further analysis of variants within the *GCKR* gene using the Combined Multivariate and Collapsing methods revealed clustering of 13 functionally damaging and deleterious rare mutations near Fructose Binding site and Glucokinase (GCK) Binding sites at the sugar isomerase domains. The *GCKR* encodes glucokinase regulatory protein that regulates GCK (a known T2D gene) by forming a complex, which plays a role in the control of blood glucose homeostasis. The lead variant with a missense mutation of Serine/Asparagine was restricted to individuals with high TG. More than 60% of the carriers were diabetic and 90% of carriers had high TG (ranging from 182 mg/dl to 560 mg/dl). However, this variant was absent in Caucasians (n=33,370), Africans (n=5,203), Hispanic/Latinos (5,789), and East Asians (n=4,327) in a large Exome Aggregation Consortium (ExAC) dataset of multiethnic populations. We have designed a transgenic zebrafish (*Danio rerio*) and will be testing the phenotypic effects of these variants to evaluate metabolic consequences *in vivo.* Taken together, our findings have the potential to find novel pathway for diabetes linked with hypertriglyceridemia.

1238T

An insertion-deletion polymorphism within *FADS2* is associated with fatty acid levels and dietary pattern in Chinese Hans. *L. Xu[1], H. Yao[1], H. Shaofeng[1], Y. Kaixiong[2], G. Zhenglong[2], L. Huaixing[1].* 1) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai, People's Republic of China; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York.

A recent study has suggested that different dietary patterns may contribute to individual variability in response to polyunsaturated fatty acid (PUFA) composition and allele frequencies of SNP rs66698963 (an insertion-deletion polymorphism) at *FADS2* locus. The I/I genotype frequency was higher in traditional vegan/vegetarian populations and was associated with 8% higher basal plasma phospholipid arachidonic acid (AA, 20:4n-6) level. In this study, we aimed to examine whether the reported insertion-deletion variant exhibits different allele frequencies and effect sizes on PUFA levels in the Beijing and the Shanghai individuals of Chinese Hans, who are supposed to have different dietary patterns, from the Nutrition and Health of Aging Population in China (NHAPC) study. The insertion-deletion polymorphism rs66698963 was successfully imputed using the 1000 Genome Project Phase 3 data as the reference panel (info=0.85). The frequencies of rs66698963-insertion (0.81 and 0.71 in Beijing and Shanghai, respectively) and its proxies (r²≥0.6) were significantly higher in participants from Beijing ($P \leq 3.07E10^{-15}$) compared with those from Shanghai. Consistent with previous findings, rs66698963 was significantly associated with multiple PUFA levels including 18:2n-6, 18:3n-6, 20:3n-6, 20:4n-6, 22:4n-6, 20:5n-3 and 22:5n-3 ($P \leq 7.95E10^{-8}$). Moreover, rs66698963 exerted significantly larger impact on 20:3n-6 level in Beijing populations ($P_{heterogeneity}=2.28E10^{-7}$) while it exerted larger impact on 18:3n-6 and 20:5n-3 levels in Shanghai populations ($P_{heterogeneity} \leq 0.016$). The deletion allele frequency is higher in Shanghai individuals (D=0.29) who consume more seafood than in Beijing individuals (D=0.19) who consume less seafood. In conclusion, we replicated the previous findings that this genetic variant is associated with fatty acid composition and the deletion allele has higher frequency in individuals who consume more seafood.

1239F

Large-scale meta-analysis of lipids traits in Hispanic population. *L.E. Petty[1], M. Graff[2], X. Guo[3], Y. Hai[3], J. Yao[3], A. Manichaikul[4], X.Q. Wang[4], A.E. Justice[2], L.S. Emery[5], T. Sofer[6], H.M. Highland[2], J.E. Below[1], Hispanic LIpids Consortium.* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 2) Dept. of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) HARBOR-UCLA Medical Center, Torrance, CA; 4) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Department of Biostatistics, University of Washington, Seattle, WA.

To date, genome wide association studies have successfully identified many (>150) loci associated with lipid levels. These studies have primarily been performed in Caucasian samples, yet dyslipidemia, a major risk factor for the development of cardiovascular disease, disproportionately affects Hispanic/Latinos (HL). Therefore, this inordinately affected ancestry group is a target population for additional, large-scale GWAS and holds promise for further discovery, as evidenced by recent, smaller GWAS in HL identifying novel loci and population-specific alleles. In this pursuit, we assembled a HL lipids consortium of >50,000 samples. Here we present preliminary results in a subset including Hispanic Community Health Study / Study of Latinos, Women's Health Initiative, Starr County, and two data sets from Mexico City (N=19,696). Imputation to 1000 Genomes Project phase 3 or phase 1 reference data was completed for each study. Association testing for four lipids phenotypes (total cholesterol, HDL, LDL, and triglycerides) using SNPTEST or MACH was performed in each study adjusting for principal components as necessary, age, sex, and any study-specific covariates. Meta-analysis was performed using METAL. Variants with IMPUTE2 quality scores less than 0.8 or MAC < 5 were excluded from meta-analysis. For total cholesterol, we identified 16 genome-wide significant loci, including 12 loci that map to genes that have not been published before in HL populations. For example, sentinel SNP rs1260326 maps to *GCKR*, which has been previously observed in HL populations. However, another sentinel SNP rs35529421 maps to *DOCK7*, which has been reported in European populations, but never HL. For HDL, LDL, and triglycerides, we identified 14, 12, and 9 significant loci, respectively, of which 8, 9, and 6 map to genes that have not been observed in HL populations. All of these findings not previously seen in HL populations have been published in other ancestry groups. Our future work will include additional HL samples to more than double the sample size. We anticipate that this increase in sample size will likely allow for identification of loci that map to novel genes for each lipid trait. Additionally, we explored linkage disequilibrium patterns around top findings and report genome-wide significant population-specific allelic effects in HLs independent of previously published alleles in European populations.

1240W

The interaction between common markers and menopausal hormone therapy influencing lipid levels. *J. Rich[1], A. Reiner[2,3], C. Kooperberg[3], P. Auer[1].* 1) School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI; 2) Department of Epidemiology, University of Washington, Seattle, WA; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Sex steroid hormones and their receptors are critical determinants of plasma lipid levels. In addition to the effects of estrogen on low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG), plasma lipid concentrations are also highly influenced by genetic variation. In this study, we sought to identify genetic variants that interact with menopausal hormone therapy to influence lipid levels. We imputed low-depth (8x) whole genome sequencing (WGS) data from the UK10K project into a total sample of 11,150 post-menopausal women that participated in the Women's Health Initiative. Exposure to hormone replacement therapy (HRT) was defined as currently taking hormone therapy versus not. We followed a two-stage approach to the gene-by-environment interaction analyses: stage 1 consisted of marginal analyses between all imputed variants with a minor allele frequency > 5%; variants with a p-value of association less than 5×10^{-7} were brought forward to the stage 2 interaction analyses. In stage 2, we tested for interaction between imputed variants and hormone replacement therapy (HRT) on LDL-C, HDL-C, and TG levels for the 2,101 variants brought forward from stage 1. The 2-stage approach identified a common marker (rs375372) that interacts with HRT to influence TG levels ($P=5.2 \times 10^{-4}$). For HRT users, the additive effect of the rs375372 alternate allele increases TG levels by 4.5%. This effect was not observed in individuals that were not taking HRT, suggesting that the effect of the rs375372 alternate allele on TG levels is specific to those that use HRT. The rs375372 variant was located in a putative 5' promoter for the *LIPC* gene, indicated by the H3K4me3 histone mark in adult Liver tissue from the RoadMap Epigenome project. *LIPC* is expressed in the liver, encodes a hepatic triglyceride lipase, and has been previously implicated in hyperlipidemia. The rs375372 variant is not in linkage disequilibrium ($R^2 > 0.6$) with any known variant association for TG. Though gene-by-environment (GxE) interactions may play an important role in explaining the heritability of lipid traits, there are limited examples of robustly associated GxE interactions. These results provide strong evidence for a new locus that interacts with an environmental exposure to influence TG.

1241T

Genome by lipidome analyses in a Singaporean cohort. *M.K. Lin[1,2], H. Begum[1,3], F. Torta[1,2], P. Narayanaswamy[1], P.A. Mundra[3], W.Y. Saw[4,5], E. Tantoso[4], P.F. Little[5], P.J. Meikle[3], Y.Y. Teo[4,5,6,7,8], M.R. Wenk[1,2].* 1) Singapore Lipidomics Incubator, Life Science Institute, National University of Singapore, Singapore; 2) Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 3) Baker IDI Heart and Diabetes Institute, Melbourne, Australia; 4) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 5) Life Science Institute, National University of Singapore, Singapore; 6) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; 7) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore; 8) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

Metabolic diseases, such as obesity and cardiovascular disorders, are associated with perturbations in lipid homeostasis. Clinical manifestations of metabolic diseases can be heterogenous, resulting from an interplay of environmental and genetic factors. The present study is based upon the hypothesis that the relevant clinical 'lipid' phenotype, high/low density lipoprotein and total triglyceride levels, are influenced by precursor molecules which define these clinical lipids, and there are underlying genetic variations influencing precursor lipid molecular levels. A Singaporean pilot cohort comprising of 105 Malay, 102 South Indian, and 109 South Chinese healthy subjects was recruited. These three ethnic groups made up approximately 98% of Singapore's ethnic-diverse population. Average age of the cohort was 51.9 years (SD ±6.0 years), with mean body mass index of 25.5 (SD ±4.4). There was no known metabolic disorders reported by the subjects during sample collection. Seemingly healthy subjects were selected based on the hypothesis that genetic features driving changes at lipid molecular level can act as biomarkers in prevention of metabolic diseases through life-style modification. Large-scale lipid profiling was performed. 283 lipid molecules were measured using targeted liquid-chromatography mass spectrometry in multiple reaction monitoring. The lipidomic data was independently validated at two distinct locations (Singapore and Australia) on different experimental platforms. All individuals were genotyped on Illumina® Omni 2.5 and exome chip. Association analyses utilizing PLINK was carried out for 2.5 million SNPs and variation in lipid molecule levels. Despite the small sample size, significant association signals (*p*-value ≤ 5.0×10^{-8}) after Bonferroni correction were detected for lipid molecules within the top ten coefficient of variation within each ethnic group. SNPs were found significantly associated to lipid classes, triglyceride and phosphatidylcholine. These SNPs are located in plausibly lipid biological pathway-related genes. Results were ethnic-specific. Results will be validated in a larger Singaporean population comprising of healthy controls and patient groups with various metabolic disorders. Implication of the validated results will help uncover metabolic biomarkers specific to Singapore's population, allowing more precise and efficient diagnoses to advise (potential) patients.

1242F

Rare alleles in Europeans are associated with low density lipoprotein cholesterol (LDL-C) and triglycerides (TG) levels in the Hutterites. *C. Igartua[1], D.L. Nicolae[1,2,3], C. Ober[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL; 3) Department of Medicine, University of Chicago, Chicago, IL.

Founder populations offer unique opportunities to study the clinical effects of alleles that are rare in outbred populations but occur at higher frequencies in these isolated populations. Whole genome sequencing in 98 South Dakota Hutterites, a founder population of European descent, revealed 1,037,333 variants that are rare (<1%) or absent in European populations, but occur at frequencies of up to 50% in the Hutterites. We conducted a GWAS of lipid levels with these "rare in European variants" (REVs). Plasma levels of LDL-C, high-density lipoprotein cholesterol (HDL-C), TG and total cholesterol (TC) were measured in 828 adult Hutterites who are related to each other in a 13-generation pedigree with 64 founders. Variants present in the whole genome sequences were previously imputed to these individuals using PRIMAL (Livne et al. *PLoS Comp Biol* 2015; 11:e1004139). We tested for associations with 322,934 REVs with MAF>1% in the Hutterites and imputation call rates >75% in this sample using a linear mixed model (GEMMA), including age and sex as fixed effects and kinship as a random effect. After Bonferroni correction for 4 phenotypes and 322,934 variants, we identified associations between increased LDL-C and 44 variants in a 7.6 Mb region flanking the LDL receptor gene (*LDLR*) on chromosome 19. These variants had minor allele frequencies of 0.03-0.04, and most are private to the Hutterites, including a missense variant in *ZNF439* ($p=5.55 \times 10^{-12}$, OR= 2.6 [2.0,3.4]), and a noncoding RNA in an intron of LOC100652768 ($p=9.84 \times 10^{-9}$, OR= 2.6 [1.9,3.6]). These variants are not in LD with common variants at *LDLR* previously associated with LDL-C (rs688, $r^2=0.02$; $p=2.92 \times 10^{-4}$, OR= 1.2 [1.1,1.3], AF 0.58 in the Hutterites). 12 REVs (MAFs 2.0%-2.4%) on chromosome 11 were associated with reduced TG levels (rs114977169, $p=9.49 \times 10^{-13}$, OR=0.28 [0.2,0.4]). These variants are in LD with a variant predicted to disrupt splicing of the first exon in *APOC3*, and which was previously associated with TG levels in the TwinsUK cohort (rs138326449; MAF=0.0025 in TwinsUK). These results provide examples of rare variants with large effects on LDL-C and TG levels in the Hutterites. The REVs flanking the *LDLR* gene have effects on LCL-C that are independent of known common variation at this locus, either through long-range regulatory effects on the *LDLR* gene itself or due to their effects on other novel genes in this region. This work was supported by R01 HL085197.

1243W

Overlap of lipid GWAS and adipose eQTL loci identifies 18 lipid loci with candidate regulatory variants. *M. Alvarez[1], E. Nikkola[1], C. Raulerson[2], K. Mohlke[2], M. Laakso[3], P. Pajukanta[1,4].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Molecular Biology Institute at UCLA, Los Angeles, USA.

High serum lipids levels are a major risk factor for cardiovascular disease (CVD) that present a major public health problem worldwide. Genome-wide association studies (GWAS) have identified >157 loci significantly associated with lipid levels. Although the molecular mechanisms underlying several of these loci have been characterized, for many loci the causal variants, genes, and functional mechanisms remain to be elucidated. We hypothesize that for some of these loci, associated variants affect transcriptional processes of genes relevant to lipid metabolism. To identify variants and genes for the uncharacterized lipid loci, we overlapped expression quantitative trait loci (eQTL) with the lipid GWAS loci. Using RNA sequencing of subcutaneous adipose tissue and genotype data of 494 individuals from the METabolic Syndrome In Men (METSIM) cohort, we assessed gene expression and allele-specific expression. To identify eQTLs we used the RASQUAL software and found 5,109 genes with a lead eQTL SNP passing a genome-wide Bonferroni threshold (p-value$<3.2 \times 10^{-6}$). To overlap eQTL and lipid-GWAS loci, we required the lead eQTL and the lead GWAS SNP to be in linkage disequilibrium (LD) ($r^2 \geq 0.75$), and observed 14, 4, 6, and 2 of HDL-C, LDL-C, total cholesterol (TC) and triglyceride (TG) loci overlapped with an eQTL, respectively. A combined total of 18 (11.5%) lipid GWAS loci overlapped with a significant eQTL. To further investigate causal regulatory SNPs, we annotated variant proxies in LD with both the eQTL and GWAS lead SNPs ($r^2 \geq 0.75$) using ENCODE and Roadmap Epigenomics data. Among the signals, we followed up on the HDL-C/TC *LILRA3* locus because of its lack of characterization and tight LD between the lead eQTL and TC GWAS SNP ($r^2=0.99$). We identified rs383925 as a likely causal variant since it lies in a region with promoter and enhancer histone marks in several cell and tissue types, including blood and adipose. The variant rs383925 is also predicted to alter the regulatory motif of PU.1, a transcription factor that may be involved in activation of macrophages. In addition, this SNP is a strong eQTL for *LILRA3* in subcutaneous adipose and blood tissue in GTEx, providing additional evidence to its association with *LILRA3* expression. As *LILRA3* encodes a soluble receptor for class I antigens and is predominantly expressed in blood, we predict this GWAS signal to act in a monocyte/macrophage specific manner in obesogenic adipose tissue to affect HDL-C and TC levels.

1244T

Evidence for modification of genetic risk in autoimmunity by exposure to pregnancy using case-only study design. *C. Adams[1], G.I. Cruz[1], X. Shao[1], H. Quach[1], K. Taylor[2,3], J. Nititham[2,3], C. Schaefer[4,5], T. Olsson[6], I. Kockum[6], L. Alfredsson[7,8], L.A. Criswell[2,3], L.F. Barcellos[1,4,5].* 1) GEGL Lab, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, San Francisco, CA; 4) Kaiser Permanente Division of Research, Oakland, CA; 5) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA; 6) Department of Clinical Neuroscience and Center for Molecular Medicine, Karolinska Institutet at Karolinska University Hospital, Stockholm, Sweden; 7) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 8) Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm.

Multiple sclerosis (MS) [MIM 126200], Rheumatoid arthritis (RA) [MIM 604302], Systemic lupus erythematosus (SLE) [MIM 152700], and Sjögren's syndrome (SS) [MIM 270150] are autoimmune diseases (AD) that predominately affect women. Pregnancy can affect the activity of AD and may also contribute to risk of AD onset later in life. Both genetic and environmental factors are believed to contribute to AD. Genes within the major histocompatibility complex (MHC) contribute substantially to the risk of AD. Genome-wide association studies (GWAS) have identified additional loci. However, whether genetic risk of developing AD is modified by pregnancy is unknown. We utilized a powerful case-only GxE study design to investigate whether exposure to pregnancy was associated with genotype for established genetic AD risk factors in MS, RA, SLE, and SS cases. All cases were adult females with confirmed disease selected from the Mother-Child Immunogenetic Study (RA n=1000; SLE n=2500) at University of California, San Francisco, the Sjögren's International Collaborative Clinical Alliance registry (SS n=3500), Kaiser Permanente Northern California (KPNC; MS n=900), and a Swedish cohort (MS n=2000). Comprehensive reproductive history, genetic ancestry, and other potential confounder and covariate data were collected for all cases. Several pregnancy exposure models comprised of different groupings of disease onset before and after first live birth pregnancy and no live birth pregnancy were tested. Logistic regression was used to test for association between pregnancy exposure and genotype for non-HLA disease risk GWAS variants and the MHC variants adjusting for genetic ancestry and confounders. Genotype data for all cases were obtained through whole-genome or ImmunoChip profiling and imputation. Results for MS, to date, demonstrate evidence for a significant association between disease onset after first live birth pregnancy and *PVT1* genotype (OR=1.53; 95% CI, 1.22-1.92; $P_{FDR-adj}$<0.05). Specifically, among KPNC MS cases, the odds of onset of MS after the first live birth pregnancy were 1.53 times the odds of onset of MS before a live birth pregnancy for each additional rs4410871 T allele. No evidence for association was found for SNPs in the MHC. *PVT1* encodes a long, non-coding mRNA that has been previously identified as a non-HLA MS risk GWAS risk locus. This study is the first to characterize the association between exposure to pregnancy and AD genetic risk factors.

1245F

Large-scale imputation of KIR genes in psoriasis case-control cohorts. *R.S. Ahn[1], D. Vukcevic[2,3,4], A. Motyer[2,3], D. Ellinghaus[5], L.C. Tsoi[6], R.P. Nair[6], J. Oksenberg[7], C. Palmer[8], J. Foerster[9], J.T. Elder[6], A. Franke[5], S. Leslie[2,3,4], W. Liao[1].* 1) Dept. of Dermatology, University of California, San Francisco, San Francisco, CA; 2) School of Mathematics and Statistics, University of Melbourne, Parkville, Australia; 3) School of BioSciences, University of Melbourne, Parkville, Australia; 4) Data Science, Murdoch Childrens Research Institute, Parkville, Australia; 5) Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany; 6) Dept. of Dermatology, University of Michigan, Ann Arbor, MI; 7) Dept. of Neurology, University of California, San Francisco, San Francisco, CA; 8) Division of Cardiovascular & Diabetes Medicine, University of Dundee, Dundee, United Kingdom; 9) College of Medicine, Dentistry, and Nursing, University of Dundee, Dundee, United Kingdom.

Killer cell immunoglobulin-like receptors (KIR) are expressed on the surface of natural killer cells and regulate innate immune response by interacting with HLA molecules expressed on the surface of antigen presenting cells. KIRs are encoded by 14 genes that are located on chromosome 19. In the autoimmune skin disorder, psoriasis, previous studies have shown that copy number of several KIR genes, including *KIR2DS1*, *KIR2DS4*, and *KIR3DL1*, are associated with psoriasis. However, these studies had limited power to detect an association because of their small sample sizes. To overcome the time and expense of directly genotyping KIR genes, we implemented a state-of-the-art method, KIR*IMP, to impute KIR copy number from SNPs on chromosome 19 from a cohort from UCSF (n = 1261) as well as US, Canadian, German, and Estonian cohorts from the PAGE consortium (n = 9482). All samples were genotyped on either the Affymetrix Axiom UK Biobank chip (UCSF) or the Illumina Immunochip (PAGE consortium). Univariate logistic regression was implemented to test for association between imputed KIR copy number and psoriasis in each cohort followed by a pooled association test. Pooled association testing revealed that *KIR2DS4* Deletion variant and *KIR3DL1* exon 4 were associated with psoriasis under a dominant model (p < 0.05). We also found suggestive evidence (p < 0.1) of association between psoriasis and *KIR2DS4* and *KIR3DL1* exon 9 under a dominant model and for *KIR2DS4* deletion variant and *KIR3DS1* under an additive model. This is the first large-scale study of imputed KIR copy number in psoriasis. We discovered that several previously reported KIR gene associations are replicated when several cohorts are pooled together. Our study suggests that future studies of typed and imputed KIR copy number in psoriasis and other autoimmune diseases are warranted.

1246W

Novel SLE risk variant at non-coding RNA (SMG7-AS1) modulating the expression of SMG7. *F. Alasti[1], C. Sun[1], B. Singh[1], P. Borden[2], P. Raj[3], J. Guthridge[1], J. James[1], P.M. Gaffney[1], J. Harley[4], E. Wakeland[2], L. Looger[2], M. Alaecon[1], S.K. Nath[1].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City (OMRF), OK, USA; 2) Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, USA; 3) Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX, USA; 4) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

Introduction: Systemic lupus erythematosus (SLE) is a progressive complex autoimmune disease characterized by the production of a range of auto-antibodies. SLE is influenced by both genetic and environmental factors; however, the exact etiopathology of SLE is not clearly understood. Several studies have identified multiple susceptibility loci for SLE, including *NCF2*. We recently reported multiple SLE risk variants at this gene. To identify additional novel variants within this region, we conducted a follow-up study in ethnically diverse populations. We subsequently found novel SNPs at neighboring genes (like *SMG7*). We hypothesize that these SNPs can modulate the expression of *SMG7*. **Methods:** The confirmation of our SNPs was done by genotyping and sequencing followed by imputation-based analysis and conditional analysis. We assessed allele specific functions using luciferase promoter assay, quantitative real-time PCR and Western blot analysis. **Results:** We discovered a novel SNP located at the non-coding part of anti-sense RNA (*SMG7-AS1*) consistently associated with individuals with European and Hispanic ancestry (OR = 0.79, 95% CI: 2.58 [2.32-2.94], P_{meta} = 1= 4.05 x 10 -61). In order to figure out if any of our associated SNPs can affect the expression of *SMG7*, we performed dual Luciferase assays in which the best candidate SNP showed strong effect on the promoter activity, to be significantly decreased by switching to the risk alleles. To examine whether the RNA expression levels of *SMG7* gene is affected by our associated SNP, we assessed a qRT-PCR on 7 individuals homozygous for the risk allele versus >10 homozygous for the protective allele. Our data showed that the RNA expression levels of *SMG7* in homozygous risk group compared to the protective group is significantly lower (P< 0.01). In addition, our protein Western blot analysis also supported our RNA quantitative results. The novel imputed SNP was confirmed by Taqman genotyping and sequencing. It is known that SMG7 has a major role at NMD pathway. To follow up our current study, we need to find out if our SNP is affecting the NMD pathway and also show how it is causing the SLE susceptibility. **Conclusion:** This study identified a novel functional SNP at SMG7-AS1 and confirmed its regulatory roles in SMG7 expression by affecting the promoter activity of *SMG7* gene. Decreased expression level of *SMG7* with the risk alleles versus the protective alleles may contribute to SLE susceptibity.

1247T

Joint pathway analysis and histone mark enrichment unravels hidden biological processes underlying celiac disease pathogenesis. *L. Bossini-Castillo[1], P. Carbonetto[2], C. Wijmenga[3], M. Stephens[4], G. Trynka[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Ancestry.com DNA, LLC, San Francisco, CA; 3) Department of Genetics, University of Groningen, University Medical Center, Groningen, The Netherlands; 4) Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America.

Celiac disease (CeD) is a complex autoimmune disorder of the gut triggered by gluten ingestion. CeD affects close to 3% of the European population. Up to 45 CeD genetic risk factors have been identified to date. Here, we show that a combined analysis of pathways and cell-specific regulatory annotations in a previously published dataset of 3,149 CeD cases and 6,325 controls (Dubois 2010 *Nat Genet*) yields novel insights into disease-relevant biological processes, and implicates disease-causal cell types. In this study, we assessed enrichment of SNPs annotated to 3,159 candidate pathways from 8 databases. The applied pathway enrichment method builds on multivariate regression to model disease risk as the combined additive effect of multiple markers (Carbonetto 2013 *PLOS Genet*). We identified five pathways with strong support for enrichment of CeD risk factors, including IL-12 signaling, expression of chemokine receptors, cytokine-cytokine receptor, E-cadherin adherens junction and Th1/Th2 differentiation. While the role of the immune pathways has been previously explored in CeD, the implication of the E-cadherin adherens junction pathway, which regulates adhesion of endothelial cells and epithelial inflammation, was novel. The enriched pathways replicated in an independent US cohort comprising 1,596 cases and 2,965 controls, with an inflation factor (lambda) of 1.13 - 1.64 for the pathway assigned SNPs. One advantage of taking a modeling approach to interrogate enriched pathways is that it can be used to inform which SNPs are associated with the phenotype. Using SNPs with posterior probability > 0.01, assigned to the significant pathways, we tested for enrichment with active promoters (using the H3K4me3 mark) in a set of 118 tissues. We observed that associations informed by different pathways pointed to distinct cell types, and these cell types were not detected from the association signal under the null hypothesis of no pathway enrichment. We observed that SNPs from E-cadherin adherens junction pathway were enriched in rectal and stomach smooth muscle, psoas muscle and different T cells, further supporting the role of adhesion of endothelial cells and epithelial inflammation in CeD pathogenesis. In summary, by analysing the accumulation of genetic variants in gene pathways jointly with cell specific regulatory annotations we were able to gain valuable novel insights into disease mechanisms and implicate relevant cell types for a complex trait.

1248F

Replication of ImmunoChip array and genome-wide linkage identified variants of Kawasaki disease in Taiwanese children. *T. Chang[1], M. Chen[2], Y. Lee[1,2], W. Chen[1], N. Chiu[2].* 1) Med Res Dept, Mackay Memorial Hospital, New Taipei, Taiwan; 2) Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan.

Genetic epidemiological studies show that genetic factors contribute significantly to Kawasaki disease pathogenesis. However, genetic studies are often inconsistent and the associations are often race or ethnicity specific. The aim of this study is to replicate 4 novel single nucleotide polymorphisms (SNPs) identified by either fine-mapping ImmunoChip array or linkage analysis, which were associated with Kawasaki disease in Caucasians. The rs7320375 *A/G*, rs4021 *A/G*, rs392775 *A/T*, and rs9847915 *A/T* SNPs were genotyped in 690 Taiwanese children with Kawasaki disease and 562 ethnically matched healthy controls by using the Pre-Developed TaqMan Allelic Discrimination Assay. In our replication study, no significant associations between these 4 SNPs and susceptibility of KD were found. In addition, no evidence for associations between these SNPs and development of coronary artery lesions (CALs) was found. In summary, our findings provide little support to validate these fine-mapping discovered novel SNPs in the susceptibility or CAL development of KD in Taiwanese children.

1249W

Low-frequency coding variants explain a fraction of multiple sclerosis heritability and identify four new susceptibility genes. *C. Cotsapas, International Multiple Sclerosis Genetics Consortium* . Neurology, Yale School of Medicine, New Haven, CT.

Multiple sclerosis (MS) is an autoimmune disease in which the myelin sheath surrounding brain neurons is destroyed by the immune system, leading to progressive physical and cognitive morbidity and shortened life expectancy. Over the last decade we have conducted a series of genome-wide association studies, discovering 200 common variant risk associations explaining over 50% of the heritability. To complement these efforts we have genotyped 250,000 low-frequency non-synonymous (NS) coding variants across all exons. Following stringent quality control we meta-analyzed 29,746 cases and 35,066 controls in 14 country-level strata, using linear mixed models to control for population stratification. We find seven low-frequency NS variants (MAF < 0.05) show convincing evidence of association ($p < 6.25 \times 10^{-7}$, Bonferroni-adjusted $p < 0.05$ for number of variants tested). These variants are located in six genes, including four outside risk loci identified by common variant GWAS. These four genes – PRF1, HDAC7, NLRP8 and PRKRA – identify new immune functions mediating risk, and implicate both variation in genome regulation and in heterologous cell killing in pathogenesis. The PRF1 A91V variant (overall MAF = 0.049, association $p = 3.9 \times 10^{-11}$) is a hypomorph, with CD8+ cytotoxic T cells from heterozygous donors showing 30% lower efficiency in effector T cell killing than donors homozygous for the major allele (PMID 25776844). We next assessed if more effects are present in our data and to estimate the heritability attributable to low-frequency variation in MS, using polygenic risk scores. We find significant evidence of polygenicity with low frequency (MAF < 0.01) NS variation, and estimate that approximately 5% of disease heritability is attributable to this fraction of the allelic spectrum. Our effect size estimates are in line with those for common variants, and we do not observe any odds ratios >1.5. Thus, we conclude that low frequency variation explains a small but significant fraction of MS heritability, and can identify new pathogenic genes that may otherwise not have been found. However, our results also suggest that assessing the contribution of low frequency variation either by genotyping or sequencing will require large sample sizes due to modest effect size. .

1250T

In sickness and in health? Mate's genes and risk of female autoimmunity in a study of 6,200 individuals. *G.I. Cruz[1], X. Shao[2], C.J. Adams[1], H. Quach[2], K. Ho[3], K. Sterba[3], J.A. Noble[4], N.A. Patsopoulos[5], M.P. Busch[6], D.J. Triulzi[7], W.S.W. Wong[8], B.D. Solomon[8], J.E. Niederhuber[8], L.A. Criswell[3], L.F. Barcellos[1,2].* 1) School of Public Health, University of California Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Lab, California Institute for Quantitative Biosciences (QB3), University of California Berkeley, Berkeley, CA; 3) Rosalind Russell / Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California San Francisco, San Francisco, CA; 4) Children's Hospital Oakland Research Institute, Oakland, CA; 5) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham & Women's Hospital, Boston, MA; Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston,; 6) Blood Systems Research Institute, San Francisco, CA; 7) Institute of Transfusion Medicine and Department of Pathology, University of Pittsburgh, Pittsburgh, PA; 8) Division of Medical Genomics, Inova Translational Medicine Institute, Falls Church, VA.

The causes of sexual dimorphism observed in autoimmune disorders (AD) are not entirely understood. Women are disproportionately affected, raising interest in female-specific factors as possible mechanisms. Pregnancy involves immunological changes in the mother. It is well known that pregnancy affects activity of pre-existing rheumatoid arthritis (RA) [MIM 604302] and systemic lupus erythematosus (SLE) [MIM 152700]. It is possible exposure to paternal antigens through the fetus could influence the development of AD in women without the disease. As not all pregnancies result in a live birth, we investigated if mating partners who carry RA- and SLE-associated risk alleles increase the risk of disease of the mother. We investigated the association between a mating partner's human leukocyte antigen (HLA) genotype and maternal AD in 155 RA and 151 SLE cases and 895 controls from the Mother-Child Immunogenetic Study (MCIS). Female cases, their children and their children's fathers were recruited at UC San Francisco; controls and their families were recruited from the Blood Centers of the Pacific, the Institute for Transfusion Medicine at the University of Pittsburgh, and from studies at the Inova Translational Medicine Institute (ITMI); a population-based sample of 5000 males matched on genetic ancestry were included for comparison. HLA alleles were imputed from genotype and whole genome sequencing data using SNP2HLA. We categorized mating partners' *DRB1* genotype as carriers of an allele or group of alleles of interest (AOI) for each disease. RA AOI tested include the *DRB1* "shared epitope" (SE) and alleles coding for high-risk amino acids (AA) at positions 11 (valine), 71 (lysine), and 74 (alanine); SLE AOI *DRB1*03:01*, **15:01*, **08:01*. We used logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the association between RA and SLE separately and a mating partner's *HLA* risk alleles. In addition, we estimated each couple's relatedness coefficient, *R,* to investigate non-random mating effects in AD. We found increased risk of RA associated with SE carrier status (OR 1.5; 1.1-2.1) and AA 71 (OR 1.4; 95% CI, 1.0-2.0); however, AA 74 was associated with reduced risk (OR 0.6; 95% CI, 0.4-0.8). We did not find any significant associations between AOIs investigated and SLE. These findings support the hypothesis that a mating partner's genotype may contribute to risk of some female AD, possibly through exposure to pregnancy.

1251F

Three novel susceptibility loci were identified through a meta-analysis of immunochip data from Korean and East Asian patients with inflammatory bowel disease. *M. Hong[1], H. Oh[1], S. Jung[1], S.J. Ahn[1], B.D. Ye[2], S.H. Park[2], S.K. Yang[2], K. Song[1].* 1) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea; 2) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine.

Most Asian GWAS on IBD genetics analyzed CD and UC separately, giving insights into the genetic architecture of each single disease, but not a broad view of the genetic basis of IBD in Asians. A recent study on IBD in Japanese successfully confirmed associations at several known loci, but did not find new loci possibly due to the limited sample size. We previously reported association analyses for CD and UC separately using immunochip data. To identify additional IBD susceptibility loci in Asian, we first performed immunochip analysis of IBD in Koreans (1332 cases, 1185 controls) followed by a meta-analysis combining Korean and East Asian IBD immunochip data sets (2824 cases, 3719 controls). In the Korean IBD Immunochip analysis, we did not find any novel loci with genome-wide significance. In a meta-analysis using immunochip data of Korean and East Asian IBD, SNPs with association $P < 1 \times 10^{-3}$ in the discovery cohort were confirmed in the Korean IBD GWAS (730 cases, 3739 controls) data set. Three novel loci reached genome-wide significance ($0.05/80,279 = 6.23 \times 10^{-7}$) in the discovery and GWAS replication cohorts combined including rs2072711 in *NCF4-CSF2RB* (OR = 1.20, P_combined = 3.34×10^{-10}), rs12928665 near *CIITA* (OR = 0.86, P_combined = 1.31×10^{-8}), and rs3138056 near *NFKBIA* (OR = 0.87, P_combined = 6.12×10^{-7}). Our results provide new biological insight to IBD in East Asians.

1252W

Genome-wide association study identifies *PRKCB* as a novel genetic factor for primary biliary cirrhosis in the Japanese population. *M. Kawashima[1,2], Y. Hitomi[2], N. Nishida[2,3], K. Kojima[4], Y. Kawai[4], H. Nakamura[5], M. Yasunami[6], K. Chayama[7], Y. Aiba[5], M. Nagasaki[4], K. Tokunaga[2], M. Nakamura[5,8], PBC Consortium in Japan (PBCCSJ).* 1) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo; 3) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine; 4) Division of Biomedical Information Analysis, Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University; 5) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center; 6) Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University; 7) Department of Gastroenterology and Metabolism, Applied Life Sciences, Institute of Biomedical & Health Sciences, Hiroshima University; 8) Department of Hepatology, Graduate School of Biomedical Sciences, Nagasaki University.

Primary biliary cirrhosis (PBC) is a chronic and progressive cholestatic liver disease, presumably caused by autoimmune reactions against biliary epithelial cells leading to liver cirrhosis and hepatic failure. PBC primarily affects women (female-to-male ratio, 7:1) between 50 and 60 years of age. The high concordance rate in monozygotic twins compared to dizygotic twins indicates strong genetic factors are involved in predisposing to PBC. ;Previously, we performed a genome-wide association study (GWAS) in 963 Japanese individuals (487 PBC cases and 476 healthy controls) and identified two significant susceptibility loci, *TNFSF15* (rs4979462) and *POU2AF1* (rs4938534). In this study, we performed extended GWAS in an additional 1,923 Japanese individuals (894 PBC cases and 1,029 healthy controls), and combined the results with the previous GWAS data. This GWAS, together with a subsequent replication study in an independent set of 7,024 Japanese individuals (512 PBC cases and 6,512 healthy controls), identified *PRKCB* (rs7404928) as a novel susceptibility gene for PBC (odds ratio [OR] = 1.26, P = 4.13 × 10^{-9}). Furthermore, primary functional variants of *PRKCB* (rs35015313) were identified by genotype imputation using a phased panel of 1,070 Japanese individuals from a prospective, general population cohort study by the Tohoku Medical Megabank Organization, and subsequent in vitro functional analyses. These results may lead to improved understanding of the disease pathways involved in PBC, forming a basis for prevention of PBC and development of novel therapeutics.

1253T

Confirmation of 5 novel susceptibility loci for Systemic Lupus Erythematosus (SLE) and integrated network analysis with 75 SLE loci. *J.E. Molineros[1], C. Sun[1], W. Yang[2], S.C. Bae[3], Y. Okada[4,5], N. Shen[6], X.J. Zhou[7], S.K. Nath[1].* 1) Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Paediatrics and Adolescent Medicine, Scientific Staff, Centre for Genomic Sciences,LKS Faculty of Medicine, The University of Hong Kong; 3) Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea; 4) Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan; 5) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Shanghai Institute of Rheumatology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; 7) Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China and Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry o.

Introduction: We recently identified 10 novel SLE susceptibility loci in Asian populations, and uncovered some novel loci that required further replication. In this follow-up study, we attempted to replicate 5 loci, and perform a joint analysis together with all previously known SLE loci to identify gene sub-networks among these loci. *Objectives*: The objectives of this study are to (a) assess the replication of five previously identified suggestive loci, followed by meta-analysis in 7 Asian and European cohorts; (b) identify how these candidate loci fit within the context of the current set of SLE loci. *Data and Methods*: We genotyped 5 SNPs from 5 loci in Hong Kong (1,456 cases and 955 controls). We augmented this data with our previously reported Asian data and published data from an European cohort, and performed a meta-analysis (9,999 cases and 27,410 controls). *Results*: All 5 loci passed the genome-wide significance threshold: *MYNN* (rs10936599 P$_{meta}$=5.14x10^{-14}), *ATG16L2* (rs11235604, P$_{meta}$=1.31x10^{-9}), *CCL22* (rs223881, P$_{meta}$=2.81x10^{-14}), *ANKS1A* (rs2762340, P$_{meta}$=1.96x10^{-11}) and *RNASEH2C* (rs1308020, P$_{meta}$=7.32x10^{-18}). Functional analysis indicated that 3 non-coding loci (rs223881, rs2762340 and rs1308020) are eQTLs targeted for multiple genes. Additionally, all these loci co-located with enhancer/promoter marks, indicating their possible roles in gene regulation. With these 5 novel loci, the total of SLE susceptibility loci increased to 75. These loci share pathway memberships, gene ontology, and protein domains that suggest common modes of action among SLE genes. All 5 loci (together with known SLE loci) had an over-representation of *E2F1* TFBS (P=3.5x10^{-5}), which implicated them in the regulation of p53 dependent/independent apoptosis. We also identified that 35 SLE loci (including *CCL22*, *MYNN* and *ATG16L2*) were directly implicated in differential expression (P<5x10^{-3}) of nearby genes (1 Mb boundary). Conclusion: We confirmed 5 novel SLE susceptibility loci that share common functional traits with many others currently established loci. We identified functional variants that directly affect gene expression, disrupt binding of key transcription factors, and share functional characteristics with known SLE loci. With this evidence, we not only increased the number of SLE loci, but also expanded knowledge of SLE pathobiology.

1254F

Exome sequencing followed by genotyping identifies a rare variant in *NLRP7* **associated with inflammatory bowel disease.** *A. Onoufriadis[1], A. Amar[1], Y. Omar[1], K. Stone[1], K. Taylor[1], M.A. Simpson[1], J. Sanderson[2], N.J. Prescott[1], C.G. Mathew[1].* 1) Medical and Molecular Genetics, King's College London, London, United Kingdom; 2) Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom.

Genome-wide association scans (GWAS) have now identified more than 200 genomic regions that influence susceptibility to the two forms of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC). Despite these successes, a substantial fraction of IBD heritability remains unaccounted for. This is likely to reflect an intrinsic limitation of GWASs, which are designed to identify common variants of small effect and are unlikely to detect associations with rare susceptibility alleles. Given that disorders segregating in pedigrees are typically caused by rare and highly penetrant mutations, we sequenced the exomes of 29 affected individuals across 13 families with multiple affected individuals. We followed a stepwise filtering process of the exome variant profiles generated, retaining all heterozygous co-segregating variants that were protein altering with a MAF<1% in either the 1000 genomes data or 1000 non-IBD individuals sequenced in-house. Based on the notion that genes carrying common disease associated variants may also harbour rare susceptibility alleles, we focused on 294 key genes that lie on 162 GWAS hits and were prioritised via gene ontology and pathway analyses (Jostins et al 2012). This analysis revealed two families with two different rare missense variants in the *NLRP7* gene co-segregating with disease, p.R801H and p.S361L. We followed up these variants (rs143169084 and rs140797839) by genotyping them in 6293 IBD cases and 7239 controls. Whilst we observed both variants at a higher frequency in IBD cases compared to controls, only p.S361L (rs143169084) was significantly associated with IBD (p=0.02822), whereas the rarer variant p.R801H (rs140797839, 0.014% controls) failed to achieve significance. The NLRP7 protein is a component of the inflammasome, which leads to induction of caspase-1 and IL-1B secretion during the innate immune response.

1255W

Molecular characterization, pathogenesis and tissue susceptibility of human systemic lupus erythematosus (SLE) by expression and eQTL analysis. *N.I. Panousis[1,2,3], G. Bertsias[4], I. Gergiannaki[4], M. Tektonidou[5], M. Trachana[6], A. Banos[7], A. Fanouriakis[8], C. Pamfil[9], E.T. Dermitzakis[1,2,3], D. Boumpas[7,8].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva (iG3), University of Geneva Medical School, Geneva, Switzerland; 3) Swiss Institute of Bioinformatics, Geneva, Switzerland; 4) Department of Rheumatology, University of Crete, Faculty of Medicine, Crete, Greece; 5) Department of Rheumatology, University of Athens, Faculty of Medicine, Athens, Greece; 6) Department of Pediatrics, Aristotle University of Thessaloniki, Faculty of Medicine, Thessaloniki, Greece; 7) Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 8) Department of Rheumatology, Attikon Hospital, Athens, Greece; 9) Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania.

Systemic lupus erythematosus (SLE) is a complex autoimmune disease characterized by remarkable immunological and clinical heterogeneity. Assessing the molecular and genetic basis of this variability is of great importance to understand the risk stratification of patients and applications of targeted therapies. By studying patterns of gene expression and genetic variation in whole blood we can develop a functional understanding of how genetic variants can alter risk of SLE and its complications. We report RNA-seq analysis of 142 genotyped and deeply phenotyped with varying levels of disease activity/severity- SLE patients and 48 control matched individuals. We found 4650 (1% FDR) differentially expressed genes (DEGs) between unaffected and SLE individuals with several molecular pathways being associated with predisposition to SLE and disease severity. By exploring DEGs within sub-categories of SLE activity and healthy individuals we defined different modules of genes associated with the pathogenesis and the disease progression. We explored organ involvement in disease progression/severity. When comparing patients with Active Lupus Nephritis (LN) with inactive LN we identified 1070 (1%FDR) DEGs with pathways such as interferon signaling, cell cycle regulation and oxidative phosphorylation to be over-expressed in active LN. PCA analysis differentiated active LN from inactive demonstrating robust differences in expression and indicating that LN is an extreme phenotype among SLE. By exploring the sub-groups of individuals according to their organ activity (i.e serological, cardiorespiratory, vasculitis) we identified genes and pathways that predispose to the specific organ affected. We further explored genetic regulation of blood expression, in the context of variants associated with gene expression in SLE. We mapped 3178 (5% FDR) genetic variants related with differences in gene expression (eQTLs). We further mapped 37 SLE-specific eQTLs, the majority of them being exome variants providing novel insights into genes and genetic variation contributing to SLE pathogenesis. Enrichment of eQTLs from the GTEx consortium and SLE GWASs revealed several loci where the eQTL and GWAS were tagging the same functional variant across different tissues, thus providing insights to the most relevant tissues implicated in the disease. All the results above provide a unique framework for SLE patient stratification, analysis of which is in progress.

1256T

Novel genetic associations of autophagy gene *MAP1LC3B* and *ARMBA1* in systemic lupus erythematosus. *Y.Y. Qi[1], S.K. Nath[2], C.L. Sun[2], Y.M. Zhang[1], F.J. Cheng[1,3], P. Hou[1], R. Mu[4], C. Li[4], J.P. Guo[4], Z.G. Li[4], F. Yu[1], M.H. Zhao[1], X.J. Zhou[1], H. Zhang[1].* 1) Renal Division, Peking University First Hospital;Peking University Institute of Nephrology; Key Laboratory of Renal Disease, Ministry of Health of China; Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Ed; 2) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA; 3) Shandong Provincial Hospital Affiliated to Shandong University, Jinan, Shandong Province, People's Republic of China; 4) Department of Rheumatology and Immunology, Peking University People's Hospital, Beijing, China.

Objective Emerging evidences from genetic, cell biology and model animal studies suggested a pivotal role of autophagy in mediating the occurrence and development of SLE. This study analyzed variations in autophagy genes for association with SLE with Chinese ancestry. **Methods** In the first stage, 342 variants from 26 autophagy genes were used to perform genetic association analysis of a GWAS dataset using ImmunoChip in 1000 Chinese individuals. The second stage involved genetic replications in additional 4824 Chinese individuals and functional annotations of associated variants. Functional annotations of the associated variants were further examined by the *in silico* method as well as by expression quantitative trait loci (eQTL) analysis with mRNA. Luciferase reporter assay and electrophoretic mobility shift assay (EMSA) were performed to confirm the functional effects in HEK293T and Jurkat cell lines. **Results** By the two-stage analysis, *MAP1LC3B* rs933717 ($P_{discovery}=7.57*10^{-3}$, $P_{replication}=5.76*10^{-5}$ and $P_{combination}=2.35*10^{-6}$) and *ARMBA1* rs12574250 ($P_{discovery}=2.39*10^{-2}$, $P_{replication}=1.62*10^{-4}$ and $P_{combination}=8.61*10^{-6}$) were observed to be consistently associated with SLE. Risk genotypes of rs933717 were correlated with higher *MAP1LC3B* transcription levels in eQTL analysis (rs933717CC vs. TT $p=0.012$; rs933717CT vs. TT $p=0.003$). And higher *MAP1LC3B* mRNA expression was observed in the peripheral blood mononuclear cells of SLE ($p=1.28*10^{-9}$). Increased transcriptional activity of rs933737C (risk allele) by luciferase reporter assay was confirmed in two cell lines ($p=0.014$ in HEK293T, $p=0.022$ in Jurkat). rs933717C (risk allele) also showed greater affinity for HEK293T and Jurkat nuclear protein-DNA complex by EMSA directly. **Conclusion** We observed novel loci implicated in autophagy pathway associated with susceptibility to SLE, which might shed new insight into etiological basis of autophagy in SLE.

1257F

Decreased severity of experimental autoimmune arthritis in peptidylarginine deiminase type 4 knock-out mice. *A. Suzuki[1], Y. Kochi[1], H. Shoda[2], K. Fujio[2], K. Yamamoto[1,2].* 1) IMS, RIKEN, Yokohama City, Japan; 2) Division of Allergy and Rheumatology, University of Tokyo.

Objective. Previously, peptidylarginine deiminase type 4 (PADI4) was identified as a susceptibility gene for Rheumatoid arthritis (RA) by genome-wide association studies. PADI4 is highly expressed in bone marrow, macrophages, neutrophils, and monocytes. Peptidyl citrulline is an interesting molecule in RA because it is a target antigen of anti-citrullinated peptide antibodies (ACPAs), and only PADs (translated protein from PADI genes) can provide peptidyl citrulline via modification of protein substrates. The aim of this study was to evaluate the importance of the PADI4 gene in the progression of RA. *Methods.* We generated Padi4 knockout (Padi4−/−) DBA1J mice. Padi4−/− DBA1J and wild-type mice were immunized with bovine type II collagen (CII) to develop collagen-induced arthritis (CIA). Expression of various inflammatory cytokines and Padi genes in immune cells was detected by real-time TaqMan assay. Cytokine concentration in sera was measured by enzyme-linked immunosorbent assay. Localization of PAD4 and PAD2 protein was indicated by immunohistochemistry. *Results.* We demonstrated that the clinical disease score was significantly decreased in Padi4−/− mice and Padi4 expression was induced by CII immunization. In Padi4−/− mice sera, serum anti-type II collagen (CII) IgM, IgG, and inflammatory cytokine levels were also significantly decreased compared with those in wild-type mice sera. Interestingly, Padi2 expression was compensationally induced in CD11b+ cells of Padi4-/- mice. *Conclusion.* On the basis of these studies, it appears that Padi4 enhances collagen-initiated inflammatory responses. Our results revealed that PAD4 affected on expression of various cytokines and also controlled Padi genes.

1258W

Rheumatoid arthritis antibody subphenotypes are associated with unique HLA-DRB1 residues at position 11. *C. Terao[1,2,3,10], B. Brynedal[4,10], Z. Chen[4], X. Jiang[4], H. Westerlind[4], M. Hansson[4], P.J. Jakobsson[4], K. Skriner[5], G. Serre[6], J. Rönnelid[7], L. Padyukov[4], P. Gregersen[8], J. Worthington[9], L. Alfredsson[4], L. Klareskog[4,11], S. Raychaudhuri[1,2,3,9,11].* 1) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 4) Unit of Rheumatology, Department of Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; 5) Charité Universitätsmedizin Berlin, Humboldt University of Berlin, Germany; 6) UDEAR, Université de Toulouse, Inserm, UPS, Toulouse, France; 7) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 8) Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Manhasset, NY; 9) Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, Oxford Road, Manchester, UK; 10) Equal first authors; 11) Equal senior authors.

Rheumatoid arthritis (RA) is a chronic autoimmune joint-destructive disease. Multiple anti-citrullinated peptide antibodies (ACPA) are seen specifically in ~70% of RA patients; the presence of ACPA is clinically used as a diagnostic tool. Residues at amino acid (AA) position 11 in the HLA-DRB1 is the strongest susceptibility position in the HLA locus to ACPA(+) RA. Here, we investigated whether different individual ACPA have distinct genetic factors. We phenotyped RA cases with a multiplex peptide array to quantify 18 fine-specific ACPA as well as 2nd generation cyclic ctrullinated peptide antibody (CCP2), the representative method to quantify ACPA. We queried 6,267 patients and 12,054 controls in 3 independent cohorts. Clustering analysis of fine ACPA demonstrated that ACPA could be classified into 2 distinct clusters. The 12 antibodies (Ab) in cluster 1 included CCP2 and an Ab recognizing a peptide derived from fibrinogen beta AA position 60 to 74 with positions 60, 72 and 74 citurillnated (Fibbeta60.74Ab) which showed the strongest correlation with CCP2. Cluster 2 included Fib72Ab, which reacts to a peptide sharing 12 AA sequences to the Fibbeta.60.74Ab peptide but strikingly differs only in citrullination at only one AA position. We imputed HLA genotypes and tested for MHC associations with specific ACPA with logistic regression models correcting for 10 principal components and cohort effects. Position 11, with 6 possible residues, had the strongest association to 14 ACPA, including all Ab in cluster 1 among cases (omnibus $p \leq 1.1 \times 10^{-33}$). Intriguingly, Fib72Ab, the cluster 2 Ab showed distinct association patterns of AA residues at the position 11 ($p = 1.1 \times 10^{-50}$) from cluster 1 associations. We previously showed that HLA-DRB1 alleles interact to confer risk of ACPA(+) RA (Lenz et al Nat Gen 2015); we assessed whether allelic interactions influence expression of specific ACPA. We identified two novel allelic combinations specifically showing interactive effects on fine-specific ACPA, including an interaction between DRB1*13:01 and 11:01 which influences expression of another Ab in cluster 2. These findings suggest that specific HLA-DRB1 alleles, and their interacting combinations lead to a different repertoire of Ab, with reactivities to different citurillnated peptide sequences. They suggest that reactivity to multiple antigen sets might independently lead to RA.

1259T

Trans-ethnic genetic association studies facilitate the mapping of novel psoriasis susceptibility loci with higher resolution. *L.C. Tsoi[1,2,3], R.P. Nair[1], M. Ghosh[4], M. Kabra[4], T. Tejasvi[1], P.E. Stuart[1], J.J. Voorhees[1], X. Wen[3], H.M. Kang[3], G.R. Abecasis[3], V.K. Sharma[4], J.T. Elder[1,5] on behalf of an international group of investigators studying psoriasis genetics.* 1) Dermatology, University of Michigan, Ann Arbor, MI, USA; 2) Dept of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA; 3) Dept of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 4) All India Institute of Medical Sciences, New Delhi, India; 5) Ann Arbor Veterans Hospital, Ann Arbor, MI, USA.

Psoriasis is a chronic autoimmune disease of skin with a complex genetic architecture, and it affects over 100 million people worldwide. Genetic association studies have revealed over 80 psoriasis susceptibility regions to date; however, most of these efforts have focused on European-origin and Chinese populations. Identification of disease variants in other populations may facilitate the prioritization of causal genetic variants by exploiting the differences in haplotype structure. To this end, we performed GWAS in an Indian sample consisting of 2,599 cases and 1,734 controls. Indicative of common disease signals and mechanisms across the two populations, effect sizes of the known psoriasis susceptibility regions were highly correlated in the Indian and European datasets (Sperman r = 0.65; $p = 3 \times 10^{-9}$). Based on this observation, we performed a meta-analysis aggregating the Indian cohort with genetic association results of a meta-GWAS of European-origin individuals consisting of over 30,000 individuals. We identified three new loci achieving genomewide significance (1p36.22, 1q24.2, and 13q14.2), including one locus that has been identified in a previous Chinese study (1p36.22) but never established at genomewide significance in European studies. Finally, we utilized a Bayesian approach to compute the posterior probability of being a causal variant for each of the known and novel loci. By identifying sets of SNPs residing in 95% credible intervals, we were able to provide higher resolution (mean reduction in 95% credible interval = 20,000 bp) and to reduce the number of potential causal variants (mean reduction = 10) in the trans-ethnic study when comparing with the European-based study for all novel loci. Our study reveals elements of common genetic architecture shared across ethnicities in psoriasis, and highlights the value of trans-ethnic meta-analysis to uncover new disease-associated variants and to enhance the fine-mapping of susceptibility loci.

1260F

Genetic factors for lupus nephritis in Asian populations. *W. Yang, H. Zhang, Y. Zhang, Y.F. Wang, T.Y. Wang, J. Yang, Y.L. Lau.* Department of Paediatrics & Adolescent Medicine, Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong.

Systemic lupus erythematosus (SLE) is a prototype autoimmune disease characterized by autoantibody production and multi-organ damage, which predominantly affects women in child bearing age. Lupus nephritis(LN) is a common manifestation of SLE and an important feature for the morbidity and mortality of the disease and a major cause of end-stage organ failure in SLE patients. Asian patients were reported to have higher rates of LN and more active glomerulonephritis compared to White patients. Previously, it was reported some genetic loci are more strongly associated with LN than with SLE patients without LN, including HLA-DRB1, STAT4, ITGAM and TNFSF4. A patient only GWAS comparing LN patients with SLE patients without LN also reported association for LN of PDGFRA, SLC5A11, ID4, and HAS2. However, all these studies were conducted on European populations and studies focusing on Asian samples who have more LN and more severe LN are scarce. Here we analyzed our Hong Kong GWAS data (paed.hku.hk/genome) by comparing LN patients with controls and SLE patients without LN with controls, and compared the results of the two analyses. We observed a clearly stronger association for patients with LN for NCF2, ITGAM, TNFAIP3, ELF1, ETS1, SLC15A4 and WDFY4. The only known susceptibility gene that may be protective against LN is CD80. Our analyses suggest potential population differences on the genetic variants that predispose to LN in Asian populations than in European populations. Studies focusing on subphenotype differences with big enough sample size and diverse ethnicity background are needed to confirm the differences and to identify novel loci for LN susceptibility. We also elaborate the pitfalls in subphenotype analysis by proposing always including controls in the analysis to reduce false positive detection and always evaluating a locus comprehensively by focusing on top and independent variants from a locus for valid conclusions on subphenotype association.

1261W

Impact of violence on allostatic load in African American young adults. *L. Jackson[1,2], F. Saadatmand[2], M. Shestov[3], G. Dunston[1], J. Wright[2].* 1) National Human Genome Center, Howard University, Washington, DC; 2) Pediatrics, Howard University, Washington, DC; 3) Genomics and Computational Biology, Biomedical Graduate Studies, University of Pennsylvania, Philadelphia, PA.

Allostatic load, the chronic stress-induced wear and tear on the body, has a cumulative deleterious effect in individuals over their lifetime. Recent studies have suggested that socioeconomic status, psychological determinants and biomedical health cumulatively contribute to allostatic load in young adults. While these finding individually suggest that African American children may be particularly susceptible to the effects of allostatic loading due to racially based discrimination and economic instability, few studies have shown the effect of exposure to violence on the allostatic load carried by young African Americans. The Biological and Social Correlates of Drug Use in African American Emerging Adults or BADU dataset is composed of young African Americans (N=557 individuals) living in the Washington DC area, collected from 2010 to 2012. Study participants were sought equally between males and females (N_{MALES}= 283, $N_{FEMALES}$=274). This dataset provides a rich source of information on the behavioral, mental, and physical health of African American young adults (18-25 year olds) living in the Washington DC area. This includes exposure to community, interpersonal and intimate partner violence. Analysis of six biomedical markers measured in BADU study participants: C Reactive Protein, Cortisol, Epstein Barr Virus IgG, IgE, IgA and IgM, known to be associated with allostatic load show that cortisol is positively correlated to reported emotional state(R=0.072) and perceived individual discrimination (R=0.059). Allostatic load appears to be high in individuals who self-report exposure to violence. These findings suggest that allostatic load is an important factor in health and disease outcomes for young African Americans.

1262T

The genetics of red blood cell density in sickle cell disease patients. *Y. Ilboudo[1], P. Bartolucci[2], F. Galacteros[2], S. Alper[3], C. Brugnara[4], G. Lettre[1,5].* 1) Biochemistry, University of Montréal, Montréal, Québec, Canada; 2) Red Cell Genetic Disease Unit, Hôpital Henri-Mondor, Assistance Publique–Hôpitaux de Paris (AP-HP), Université Paris Est IMRB - U955 - Equipe n°2, Créteil, France; 3) Division of Nephrology, Beth Israel Deaconess Medical Center, Boston, USA Department of Medicine, Harvard Medical School, Boston, USA; 4) Department of Laboratory Medicine, Boston Children's Hospital, Boston, Massachusetts, USA; 5) Department of Medicine, Université de Montréal, Montréal, Québec, Canada.

Sickle cell disease is a monogenic blood disorder caused by mutations in the β-globin gene. Although a single gene is involved in this disorder, the clinical symptoms are heterogeneous. Dehydrated red blood cell density (DRBC) is a typical attribute of sickle cell disease (SCD) patients. Recent studies have linked higher percentage of DRBC with increased incidence of SCD-related pathophysiologies caused by the deformation of erythrocytes in blood vessels. Based on these novel findings, we sought to identify the genetic determinants of DRBC in SCD patients. We conducted a genome-wide association study (GWAS) in 405 SCD patients from the GEN-MOD study. These participants are of recent African descent and were recruited in France. Genotyping was done on the Illumina HumanOmni2.5Exome array and imputation was performed using 1000 Genomes Phase 3 haplotypes (v.5). We also performed gene-based tests (SKAT & VT), focusing on non-synonymous and splice site variants with a minor allele frequency <5%. DRBC was analyzed using an additive linear regression model, correcting for age, sex, and the first 10 principal components. Finally, we selected 65 patients from the GEN-MOD study for whole-exome sequencing (WES), and investigated the association between DRBC and rare detrimental mutations in strong candidate genes that play a role in red blood cell biology. The GWAS analyses were inconclusive as none of the variants reached array-wide significance ($P<2x10^{-7}$). This lack of association probably reflects our modest sample size to find association between DRBC and common variants of weak effects. The WES experiment identified 230 mutations in 50 candidate genes. Of interest, we found that the patient with the highest percentage of DRBC (37%) had a missense mutation in *ATP1B2*, the beta2 subunit of the ouabain-sensitive Na+/K+ -ATPase. This gene is also known as *AMOG* (Ca2+-dependent adhesion molecule of glia) and is expressed abundantly in the brain and the retina. It binds to retinoschisin, a retinal degeneration gene for X-linked human juvenile retinoschisis. We present the first genetic study of DRBC, a biomarker and potential modifier of severity in SCD. We found no evidence of common variants with strong effect on DRBC variation. Our candidate-gene approach based on WES highlighted several promising candidate mutations, although functional validation is required to confirm their mechanistic effect.

1263F

Polymorphisms in *IL18* gene are associated with person-to-person variation in IL18 levels in individuals with HCV and/or HIV infection. *C. Vergara[1], C.L. Thio[1], R. Latanich[1], G. Kirk[2], S. Mehta[2], A. Cox[1], A. Kim[13], G. Lauer[13], M. Busch[3], E. Murphy[3,4], M. Villacres-Cevallos[5], M. Peters[4], A. French[6], E. Golub[2], J. Eron[7], C.D. Lahiri[8], S. Shrestha[9], D. Gustafson[10], M. Young[11], K. Anastos[12], B. Aouizerat[14], D. Thomas[1], P. Duggal[2].* 1) Division of Infectious Diseases, Johns Hopkins University, Baltimore, MD. USA; 2) Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD.,USA; 3) Blood Systems Research Institute, San Francisco, CA.,USA; 4) University of California, San Francisco, CA.,USA; 5) Keck School of Medicine of the University of Southern California, Los Angeles, CA., USA; 6) CORE Center/Stroger Hospital of Cook County, IL., USA; 7) University of North Carolina at Chapel Hill, NC.,USA; 8) Emory University School of Medicine, Atlanta, GA, USA; 9) The University of Alabama at Birmingham, AL., USA; 10) State University of New York - Downstate Medical Center, NY, USA; 11) Georgetown University Medical Center, Washington, DC.,USA; 12) Albert Einstein College of Medicine and Montefiore Medical Center, NY.USA; 13) Massachusetts General Hospital, Boston, MSS.,USA; 14) New York University, NY.,USA.

Inflammasomes are multi-protein cytosolic complexes that integrate several pathogen-triggered signaling cascades ultimately leading to caspase-1 activation and the generation of the pro-inflammatory cytokine interleukin (IL)-18. Hepatitis C virus (HCV) and human immunodeficiency virus-1 (HIV) infections are associated with elevated plasma levels of IL18 suggesting inflammasome activation. However, there is marked person-to-person variation in the inflammasome response to HCV and HIV as demonstrated by varying blood levels of IL18. We hypothesized that there is a genetic basis for this variation. To test this hypothesis, we analyzed the associations of plasma IL18 levels and polymorphisms in 10 genes that were identified to participate in various stages of the inflammasome signaling cascade and/or IL18 production. We studied 1538 subjects with active HCV and/or HIV infection including 386 individuals of European ancestry, 796 individuals of African ancestry and 356 individuals of Hispanic ancestry. Samples were genotyped using the Illumina Human Omni-Quad and Omni.2.5. IMPUTE2 was used for imputation to 1000 genomes Phase III and PLINK was used test the association of variants with LogIL18 levels using an additive model. HCV and HIV infection status and HIV RNA, plus 2 principal components were included as covariates. The results obtained in the three ethnic groups were meta-analyzed with META. We observed significantly associated variants in the *IL18-BCO2* region on chromosome 11 in two imputed intronic SNPs: rs111311302 and rs80011693 in the meta-analysis. The presence of one G allele in rs111311302 confers a decrease of 0.06 units of LogIL18 levels in comparison with the C allele (β for G allele = -0.06, SE: 0.02, P value= $2.7 x 10^{-4}$). Both variants are in high linkage disequilibrium (LD), r2=0.98-1 in the three ethnic groups and the same values were observed for rs80011693 (T>C). No associations were detected between IL18 levels and polymorphisms in *IL-1β*, *ASCC1*, *NLRP3*, *NFKB1*, *NFKB2*, *CASP1*, *TLR7*, *TLR8* and *MYD88*. In conclusion, person to person variation in DNA of IL18 appears to modify IL18 production in response to HIV and HCV infections. While additional research is needed to demonstrate the mechanism, these data may help explain variability in the long-term inflammatory outcomes of chronic viral infections.

1264W

Integration of GWAS and lung eQTL to identify asthma genes. *J. Berube[1], M. Lamontagne[1], M. van den Berge[2], D.D. Sin[3], K. Hao[4], D.C. Nickle[5], W. Timens[2], P.D. Paré[3], M. Laviolette[1], Y. Bossé[1,6].* 1) Institut universitaire de cardiologie et de pneumologie de Québec, Quebec City, Canada; 2) University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, Netherlands; 3) The University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada; 4) Icahn School of Medicine at Mount Sinai, New York, NY, United States; 5) Merck & Co, MRL, Seattle, Washington, United States; 6) Department of Molecular Medicine, Laval University, Quebec City, Canada.

   Several genetic variants associated with asthma have been identified through GWAS. An integration of asthma GWAS data with lung expression quantitative trait loci (eQTL) has the potential to reveal the most likely causal genes in asthma susceptibility loci. Here, we leverage the results of the largest publically available GWAS data on asthma and a large-scale eQTL study performed in human lung tissues to elucidate causal genes in asthma susceptibility loci. First, the summary statistics of the 10,365 asthmatic subjects and 16,110 controls from the GABRIEL consortium were obtained for five loci that showed genome-wide significance and two additional borderline significant loci. Missing summary statistics were imputed using the ImpG software and the 1000 genomes European genotypes. Second, the lung eQTL dataset summary statistics were calculated 1 Mb up- and down-stream of the sentinel asthma associated SNPs for 1,038 individuals with genome-wide genotyping and gene expression data. Colocalisation analyses were then performed using the R package COLOC with a $p_{12}=10^{-5}$. For the 17q21 locus, the gene *GSDMA* demonstrated a 96.9% posterior probability that the asthma and the lung eQTL signals colocalized, suggesting that the genetic association and lung eQTL signals shared causal variants. The most likely SNP is rs4580194 with an imputed p-value for asthma of 1.04e-11 and an eQTL p-value with *GSDMA* of 1.54e-83. For the other asthma loci on chromosomes 2 (*IL18R1*), 5 (*SLC22A5* and *IL13*), 9 (*IL33*), 15 (*SMAD3*), 15 (*RORA*) and 22 (*IL2RB*), the highest posterior probabilities for genes surrounding asthma SNPs were between 2 and 3%. This study supports *GSDMA* as the most likely asthma gene on chromosome 17q21. Other GWAS-nominated asthma loci will require further investigation.

1265T

Genome-wide association analysis reveals variants on chromosome 19 that contribute to childhood risk of chronic otitis media with effusion. *E. Einarsdottir[1,2], L. Hafrén[3], E. Leinonen[1], M.F. Bhutta[4], E. Kentala[3], J. Kere[1,2], P.S. Mattila[3].* 1) Folkhälsan Institute of Genetics, and Molecular Neurology Research Program, University of Helsinki, Helsinki, Finland; 2) Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 3) Department of Otorhinolaryngology, Helsinki University Hospital, Helsinki, Finland; 4) Children's Surgical Centre, Phnom Penh, Cambodia.

   We performed a genome-wide association (GWA) study to identify genetic risk factors underlying childhood otitis media (OM [MIM166760]). The study included 829 Finnish affected children and data from 2118 controls genotyped previously. The most significant findings were validated on an independent platform. An association was found to a locus on chromosome 19 (best P = 1.77×10^{-7}, OR = 1.59). This locus harbors six genes in strong linkage disequilibrium. In a sub-phenotype analysis of the 509 patients with chronic otitis media with effusion (COME), one marker reached genome-wide significance (P = 2.92×10^{-8}). This association, albeit to the opposite allele, was replicated in a UK family cohort of 4860 subjects (best P = 1.62×10^{-4}, OR = 0.71). Our study is the largest reported association study performed on OM. Its main strength is the large sample size and the homogeneous Finnish population from which it is derived. We hypothesize that the genomic region on chromosome 19 is important for OM, and particularly COME, risk in both the Finnish and UK populations. Although the precise risk variants and mechanism remain unclear, our study suggests that the identified region on chromosome 19 includes a novel and previously uncharacterized risk locus for OM.

1266F

GWAS for serum galactose-deficient IgA1 (Gd-IgA1) points to critical genes of the O-glycosylation pathway. *K. Kiryluk[1], Y. Li[1], Z. Moldoveanu[2], H. Suzuki[3], J. Xie[4], N. Mladkova[1], R. LeDesma[1], P. Hou[5], I. Ionita-Laza[1], N. Chen[4], H. Zhang[5], J. Novak[2], A.G. Gharavi[1].* 1) Columbia University, New York, NY; 2) University of Alabama at Birmingham, Alabama, USA; 3) Juntendo University Faculty of Medicine, Tokyo, Japan; 4) Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 5) Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China.

Background: IgA1 O-glycosylation defects are universally detected among patients with IgA nephropathy (IgAN), the most common form of primary glomerulonephritis worldwide. Serum level of galactose-deficient IgA1 (Gd-IgA1) represents a known heritable biomarker for IgAN, but specific genetic factors involved in its determination are not known. **Methods:** We performed a quantitative GWAS for serum Gd-IgA1 in 2,633 individuals of Asian and European ancestry. In the discovery phase, we used HAA-lectin based ELISA to analyze sera of 1,195 individuals of East Asian and European ancestry genotyped with Illumina 610-quad and 550v3 platforms, respectively. We used a linear model with individual SNPs coded as additive genetic predictors and the outcome defined as standardized residuals of serum Gd-IgA1 after adjustment for age and sex. Each cohort was additionally adjusted for significant PCs of ancestry. Suggestive loci were followed by targeted genotyping in 1,438 additional individuals of East Asian and European ancestry. Subsequently, all cohorts were meta-analyzed jointly to identify novel genome-wide significant loci. **Results:** The strongest association signal was located within a 200-kb interval on chr.7p21.3 containing *C1GALT1* gene (rs13226913, P = 3.2 x 10^{-11}). The protein encoded by *C1GALT1* generates the common core 1 O-glycan structure by the transfer of galactose (Gal) from UDP-Gal to GalNAc-alpha-1-Ser/Thr. Core 1 is a precursor for O-glycans in the hinge region of circulating IgA1, as well as in many extended mucin-type O-glycans on cell surfaces. Notably, rs13226913 has a highly significant cis-eQTL effect on *C1GALT1* mRNA levels (P = 3.9 x 10^{-23}, the Gd-IgA1-increasing allele is associated with lower levels of mRNA expression). The second locus comprises a 100-kb interval on chr.Xq24 and contains *C1GALT1C1* gene (rs5910940, P = 2.7 x 10^{-8}), the molecular partner of C1GALT1 essential for its proper enzymatic activity. We demonstrate that non-coding variants in *C1GALT1* and *C1GALT1C1* genes jointly explain up to 7% of the overall variability in circulating Gd-IgA1. **Conclusions:** In the first GWAS for serum Gd-IgA1 we discovered two genome-wide significant loci that encode the molecular partners involved in the critical step of IgA1 *O*-glycosylation. Our findings provide new insights into the genetic regulation of O-glycan synthesis and are potentially relevant to several human diseases, including IgAN, inflammatory bowel disease, hematologic disease, and cancer.

1267W

The psoriasis risk allele *HLA-C*06:02* shows evidence of association with chronic and recurrent tonsillitis. *L.L.E Koskinen[1], K. Haapasalo[1], J. Suvilehto[2], P. Jousilahti[3], A. Wennerström[3], S. Suomela[4], R. Trembath[5], J. Barker[5], J. Vuopio[6,7], J. Kere[8,9], T.S. Jokiranta[1], P. Saavalainen[1].* 1) Research Programs Unit, Immunobiology, University of Helsinki, Helsinki, Finland; 2) Department of Otorhinolaryngology, Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 4) Department of Dermatology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 5) Division of Genetics and Molecular Medicine, Faculty of Medicine and Life Sciences, Kings College London, SE1 9RT, UK; 6) Medical Microbiology and Immunology, University of Turku, Turku, Finland; 7) Bacterial Infections Unit, National Institute for Health and Welfare, Helsinki, Finland; 8) Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 9) Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland.

Pharyngeal tonsillitis is one of the most common upper respiratory tract disorders in humans where Group A streptococcus is considered to be clinically the most important bacterial pathogen. While many patients experience only occasional throat infections with a relative short duration, a subset of patients suffer from chronic or recurrent tonsillitis. The reason for this is not yet fully understood. Susceptibility to tonsillitis has been suggested to have a genetic component. To search for new genetic biomarkers in pharyngeal tonsillitis, we performed an association study using the ImmunoChip (Illumina Infinium SNP array). Over 190,000 SNPs were genotyped in 96 Finnish patients subjected to tonsillectomy due to chronic or recurrent tonsillitis. 504 controls from the DILGOM/FINRISKI 2007 cohort were previously genotyped with ImmunoChip. The data was analyzed for association with PLINK. Classical alleles and amino acids of the HLA class I and II genes were imputed from the SNP genotypes with SNP2HLA. ImmunoChip data (from the Genetic Analysis of Psoriasis Consortium) and *HLA-C*06:02* genotype information from 163 Finnish patients with psoriasis were used in comparing the results with the tonsillitis data. For tonsillitis, the strongest association was identified on chromosome six at the MHC locus nearby genes *RNF39*, *TRIM31*, *HLA-C* and *HLA-B* (p=3.07E-6). Association analysis of the imputed HLA alleles suggested *HLA-C*06:02* as a risk factor for tonsillitis (p=5.0E-4, OR 2.3). Haplotype analysis pointed towards *HLA-C*06:02/HLA-B*5701* as the haplotype conferring the strongest risk for tonsillitis (p=8.0E-3, OR 5.5). Comparison of the tonsillitis and psoriasis genotype data at the HLA locus showed that the same haplotype increased the risk of both diseases.   We identified evidence of association with recurrent/chronic tonsillitis at the HLA locus, with the *HLA-C*06:02* haplotype conferring the highest disease risk. Interestingly, this haplotype is a known risk haplotype in psoriasis also. Our finding suggests a novel genetic explanation for the known link between streptococcal throat infections and onset and pathogenesis of psoriasis. The result supports the theory of molecular mimicry between streptococcal M protein and skin keratins, and points to presentation of shared antigenic T cell epitopes by HLA-Cw6 in both diseases.

1268T

Combining eQTL analyses with CRISPR/Cas9 experiments to fine-map causal variants that regulate *ATP2B4*, a gene associated with red blood cell traits and malaria resistance. *S. Lessard[1], E. Stern[2,3], D.E. Bauer[2], G. Lettre[1].* 1) Department of Medicine, Université de Montréal, Montreal Heart Institute, Montreal, Quebec, Canada; 2) Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Harvard Stem Cell Institute, Boston, Massachusetts; 3) Tufts University School of Medicine, Boston, Massachusetts.

A major challenge arising from the success of genome-wide association studies (GWAS) is to link associated non-coding variants to causal genes. This information is essential to translate genetic discoveries into therapeutic breakthroughs. Here, we explore how to combine expression quantitative trait loci (eQTL) mapping and the CRISPR/Cas9 genome-editing tool to dissect the *ATP2B4* GWAS locus associated with red blood cell (RBC) traits and malaria resistance.Using RNA-seq and genotype data from human erythroblasts, we identified 534 genes with strong evidence of allelic imbalance (AI, binomial P<1.6x10^{-5}). For 215 of these genes, we identified at least one significant eQTL at a false discovery rate <5%. Although most of these eQTL associations are replicated in other tissues from GTEx, our results highlighted 64 eGenes that are specific to erythroblasts. We found an enrichment of erythroblast eQTL in erythroid enhancers and ChIP-seq peaks for the erythroid transcriptional master regulator GATA1.Of particular interest, we identified a cluster of SNPs on chromosome 1 in strong linkage disequilibrium that are eQTL for *ATP2B4* (P=5x10^{-6}), which encodes the main plasma membrane calcium-transporting ATPase in RBCs. These SNPs overlap an erythroid-specific DNase I hypersensitive site (DHS) and are associated with RBC traits and malaria resistance. Using the CRISPR/Cas9 system in a human erythroid cell line, we deleted this DHS and observed a dose-dependent effect on *ATP2B4* expression levels, where cells with bi-allelic deletions displayed almost no *ATP2B4* expression. Smaller deletions narrowed down the essential regulatory sequence to a 98 bp region that contains 5 GATA1 binding motifs overlapping several eQTL. Individual disruptions of these motifs suggested cooperative regulation of *ATP2B4* by constituent GATA1 binding sites. Overall, this data suggests that the biological basis of the *ATP2B4* GWAS associations is through modulation of core erythroid enhancer function by common haplotypes. We are now exploring the effect of *ATP2B4* deficiency on calcium transport, hemoglobin concentration, and plasmodium invasion.Our study provides a proof-of-concept that eQTL mapping followed by genome editing in relevant cells facilitates the molecular characterization of GWAS signals. The *ATP2B4* minimal regulatory sequence would have been missed if we had not mapped eQTL specifically in erythroid cells, emphasizing the importance of employing the appropriate cell model.

1269F

A synonymous variant in *IL10RA* affects RNA splicing in pediatric patients with refractory inflammatory bowel disease. *S. Oh[1], J. Baek[2], H. Liany[3], J.N. Foo[3], K.M. Kim[1], S.C. Yang[2], J.J. Liu[3], K. Song[2].* 1) Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, South Korea; 2) Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea; 3) Human Genetics Group, Genome Institute of Singapore, Singapore.

Interleukin-10 receptor (*IL10R*) mutations are associated with severe childhood inflammatory bowel disease (IBD). Two unrelated patients who died of very early-onset severe IBD and sepsis were identified to harbor the same compound heterozygous mutations in *IL10RA* (p.R101W; p.T179T). A third patient was found to be homozygous for p.T179T. The missense change of p.R101W has been reported. The synonymous change of p.T179T, with a minor allele frequency of 0.035% in the population, was novel. The p.T179T mutation was located before the 5′ splice donor site, leading to exon skipping and out-of-frame fusion of exons 3 and 5, causing altered STAT3 phosphorylation in IL10-induced peripheral blood mononuclear cells. The patient developed colitis at six years of age, the oldest reported age of onset among patients with *IL10RA* mutations, and did not suffer from perianal disease. We report three pediatric patients with a rare, synonymous p.T179T variant causing a splicing error in *IL10RA*.

1270W

Human genetic determinants of the humoral immune response to common pathogens. *P. Scepanovic[1,2], C. Alanio[3,4], C. Hammer[1,2], D. Duffy[3,4], J. Antonsson[5,6], L. Abel[7,8], L. Quintana-Murci[9,10], M. Albert[3,4,11], J. Fellay[1,2], The Milieu Intérieur Consortium.* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Immunobiology of Dendritic Cell Unit, Institut Pasteur, Paris, France; 4) Center for Human Immunology, Institut Pasteur, Paris, France; 5) Department of Automatic Control, LTH, Lund University, Lund, Sweden; 6) International Group for Data Analysis, Pasteur Institute, Paris, France; 7) Laboratory of Human Genetics of Infectious Diseases, Necker branch, Inserm U1163, Paris Descartes University, Imagine Institute, Paris, France; 8) St Giles laboratory of Human Genetics of Infectious Diseases, Rockefeller branch, The Rockefeller University, New York, NY, USA; 9) Unit of Human Evolutionary Genetics, Department of Genomes and Genetics, Institut Pasteur, Paris, France; 10) Centre National de la Recherche Scientifique, URA 3012, Paris, France; 11) Department of Cancer Immunotherapy, Genentech, South San Francisco, CA, USA.

   Introduction. Immunological responses to latent infections and vaccinations vary among individuals, and this is thought to be highly influenced by host genetics. To unravel the mechanisms of genetic control of the immunoglobulin G (IgG) responses to common latent pathogens and vaccines, we conducted genome-wide association studies (GWAS) in a well characterized population of 1000 healthy individuals recruited by the *Milieu Intérieur* Consortium. **Methods.** The *Milieu Intérieur* (MI) project is a population-based study that aims at assessing the determinants of immunological variance within a healthy population. To achieve this, 1000 healthy individuals, with a 1:1 sex ratio, and stratified across five-decades of life (age 20 – 69) were recruited. Serum from these donors were used for qualitative (serostatus: seropositive/seronegative) and quantitative assessment of IgG responses against Cytomegalovirus (CMV), Epstein-Barr virus (EBV), Herpes simplex virus 1 & 2 (HSV-1 & 2), Varicella zoster virus (VZV), *Helicobacter pylori*, *Toxoplasma gondii*, Influenza A virus, Measles, Mumps, Rubella, and Hepatitis B virus. We used the Illumina OmniExpress and HumanExome arrays to genotype the 1,000 MI donors. After imputation with IMPUTE2, logistic or linear regressions were performed to detect associations between 5 million human polymorphisms and antibody responses.   **Results and discussion.** No genome-wide significant associations ($P < 5 * 10^{-8}$) were found for serostatus. In contrast, genome-wide significant associations were observed for the levels of IgG mounted against EBV, Rubella, and HSV-1. For EBV (EBNA antigen) and Rubella the most relevant association concerns the MHC locus on chromosome 6 ($P = 1.1 * 10^{-9}$ and $P= 7.7 * 10^{-10}$ respectively) and confirms previous observations. By imputing classical HLA alleles and amino acids, we found that these associations rely on variations in amino acid composition of HLA-DRβ1 and HLA-DPβ1 molecules respectively. In parallel, we report new associations of (i) EBV EBNA IgG levels with *LOC646241* ($P = 3.4 * 10^{-9}$), (ii) EBV VCA IgG levels with *SND1* ($P = 2.7 * 10^{-9}$) and *LINC01090* ($P = 3.11 * 10^{-9}$), and (iii) HSV1 IgG levels with *CTNNA3* ($P = 7 * 10^{-12}$). Together our results provide new possible insights into mechanisms determining response to these viral antigens and encourage further genetic and functional work.

1271T

Combined QTL mapping and RNA seq in a mouse cross followed by association studies of the human orthologous of the differentially regulated QTL genes identifies new risk genes of periodontitis. *A.S. Schaefer[1], A. Shusterman[2], M. Munz[1,3], Y. Nevo[4], S. Elgavish[4], Y. Jockel-Schneider[5], C. Bruckmann[6], K. Berger[7], P. Hoffmann[8], M. Laudes[9], K. Divaris[10], S. Offenbacher[11], S. Jepsen[12], J. Erdmann[3], W. Lieb[13], A. Franke[14], S. Schreiber[9,14], H. Dommisch[1], E. Weiss[15], Y. Houri-Haddad[2], F. Iraqi[16].* 1) Department of Periodontology and Synoptic Dentistry, Charité - University Medicine Berlin, Germany; 2) Department of Prosthodontics, Hadassah Medical Center, Israel; 3) Institute for Integrative and Experimental Genomics, Lübeck, Germany; 4) Bioinformatics Unit of the I-CORE Computation Center at the Hebrew University and Hadassah, Israel; 5) Department of Periodontology, Clinic of Preventive Dentistry and Periodontology, University Medical Center of the Julius-Maximilians-University, Würzburg, Germany; 6) Department of Conservative Dentistry and Periodontology, Clinic of Dentistry, Bernhard Gottlieb University, Vienna, Austria; 7) Institute of Epidemiology and Social Medicine, University Münster, Germany; 8) Institute of Human Genetics, University of Bonn, Germany und Human Genomics Research Group, Departemnt of Biomedicine, University Hospital of Basel, Switzerland; 9) Clinic of Internal Medicine I, University Clinic Schleswig-Holstein, Kiel, Germany; 10) UNC School of Dentistry, Department of Pediatric Dentistry, Chapel Hill, NC, UNC Gillings School of Global Public Health, Department of Epidemiology, Chapel Hill, NC, USA; 11) UNC School of Dentistry, Department of Periodontology, Chapel Hill, NC, USA; 12) Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany; 13) Institute of Epidemiology, Christian-Albrechts-University, Kiel, Germany; 14) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 15) Dental School, Tel Aviv University, Israel; 16) Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Israel.

   Genome-wide association testing of single nucleotide polymorphisms (SNPs) is limited to detect the complete genetic variance of human diseases. One reason is the correction for multiple testing of large numbers of independent markers. In this instance, candidate gene studies that focus on confined chromosomal regions may be advantageous. An ideal mapping strategy takes account of the advantages of both approaches, but requires a system for an unbiased *de novo* generation of hypotheses on candidate genes. Herein, we describe a systems genetics approach, which exploits murine genome variability to systematically generate candidate genes that can be tested in human samples. Periodontitis (PD) has a high global prevalence rate of 11% for the severe forms but the genetic etiology is currently very poorly understood. In a previous study, we crossed the mouse lines A/J and BALB/cJ (resistant and susceptible to bacteria induced PD) into the F$_2$ generation and identified two QTLs of genomewide significance. Here, we performed RNA sequencing of hemi-maxilla of the extreme phenotypes of the F$_2$ offspring (HiSeq system, Illumina). Four differently expressed genes at the previously identified QTLs were detected using the software tools Cufflinks and DESeq. (*Ugt2a1*, *Pf4*, *Sult1d1*, *Stoml3*). The human orthologous were analysed for disease associated SNPs in imputed genotypes (1000G Phase3 EUR) of a German case-control sample of the very severe, rare, and early-onset phenotype aggressive PD (AgP; 717=cases, 4,079=controls, genotyped on OmniExpress Bead Chips, Illumina). Two haplotype blocks suggested association with PD and were located upstream *UGT2A1* (tagging SNP rs146712414, $P=9.1x10^{-5}$, odds ratio [OR]=1.34, 95% confidence interval [CI]=1,16-1.56) and at *PF4* (SNP rs1595009, $P=2.8x10^{-4}$, OR=0.77 [0.67-0.89]). These associations were tested in a sample of white US-Americans, diagnosed for the widespread but moderate chronic PD, (CP; 1,961=cases, 1,864=controls). With this sample, only the haplotype at *PF4* was associated with CP (rs1595009, p=0.029; $P_{pooled}=1.2x10^{-4}$). eQTL analysis (Haploreg, GTEx) showed that the risk G-allele was highly significant associated with elevated expression of the distant gene *PF4V1* ($p=3.7^{-10}$) and *CXCL5*, located *in cis* ($p=6.5x10^{-6}$). Our data show that expression patterns in mouse models can recapitulate human conditions and allow narrowing down a limited number of genes, which can be tested successfully for disease association in the human genome.

1272F

The role of rare variants in a pediatric inflammatory bowel disease cohort. *K. Shaw[1], D. Cutler[2], D. Okou[3], L. Denson[4], S. Kugathasan[3], M. Zwick[2].* 1) Genetics and Molecular Biology Program, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University, Atlanta, GA; 3) Department of Pediatrics, Emory University, Atlanta, GA; 4) Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

The two most common forms of inflammatory bowel disease (IBD) are Crohn's disease (CD) and ulcerative colitis (UC). CD and UC are severe chronic diseases characterized by relapsing-remitting gastrointestinal inflammation. Around 5% of existing IBD cases in the United States are patients under the age of 20. Studies of these pediatric cohorts can provide unique insights into the genetic architecture of IBD. Large genome-wide association studies of IBD have found 200 loci associated with disease but explain only 13.1% of variance in disease liability for CD and 8.2% for UC. In this study we found rare, nonsynonymous variants at well-conserved loci (CADD score >10) that were significantly more frequent in exome sequencing of 376 pediatric IBD patients than in publicly available Exome Aggregation Consortium (ExAC) data. We calculated the proportion of variance explained by these sites and performed pathway enrichment analysis of genes containing significant sites. We found 178 sites enriched in our pediatric IBD patients. Some sites overlapped with known CD and UC-associated genes, including NOD2 and several genes in the HLA system. Pathway enrichment analysis implicated many immune-related pathways consistent with our understanding of IBD such as detection of bacterium, antigen processing and presentation, toll-like receptor signaling, and interferon-gamma and interleukin-6 production. Other enriched pathways included those involved in metabolism of drugs and xenobiotics, as well as trafficking of metabolites.

1273W

Copy number variation of *ADAM3A* gene in systemic lupus erythematosus. *F.B. Barbosa[1], M. Simioni[2], C.E.V. Wiezel[1], F.R. Torres[2], M.C. Molck[2], E.A. Donadi[3], V.L. Gil-da-Silva-Lopes[2], A.L. Simões[1].* 1) Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto - SP, Brazil; 2) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Campinas - SP, Brazil; 3) Division of Clinical Immunology, Department of Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto - SP, Brazil.

Systemic lupus erythematosus (SLE [MIM: 152700]) is an autoimmune disease with a strong genetic background characterized by chronic inflammation and autoantibody production. Copy number variations (CNVs), deletions or duplications usually greater than 1 Kb in genomic DNA, can contribute to the variability of the risk for complex diseases including SLE. Blood cell lymphocytes from 23 unrelated SLE patients and 110 healthy subjects were submitted to the genome-wide human Cytoscan HD Array (Affymetrix®) to screen for CNVs in genomic DNA. Low copy number (<2 copies) of *ADAM3A* gene was identified as potential candidate to SLE susceptibility (p = 0.0297, Odds Ratio [OR] = 3.6, 95% confidence interval [CI] = 1.0—16.0). *ADAM3A* gene is inserted in a cluster with other genes of *ADAM* family involved in several inflammation processes. *ADAM3A* have been related to some types of cancer and autoimmune diseases. The purpose of this study was to confirm the role of the *ADAM3A* CNV in SLE using a target-specific methodology in a larger sample size. A total of 134 unrelated SLE patients and 201 healthy controls were genotyped by SYBR green-based quantitative PCR (qPCR) using the StepOnePlus instrument and software (Applied Biosystems). Fisher's Exact test was applied for comparing the frequencies of *ADAM3A* copy number states between SLE and healthy controls using R. The genotypes of individuals determined by both Cytoscan HD Array and qPCR were 100% and 90% concordant in case and control groups, respectively. Individuals with discordant genotypes in the two methodologies were excluded from the statistical analysis. The copy number state of *ADAM3A* varied from 0 to 3 in the SLE group, while it ranged from 0 to 8 copies in controls. No evidence of association of the *ADAM3A* deletion with SLE susceptibility was found (p = 0.4762, OR = 0.8, 95% CI = 0.5—1.4). However, high copy number of *ADAM3A* (>2 copies) were described as a protective factor against susceptibility to SLE (p = 0.0277, OR = 0.3, 95% CI = 0.1—0.9). Our findings suggest that common duplications in immunoregulatory genes may be a protective factor against the development of SLE. Support: FAPESP (2013/17062-9, 2011/23794-7), CNPq (312547/2009-9, 304455/2012-1) and FAEPA.

1274T

Revertant mutation in codon 215 of HIV-1 in Mexico between 2000-2009 and 2010-2014 in patients treatment-naive and with antiretroviral therapy.

J. Cardenas-Bedoya[1,2], M. Escoto-Delgadillo[1], B.M. Torres-Mendoza[1,3], D.A. Carbaja-Uribe[1], E.G. Hernandez-Lopez[1,5], E. Vazquez-Valls[1,4]. 1) Immunodeficiencies and Human Retrovirus Laboratory, Western Biomedical Research Center, IMSS , Guadalajara, Independencia, Mexico; 2) Ph.D in Human Genetics, Department of Molecular Biology and Genomics, University Center for Health Science, University of Guadalajara, Guadalajara, Jalisco, Mexico; 3) Department of Medical Clinics, University Center for Health Science, University of Guadalajara, Guadalajara, Jalisco, Mexico; 4) Directorate of Education and Research, U.M.A.E. Specialties Hospital, Western National Medical Center, Mexican Institute of Social Security, Guadalajara, Jalisco, Mexico; 5) Ph.D in Pharmacology, University Center for Health Science, University of Guadalajara, Guadalajara, Jalisco, Mexico.

Background: The high variability in the genome of Human Immunodeficiency Virus (HIV) is due to its rate of replication and inability of the reverse transcriptase to correct errors. These errors generate mutations that in some cases back to their wild type. Such as resistance mutations like T215Y/F and its revertants in treat-naive and with antiretroviral therapy (ART) patients. The objective was to compare the frequencies of variants of the reverse transcriptase gene of HIV-1 at codon 215 revertants mutations between 2000-2009 and 2010-2014 in the patients with ART or naive-treat. **Methods:** Were selected plasma samples from 473 patients with HIV-1 from 2000 to 2014, who attended the Laboratory of Immunodeficiencies and Human Retrovirus, Western Biomedical Research Center, Mexican Institute of Social Security. They were divided into four groups of patients: 1) ART-naive from 2000 to 2009 (n=95); 2) ART-naive from 2010 to 2014 (n=134); 3) ART-treat from 2000 to 2009 (n=121) y 4) ART-treat from 2010 to 2014 (n=123). Extraction of viral RNA was performed, an RT-PCR reaction and subsequently sequencing for reverse transcriptase. Genotyping was performed with Trugene HIV-1 system with kit the OpenGene DNA Sequencing System (Siemens; Ontario, Canada). The sequences were analyzed with the database of Stanford University and the revertants mutations were determined. Chi-square test was used to compare frequencies in SPSS version 22. **Results:** We found the frequency of mutations at codon 215 between 1 and 2 of 3.16% and 4.48% respectively. For T215Y only found in the group 1 at 1.05%. The variants of the revertant mutation for group 1 was T215A/C with a frequency of 1.05% for each variant, and group 2 was T215S/D/L/P with 1.49% for S/L and 0.75% for D/P. Group 3 had a mutation frequency of 48.76%, where T215Y was 33.88%, T215F 6.61% and the revertant mutation T215C 2.48%. T215V of 1.65% and T215S/D/L/H of 0,83% each one; in the group 4 frequency was 47.97% mutations, where T215Y was 23.58%, T215F 17.07% and the revertants mutations T215D/E 2.44% and T215C/V/I 0.81%. There are significant differences in T215Y/F between 1-3 and 2-4 groups (p<0.05). In revertant mutations there were no differences between periods, as well as naive and treated with antiretroviral drugs. **Conclusion:** The frequency of mutation variants have changes T215 showing that the virus is returning to its wild type.

1275F

Identification of the primary functional variants for primary biliary cirrhosis (PBC) susceptibility loci by high-density association mapping and *in silico/in vitro* functional analyses.

Y. Hitomi[1,10], K. Kojima[2,3,10], M. Kawashima[1,4], Y. Kawai[2,3], N. Nishida[1,5], Y. Aiba[6], M. Yasunami[7], M. Nagasaki[2,3,8], M. Nakamura[6,9], K. Tokunaga[1]. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 3) Graduate School of Medicine, Tohoku University, Sendai, Japan; 4) Japan Science and Technology Agency (JST), Tokyo, Japan; 5) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 6) Clinical Research Center, National Hospital Organization, Nagasaki Medical Center, Omura, Japan; 7) Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 8) Graduate School of Information Sciences, Tohoku University, Sendai, Japan; 9) Department of Hepatology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 10) These authors contributed equally to this work.

Primary biliary cirrhosis (PBC) is a chronic and cholestatic liver disease that is caused by autoimmune destruction of intrahepatic small bile ducts, eventually leading to liver cirrhosis and hepatic failure. Recently, *Human Leukocyte Antigen (HLA)*, *tumor necrosis factor superfamily member 15 (TNFSF15)*, *POU class 2 associating factor 1 (POU2AF1)*, and other genes have been identified as susceptibility genes to PBC in the Japanese population by our previous genome-wide association study (GWAS). However, especially for a gene region that is located in a large linkage disequilibrium (LD) block such as chromosome 17q12-21 (*ORMDL3-GSDMB-ZPBP2-IKZF3*), primary functional variants in these gene loci and their molecular mechanisms for disease susceptibility to PBC have not been clarified. In the present study, in order to identify the primary functional variants in chr.17q12-21, high-density association mapping was carried out based on SNP imputation analysis by IMPUTE2 (ver. 2.3.1), using the whole-genome sequence data from a reference panel of 1,070 Japanese individuals (1KJPN), together with genotype data from our previous GWAS (PBC patients: n = 1,389; healthy controls: n = 1,508). Among 95 SNPs that showed genome-wide significant associations ($P < 5.0 \times 10^{-8}$) by this high-density association mapping, candidate functional SNPs that potentially regulate the gene expression efficiencies were selected by *in silico* analysis based on the Regulome DB database and the UCSC genome browser. Primary functional variants for disease susceptibility were identified by *in vitro* functional analysis (Luciferase assay and EMSA) using the human bile duct cell line HuCCT1 and the human T cell line Jurkat. Additionally, e-QTL data from the GTEX-portal database supported the regulations of expression efficiencies by these functional variants ($P = 3.80 \times 10^{-21}$ in whole-blood, n = 338). These results suggested that high-density association mapping based on SNP imputation analysis and on *in silico / in vitro* functional analysis can identify primary functional variants in disease susceptibility genes. This study illustrated a systematic methodology for identification of primary functional variants from a pool of suggestive SNP associations, and this methodology may eventually contribute to novel diagnostic and therapeutic methods for PBC.

1276W

Prediction of Type 1 diabetes status with high precision, based on GWAS identified SNPs. *J.D. Roizen[1], A.M. Casella[1], J.P. Bradfield[2], H. Hakonarson[3].* 1) Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Quantinuum Research LLC, San Diego, CA, United States; 3) 3Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States.

Type 1 diabetes mellitus (T1DM) is the third most common chronic disease of childhood and the most common autoimmune disease in pediatrics with a prevalence of 1 in 500 children in the United States. Among childhood diseases not due to a single gene mutation, T1DM has a uniquely high proportion of genetically inherited risk. Genome wide association studies have identified 68 variants associated with increased risk for T1DM; approximately 50% of the genetic risk for T1DM is due to high risk HLA alleles (DR3 or DR4 in combination with DQ8). The non-HLA contribution of genetic predisposition towards diabetes has been well defined, however, they have been unused either for risk identification to identify prospective study participants or for risk stratification of study participants. Thus, a critical gap in type 1 diabetes research is that known inheritable contributors to type 1 diabetes have not been used to predict type 1 diabetes genetic risk. We used the currently identified single nucleotide polymorphisms (SNPs) identified as associated with T1DM risk to generate a risk score for individuals in a large T1DM dataset genotyped on Illumina human genotyping arrays (n = 18186; 6490 cases, 11696 controls). We used standard quality control measures for genotyping (minor allele frequency > 0.01, Hardy-Weinberg equilibrium 0.000001, and maximum per-person missing genotyping frequency of 0.05). We imputed HLA genotypes using SNP2HLA for imputation of non-HLA SNPs we used IMPUTE2. We used the approach described by Hu et al., including the use of three critical amino acids in the HLA for the aggregate of HLA risk score. We used the ProbABEL package in R to score all individuals and evaluate a SNP-based risk score. Applying this score on this dataset with the initial five eigenvectors to account for genetic background generated a receiver-operator curve (ROC) with an area under the curve (AUC) of 0.915; a re-regression of the betas for these inputs generates and ROC with an AUC of 0.926. Thus, we report a substantial improvement in our ability to predict T1DM and a milestone in GWAS – moving from identifying alleles to using alleles to highly-accurately stratify risk. The objective of this study has been a long-standing goal of the diabetes research community and is an essential element to intervene with preventive or curative therapy.

1277T

Finemapping reduces putative causal variants to five or fewer in 8 and 13 of rheumatoid arthritis and type I diabetes associated loci and identifies individual causal variants in 4 loci. *H. Westra[1,2], Y. Luo[1,2], S. Eyre[3], J. Worthington[3], J. Martin[4], T. Huizinga[5], L. Klareskog[6], S. Rantapaa-Dahlqvist[7], S. Onengut[8,9], A. Lee[10], J. Todd[11], W-M. Chen[8,9], A. Quinlan[8,12], P.K. Gregersen[10], S.S. Rich[8,9], S. Raychaudhuri[1,2,13,14].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Arthritis Research UK Centre for Genetics and Genomics, Musculoskeletal Research Centre, Institute for Inflammation and Repair, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK; 4) Instituto de Parasitologíay Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain; 5) Department of Rheumatology, Leiden University Medical Centre, Leiden, the Netherlands; 6) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 7) Department of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden; 8) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 9) Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA; 10) The Feinstein Institute for Medical Research, Northwell Health, Manhasset, New York, USA; 11) Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0XY, UK; 12) Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA; 13) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Smith Building, 1 Jimmy Fund Way, Boston, MA 02115, USA; 14) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

Rheumatoid arthritis (RA) and Type 1 diabetes (T1D) are seropositive autoimmune diseases in which destruction of joints (RA) and pancreatic beta-cells (T1D) respectively lead to chronic diseases with disability and decreased life expectancy. Genome-wide association studies (GWAS) have identified 101 loci associated with RA (Okada *et al*, *Nature*, 2014) and 53 loci associated with T1D (Onengut-Gumuscu *et al*, *Nat Genet*, 2015). However, array-based genotyping used in GWAS captures limited sets of variants, potentially missing causal variants. We therefore performed imputation on the 184 non-MHC ImmunoChip loci using the European (EUR) and cosmopolitan (COSMO) populations in 1000 genomes, and the Haplotype Reference Consortium (HRC) as reference. To ensure selection of the best imputation strategy for finemapping, we objectively assessed imputation quality by sequencing 900 1kb regions within 55 ImmunoChip loci in 744 individuals. Out of the 510 variants with minor allele frequency (MAF) > 1%, 72% (EUR), 72% (COSMO), and 50% (HRC) were imputed, of which 76% (EUR), 77% (COSMO) and 78% (HRC) were imputed with high quality (correlation of genotyped vs imputed dosage R^2 > 0.8). Based on these findings, we used the COSMO imputed ImmunoChip data for all 11,475 cases and 15,870 controls for RA, and 12,097 cases and 14,342 controls for T1D. We performed logistic regression on the imputed dosages (MAF > 0.5%, Hardy-Weinberg p > 10^{-5}; INFO > 0.3) and examined 33 RA loci RA and 48 T1D loci associated with disease in our samples (disease association p < 10^{-5}). Performing conditional logistic regression on these loci, we showed that 4 RA loci and 9 T1D loci harbor multiple independent variants (p < 10^{-5}). We then performed Bayesian fine-mapping (Wakefield, *Am J Hum Genet* 2007; Maller *et al*, *Nat Genet,* 2012), created credible sets of variants spanning 90% of the posterior probability, and identified 8 loci for RA and 13 loci for T1D having ≤ 5 variants in their credible set. This method identified a single, putatively causal variant at loci contributing to RA (*REL*) and T1D (*IL2RA*, *PTPN2, SH2B3*). In conclusion, our approach of dense genotyping, high-quality imputation, Bayesian fine-mapping, and logistic regression has pinpointed a set of putative causal variants in 8 loci for RA and 13 loci for T1D. These variants will guide future validation and functional studies. .

1278F

Body mass index is causally associated with pediatric and adult multiple sclerosis onset: A study of over 20,000 individuals using Mendelian randomization. *M. Gianfrancesco[1], X. Shao[1], J. Graves[2], A. Waldman[3], T. Lotze[4], T. Schreiner[5], B. Greenberg[6], B. Weinstock-Guttman[7], G. Aaen[8], J.M. Tillema[9], J. Ness[10], M. Candee[11], L. Krupp[12], M. Gorman[13], L. Benson[13], T. Chitnis[14], S. Mar[15], I. Kahn[16], M. Glymour[2], I. Jonsdottir[17], K. Stefansson[17], P. Strid[18], J. Hillert[18], A. Hedstrom[18], T. Olsson[18], I. Kockum[18], C. Schaefer[19], E. Waubant[2], L. Alfredsson[20], L.F. Barcellos[1].* 1) University of California, Berkeley, Berkeley, CA; 2) University of California, San Francisco, San Francisco, CA; 3) Children's Hospital of Philadelphia, Philadelphia, PA; 4) Texas Children's Hospital, Houston, TX; 5) University of Colorado School of Medicine, Aurora, CO; 6) University of Texas Southwestern, Dallas, TX; 7) SUNY Buffalo, Buffalo, NY; 8) Loma Linda University, Loma Linda, CA; 9) Mayo Clinic, Rochester, MN; 10) University of Alabama at Birmingham, Birmingham, AL; 11) University of Utah, Salt Lake City, UT; 12) SUNY Stony Brook, Stony Brook, NY; 13) Boston Children's Hospital, Boston, MA; 14) Brigham and Women's Hospital, Boston, MA; 15) Washington University St. Louis, St. Louis, MO; 16) Children's National Medical Center, Washington, DC; 17) deCODE Genetics, Reykjavík, Iceland; 18) Karolinska Institutet, Stockholm, Sweden; 19) Kaiser Permanente Division of Research, Oakland, CA; 20) Centre for Occupational and Environmental Medicine, Stockholm, Sweden.

Multiple sclerosis (MS) [MIM 126200] is an autoimmune disease with both genetic and environmental risk factors. Recent studies indicate childhood and adolescent obesity double the risk of MS, but this association may reflect unmeasured confounders rather than causal effects of obesity. We utilized separate-sample Mendelian randomization (MR) to estimate the causal effect of body mass index (BMI) on both pediatric and adult onset MS susceptibility. For non-Hispanic white members of Kaiser Permanente, Northern California (1,104 adult MS cases and 10,536 controls) and a replication dataset from Sweden (5,133 MS adult cases and 4,718 controls), a weighted genetic risk score (GRS) was calculated as an instrumental variable using 97 variants previously established to predict BMI (R^2= 2.7%). Models were adjusted for birth year, sex, education, smoking, ancestry, and genetic predictors of MS including *HLA-DRB1*15:01* and 110 non-HLA variants. We also examined this relationship in non-Hispanic white pediatric cases and controls from over 15 sites across the U.S. (total sample size: 394 pediatric MS cases, 10,875 controls). Adjusted MR estimates suggested higher genetically induced BMI predicted greater MS susceptibility in adult KPNC and Swedish datasets (odds ratio [OR] = 1.13, 95% CI 1.04, 1.22 and OR=1.09, 95% CI 1.03, 1.15, respectively). A significant causal association between BMI GRS and pediatric MS was also demonstrated (OR = 1.20, 95% CI 1.06, 1.36; p=0.004) after adjusting for sex, ancestry, *HLA-DRB1*15:01,* and 110 non-HLA MS risk variants. Although the mechanism remains unclear, these findings support a causal effect of increased BMI on both adult and pediatric onset MS and suggest a role for inflammatory pathways that characterize both autoimmune disease and obesity.

1279W

Identifying T1D risk loci with parent-of-origin effects. *S. Onengut-Gumuscu[1,2], W-M. Chen[1,2], E. Farber[1], P. Concannon[3,4], S.S. Rich[1,2], Type 1 Diabetes Genetics Consortium (T1DGC).* 1) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 2) Public health Sciences, University of Virginia, Charlottesville, VA; 3) Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL; 4) Genetics Institute, University of Florida, Gainesville, FL.

Type 1 diabetes (T1D) is an autoimmune disease that arises from the action of both genetic and environmental factors. T1D is the most common type of diabetes in children affecting up to 1 in 300 individuals. Fathers with T1D are more likely to transmit the disease to their offspring than mothers with T1D. On average, 5-10% of the offspring of affected fathers develop T1D, while only 2-5% of offspring of affected mothers develop T1D. This long-recognized epidemiologic observation suggests that DNA regulatory (epigenetic) factors could play an important role in T1D risk. In addition to parent-of-origin (POO) effects in T1D, these effects have been implicated in bipolar disease, cancer, and type 2 diabetes. To date, there are few large-scale family-based studies of autoimmune disorders that provide the data necessary to determine differential parental transmission of risk alleles. Over 2,000 affected sibpair (ASP) families were collected by the Type 1 Diabetes Genetic Consortium (T1DGC) with ImmunoChip (and other) genotyping data. We performed association analysis of T1DGC ASP families using the parent-of-origin test (TDT-POO) as incorporated in PLINK. These results support POO effects at several T1D loci, including 14q32.2, a well-established T1D risk locus, positioned in an imprinted region of the genome (*MEG3*). The 14q32.2 T1D locus shows paternal POO effects (P_{PAT} = 2.43x10⁻⁵, P_{MAT} = 0.7). The 16p13 locus is highly enriched for transcription factor binding events and has been shown to contain T1D-associated risk variants that regulate gene expression (*CLEC16A/DEXI*). Transmission of alleles from parents to affected offspring exhibit disease association with paternal transmission (P_{PAT} = 8.55x10⁻⁵, P_{MAT} = 0.01). The 20p13 locus (*SIRPG*) exhibits T1D risk associated with maternal transmission (P_{PAT} = 0.6, P_{MAT}= 3.11x10⁻⁶). Further studies examining the SNPs implicated in POO effects and the role of those SNPs on gene regulation and signaling are needed to understand the mechanisms underlying T1D risk.

1280T

Dissecting complex asthma phenotypes by mapping the *IL1RL1* 2q12 region to three genes and two cell types. *S. Ghosh[1], D. Mayhew[1], A. Yeo[2], M. Magid-Slav[1], G. Livi[3], Q. Lu[3], J. Betts[2].* 1) GlaxoSmithKline, King of Prussia, PA; 2) GlaxoSmithKline, Stevenage, Herts, UK; 3) GlaxoSmithKline, Collegeville PA.

Asthma is a heterogeneous disease characterized by recurrent wheezing, which is a manifestation of reversible airways narrowing, commonly to environmental stimulants. Many genes associated with asthma transmit danger signals of airways epithelial damage to the immune system that underpin the pathways for allergic asthma. *IL33*, which helps induce T helper type 2 [Th2] cytokines and its receptor, *IL1RL1* (expressed on mast cells, Th2, Tregs, macrophages), represent such a system. However, non-allergic asthma, about 50% of cases, can be driven by neutrophilic, Th17-dependent pathways that arise when infective agents activate IL1 and IL18 cytokines *via* the inflammasome. In the absence of a large meta-analysis and genetic fine mapping of European asthma samples, we used *in silico* approaches to identify potential effector genes in the strongly associated, noncoding, 2q12 region surrounded by the *IL1RL1, IL18R1* and *IL18RAP* genes. The cis-regulatory elements controlling expression of *IL1RL1*, *IL18R1*, and *IL18RAP* are within two adjacent, asthma-associated, linkage disequilibrium regions using the dataset of FANTOM5. The first associated region had enhancers active in neutrophils (correlating with *IL18RAP* and *IL18RL1* expression) and mast cells (correlating with *IL1RL1* expression). In contrast, the second associated region had an eQTL (rs13015714) for *IL18RAP* as well as asthma-associated SNPs (e.g. intronic rs3771166) that are in perfect LD with *IL1RL1* coding SNPs. We are locating open chromatin in the two asthma regions in stimulated neutrophils and mast cells, respectively, to narrow down to the etiological SNPs. Finally, we will use genetically modified cell lines that secrete any/all of these cytokines. Our aim is replace protective with susceptibility variants and measure real-time expression of each of these cytokines as well as signal transduction of IL1RL1 when coding SNPs are thought etiological. Thus, three genes in two cell types are implicated for the 2q12 association region, explaining the Th2 (mast cell) and Th17 (neutrophilic) phenotypes that are part of the clinical heterogeneity in asthma. Performing a similar dissection in other asthma-associated regions will clarify the degree to which these and other cellular mechanisms contribute to this disease. This will ultimately identify asthma subtypes for precision therapy.

1281F

A pilot study of blastomycosis susceptibility in a high risk population. *M.B. O'Neill, D. Xiong, B.S. Klein, C.S. Pepperell.* University of Wisconsin - Madison, Madison, WI.

Blastomycosis is a rare infection caused by the dimorphic fungus *Blastomyces dermatitidis*. The outcomes of infection with *B. dermatitidis* vary dramatically - many people are unaffected while others develop an invasive fungal infection that can result in respiratory failure, neurological damage and death. During recent blastomycosis outbreaks in Wisconsin, marked disparities in incidence rates were identified between residents of Hmong and European ethnicity (168 vs. 13 per 100,000, respectively). There is no evidence of differential exposure to the fungus nor of differences in co-morbidities that alter susceptibility, which suggests the disparity may have an underlying genetic basis. Hmong populations are expected to have high levels of autozygosity due to historically high rates of within-group marriage (endogamy) and recent population bottlenecks caused by repeated migrations, geographic dispersal, and population losses from military conflict. We hypothesize that individuals of Hmong-ancestry have increased susceptibility to blastomycosis as a result of homozygous variants that affect immune responses to the fungus, which are rare or absent in European-ancestry populations. In a pilot study, we used a disease mapping strategy designed to exploit the predicted high rates of autozygosity in Hmong populations to uncover potential loci mediating variable susceptibility to infection. PLINK was used to identify runs of homozygosity (ROH) in the whole genome sequence of five persons of Hmong-ancestry who developed blastomycosis during an outbreak in Marathon County, Wisconsin. Loci within ROH shared by all infected individuals ('consensus ROH') were annotate and cross-referenced with the 1000 Genomes, Genotype-Tissue Expression (GTEx), and Immune Variation (ImmVar) projects to determine the frequency of each single nucleotide polymorphism (SNP) across human populations, as well as their predicted or demonstrated effects on function. Variants within consensus ROH that were fixed in the five Hmong cases and found at 0-20% in the European superpopulation of the 1000 Genomes Project were inspected further for promise as disease susceptibility loci; SNPs identified as expression QTLs in the GTEx or ImmVar projects, SNPs annotated as regulatory, and/or SNPs with high FUNSEQ scores were deemed candidate disease susceptibility loci.

1282W

Whole genome sequence analysis of asthma in families from Barbados.
N. Rafaels[1], M. Daya[1], M. Boorgula[1], S. Chavan[1], C. Foster[2], P. Maul[3], T. Maul[3], H. Watson[4], I. Ruczinski[5], T.H. Beaty[6], R.A. Mathias[2], K.C. Barnes[1]. 1) University of Colorado Anschutz Medical Campus, Aurora, CO; 2) Johns Hopkins Asthma & Allergy Center, Johns Hopkins University School of Medicine; 3) Genetics and Epidemiology of Asthma in Barbados, The University of the West Indies; 4) School of Clinical Medicine and Research, Queen Elizabeth Hospital, The University of the West Indies; 5) Department of Biostatistics, School of Public Health, Johns Hopkins University; 6) Department of Epidemiology, School of Public Health, Johns Hopkins University.

Asthma is a complex disease with striking disparities between ethnicities, and asthmatics of African descent have more severe asthma, poorer response to therapy, and higher IgE levels compared to asthmatics of European ancestry. Genetics of asthma research has been conducted in the Barbados Asthma Genetics Study (BAGS) for >2 decades focused on a population of African descent. As part of the NHLBI's Trans-Omics for Precision Medicine (TOPMed) Program, high coverage whole genome sequencing (WGS; 39.4 average depth) was performed on samples collected from 190 families (mean pedigree size of 5.7 individuals, ranging from 1 to 56 individuals). Sequencing was 99.5% concordant with samples genotyped on the Illumina OMNI 2.5 array. The average Yoruba (YRI) component in Barbados is 88%. There were 41.9 million (M) variants observed, of which 6.3M (15%) were private. 59.8% (~25.1M) with an minor allele frequency (MAF) ≤1%, 17.2% (~7.2M) with a MAF 1-5%, and only 23.0% (~9.6M) were common (MAF>5%). On average, individuals had 4.3M variants different from reference (NCBI37), of which 1.6M (38.4%) were in genes, and ~24,000 (0.6%) were coding. This data set provides us with the opportunity to investigate the role that structural and rare variants may have in allergic disease and asthma. To illustrate the utility of the data set and to hone in on rare variation associated with disease, we examined one of the largest and well-characterized pedigrees of 31 individuals to determine how private variants may play a role in asthma pathogenesis. Among 226,498 variants private to this family, a novel missense variant predicted to be damaging by both polyphen2 and SIFT in the ADAMTS12 gene, a gene previously associated with asthma in linkage studies, is found in 15 of 31 individuals in this family. We discuss our quality control and analysis pipeline and showcase preliminary results.

1283T

Exploring causal regulatory variation underlying GWAS signals for generalized vitiligo.
G. Andersen[1], Y. Jin[1,2], D. Yorgov[3], T. Ferrara[1], S. Santorico[1,2,3], R. Spritz[1,2]. 1) Human Medical Genetics and Genomics, University of Colorado, Anschutz Medical Campus, Aurora, CO; 2) Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA; 3) Mathematical and Statistical Sciences, University of Colorado, Denver, USA.

From meta-analysis of three genome-wide association studies (GWAS) of vitiligo in European-derived whites, we have identified and replicated 48 loci, establishing a pathobiological framework for melanocyte-directed autoimmunity. At 42 (87.5%) of the vitiligo-associated loci, the most significant "lead SNPs" reside in non-coding regions, suggesting that many may alter gene expression. To investigate potential regulatory roles of variants at vitiligo susceptibility loci, we performed conditional logistic regression to identify likely causal variants, studied variant annotation from ENCODE regulatory datasets, applied PrediXcan to assess predicted gene expression differences between vitiligo cases and controls, and assessed co-localization of vitiligo associations and expression quantitative trait loci (eQTLs) with a Bayesian analysis. Specifically, we carried out conditional logistic regression to identify potential causal variation at each vitiligo susceptibility locus, and we annotated the most-associated "lead variants" against ENCODE datasets for multiple immune cell types and melanocytes. These analyses showed that many lead variants reside in DNase hypersensitivity sites and/or chromatin marks that represent active transcriptional enhancers and promoters. We used a Bayesian approach to assess co-localization of the vitiligo association peaks and eQTL peaks in purified blood monocytes, which mediate many aspects of the autoimmune response. This analysis indicated that 8 of the vitiligo associations correspond to *cis*-eQTLs that control expression of specific genes in the vicinity. To assess correspondence of the confirmed vitiligo association signals with differential gene expression in vitiligo cases versus controls, we used PrediXcan to predict expression of 11,553 genes for all 2,853 cases and 37,412 controls in our GWAS meta-analysis cohort. PrediXcan identified 83 genes with significant differential predicted expression in vitiligo cases versus controls following Bonferroni correction. Over 90% of these were located within 1 Mb of a confirmed vitiligo susceptibility locus. Together, our results indicate that for vitiligo as for other complex diseases, there is enrichment of causal variations that alter gene expression. This may bode well for identifying potential therapeutic targets, as pharmacologic modulation of dysregulated biological pathways may prove more tractable than attempting to target proteins impacted by amino acid substitutions.

1284F

First GWAS of inflammatory bowel disease in African Americans identifies *ZNF649* **and** *LSAMP* **as novel, African specific loci, and replicates and defines established inflammatory bowel disease loci for African Americans.** *S.R. Brant[1], D.T. Okou[2], C.L. Simpson[3], D.J. Cutler[2], T. Haritunians[4], J.P. Bradfield[5], P. Chopra[2], J. Prince[2], F. Begum[6], J.H. Cho[7], H. Hakonarson[5], M.E. Zwick[2], D.P.B. McGovern[4], S. Kugathasan[2].* 1) Meyerhoff IBD Center, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Emory University School of Medicine, Atlanta, GA; 3) The University of Tennesse Health Science Center, Memphis, TN; 4) F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 5) Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA; 6) Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 7) Dept of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

   INTRODUCTION: The inflammatory bowel diseases (IBD), comprising ulcerative colitis (UC) and Crohn's disease (CD), are chronic gastrointestinal diseases associated with significant morbidity and rising prevalence, particularly in the African American (AA) population. We previously found evidence in AAs, by genotyping IBD cases and controls using the Immunochip platform for 8 of the 200 established European IBD, but none was genome wide significant, and we found no significant evidence for novel loci. We therefore performed this first AA IBD GWAS. METHODS: We genotyped two AA cohorts: 1258 AA IBD cases coordinated by Johns Hopkins and Cedars Sinai IBD Genetic Centers genotyped on Omni2.5 and used dbGAP data from the Health Retirement Study on 1678 AA controls; and 1087 IBD cases coordinated by Emory University IBD Genetics Center genotyped on Affymetrix Axiom Genome-Wide Pan AFR World Array and 3324 previously genotyped Kaiser controls. We imputed 11.4 million common SNPs and performed meta-analysis of the 2 GWASs (2345 IBD, 5002 controls). RESULTS: Genome wide significant (GWS) associations were detected (p<5.0x10⁻⁸) for SNPs at *HLA-DRB1*, *ZNF649* (19q31), and *LSAMP* for UC, and *USP25* for IBD. Furthermore, GWS associations spanning *BTNL2* to *HLA-DQA1* were identified for all IBD, by conditioning on the top UC HLA association. *ZNF649 (*a gene influencing NF-kB and TNFa activation) and *LSAMP* (expressed in colon) represent novel IBD loci. Their associated SNPs are monomorphic in non-African populations, suggesting these loci are African specific. We found strong evidence (p<1.5x10⁻⁶) for SNPs at 8 additional established IBD loci (4 <u>novel</u> for the AA population) proximal or at genes <u>KCNQ2</u> for UC*; NOD2* and *IL23R for CD;* and *PTGER4,* <u>TNC</u>*, IL12B,* <u>CXCR6</u>, and <u>ADCY3</u> for IBD. Several, like *TNC* (p=9.6x10⁻⁸) a gene regulating intestinal barrier function and T-cell activation 200kb from the major Asian CD locus *TNFSF15,* showed unique patterns of association with African-specific alleles. Among 231 SNPs maximally and independently associated at established European loci, 104 showed evidence of replication (p<0.05) in AAs. CONCLUSIONS: The detection of novel IBD loci, African-specific ancestral alleles for both novel and established loci, and unique patterns of association, demonstrates the importance of high-density, gene mapping for IBD in the AA population. It remains vital to expand GWASs to understudied and unique populations to define genetic architecture for complex diseases.

1285W

Identification of three new susceptibility loci for inflammatory bowel disease in Koreans. *H. Oh[1], M. Hong[1], S. Jung[1], S.J. Ahn[1], C.H. Lee[2], E. Kim[2], B.D. Ye[3], S. Park[3], J. Liu[4], D.P.B. McGovern[5], B. Han[2], S.K. Yang[3], K. Song[1].* 1) Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, South Korea; 2) Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center; 3) Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine; 4) Genome Institute of Singapore, Singapore; 5) The F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute; Cedars-Sinai Medical Center, Los Angeles, USA.

   Recent meta-analysis of the genome-wide association studies (GWAS) identified over 200 IBD associated regions. However, identified common variants account for only a fraction of IBD heritability. Moreover, despite of observed differences in clinical characteristics of IBD among different ethnicities, there have been limited studies in non-European populations. We performed a genome-wide association study on IBD in Koreans. We used 1505 cases (922 CD, 583 UC) and 4041 controls for the discovery stage followed by replication in an additional 1,989 cases (993 CD, 996 UC) and 3,491 controls. We identified three new susceptibility loci for IBD at genome-wide significance: a SNV at 1q21 (odds ratio [OR] = 1.35, 95% CI = 1.22-1.49, combined *P* = 1.96 x 10⁻⁹), a SNV at 22q13 (OR = 1.23, 95% CI = 1.14-1.32, combined *P* = 2.48 x 10⁻⁸), and a SNV at 9p21 (OR = 1.17, 95% CI = 1.11-1.24, combined *P* = 5.45 x 10⁻⁸). In addition, we confirmed significant associations of 28 previously known IBD loci (*P*<2.16 x 10⁻⁴). Together, the loci we identified and replicated accounted for 6.82% and 5.53% of the total disease variance for CD and UC risk in Koreans, respectively. Estimate of the genetic correlation in risk between CD and UC was 0.47 (SE 0.08). Our study provides new biological insight to IBD in East Asians and supports the complementary value of genetic studies in different populations.

1286T

Cross-talks between autism and atopy: Perspectives from shared genetic etiology. *P. Lin, C. Cassidy, S. Berry, T. Mersha.* Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

   Individuals with autism spectrum disorder (ASD) might have a higher risk of atopic diseases. It has been speculated that immune dysregulation may play a role in the link between ASD and atopy. In this study, we searched for 16 million PubMed abstracts to extract lists of shared risk genes for ASD and three atopy-related diseases (i.e., asthma, COPD, and psoriasis). We have calculated inclusion index (ratio of overlapped genes to the reference list) to evaluate the relative contribution of overlapped genes to the etiologies of ASD. Additionally, we assess whether overlapped genes were over-represented in ASD-related gene networks, Finally, we used Panther web-tool to identify functional ontologies of the overlapped genes. We found more shared networks than expected by chance alone (hypergeometric p-value < 5×10^{-6} for all disease pairs, i.e., ASD-asthma, ASD-COPD, ASD-psoriasis). The inclusion indexes for all three of the "ASD-atopic disease" pairs were 0.25~0.3. Furthermore, the results reveal that the overlapped genes for ASD and asthma were significantly over-represented in the endogenous cannabinoid signaling pathway (fold enrichment = 19.3, p-value = 1.4×10^{-3}). The overlapped genes for ASD and COPD were significantly over-represented in the hopxia response via HIF activation pathway (fold enrichment = 26.8, p-value = 2.7×10^{-3}). The overlapped genes for ASD and psoriasis were significantly over-represented in the endogenous interferon-gamma signaling pathway (fold enrichment = 18.1, p-value = 1.2×10^{-2}). An improved understanding of these shared networks/pathways could provide valuable insights into the causal pleiotropic effects that may contribute to comorbidity between ASD and atopy.

1287F

Whole genome sequencing-based detection of a common African structural variant of the alpha-globin gene cluster and its interaction with sickle cell trait on red cell and kidney phenotypes: The NHLBI TOPMed Project. *A.P. Reiner[1,10], R.E. Handsacker[2,3], S.A. McCarroll[2,3], P.L. Auer[10,11], L.A. Lange[7], G. Lettre[4,5], A. Correa[9], C. Jaquish[12], H. Kang[8], G. Abecaisis[8], C. Laurie[1], J.D. Smith[1], D.A. Nickerson[1], R. Naik[6], J.G. Wilson[9], NHLBI TOPMed Project.* 1) University of Washington, Seattle, WA, 98195, USA; 2) Harvard Medical School, Boston, Massachusetts, USA; 3) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Université de Montréal, Montréal, QC H3T 1J4, Canada; 5) Montreal Heart Institute, Montréal, QC H1T 1C8, Canada; 6) Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 7) University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; 8) University of Michigan, Ann Arbor, MI 48108, USA; 9) University of Mississippi Medical Center, Jackson, MS 39216, USA; 10) Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; 11) University of Wisconsin-Milwaukee, Milwaukee, WI 53205, USA; 12) National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.

   Heterozygosity or carrier status for two of the clinically most well-characterized recessive genetic disorders among African ancestral populations, sickle cell *HBB* rs334 (p.V6L) on chromosome 11 (sickle cell trait or SCT) and a common 3.7 kb deletion structural variant (SV) of the alpha-globin gene cluster on chromosome 16p (alpha thalassemia trait), have not been adequately characterized with respect to phenotypic consequence in large studies. Only recently was SCT associated with a 1.5- to 2-fold risk of chronic kidney disease (CKD) among African Americans. Notably, the presence of the alpha-globin 3.7kb deletion lowers the concentration of sickle hemoglobin, and thereby can modify the clinical manifestations of sickle cell disease. Whether the alpha-globin 3.7 kb deletion modifies the risk of CKD among SCT carriers is unknown. We applied the Genome STRiP multi-sample algorithm for simultaneous discovery and genotyping of structural variations to 30x whole genome sequence (WGS) generated in the NHLBI TOPMed project on 3,020 African Americans from the Jackson Heart Study (JHS). We further examined the association of SCT in the presence or absence of the 3.7 kb deletion variant on red cell and CKD phenotypes. Overall, 67% of JHS African Americans had two diploid copies of the SV, 28% were heterozygous, and 4% were homozygous for the 3.7 kb deletion, while 1% carried extra copies (3 or 4). When adjusted for age, sex, and genetic ancestry, carrying one or two copies of the 3.7 kb deletion was strongly associated with anemia and microcytic red cells (mean cell volume or MCV<80 fl) (p<0.00001). SCT was associated with a 1.8-fold increased risk of low MCV (p=0.001), regardless of alpha-globin gene status. By contrast, SCT was associated with 1.5-fold increased risk of anemia only among individuals carrying the normal diploid copy number of the alpha-globin SV (p for interaction <0.05). Similarly, among the 1,800 participants with normal diploid copy number of the alpha globin SV, SCT was associated with 2.4-fold increased risk of CKD (defined as eGFR<60mL/min); but there was no increased risk of CKD associated with SCT (relative risk=1.0) among the 876 individuals carrying at least one copy of the alpha-globin deletion (p for interaction <0.03). If validated in independent samples, the observation that a common alpha-globin deletion reduces the risk of CKD associated with SCT has potentially important implications for public health screening and therapy.

1288W

Recessive coding and regulatory mutations in *FBLIM1* underlie the pathogenesis of Chronic Recurrent Multifocal Osteomyelitis (CRMO). *A. Cox[1], B. Darbro[1], R. Laxer[2], X. Bing[1], A. Finer[1], A. Bassuk[1], P. Ferguson[1].* 1) Pediatrics, University of Iowa, Iowa City, IA; 2) Hospital for Sick Kids, Toronto, ON, Canada.

Chronic recurrent multifocal osteomyelitis (CRMO) is a rare, autoinflammatory bone disease presenting in infancy and childhood. While CRMO is characterized by painful bone lesions, it is often comorbid with psoriasis or Crohn disease and many patients have close relatives with either of the more common disorders. Some syndromic cases of CRMO are caused by truncation mutations in the interleukin-1 receptor agonist (*IL-1RA*) gene and these children respond very well to treatment with recombinant *IL-1RA*. However, most cases of CRMO are non-syndromic with an unknown genetic cause, although many patients respond to TNF-α blocking agents, implicating the pathway in the disease. Via whole-exome sequencing, we determined a rare, coding, homozygous mutation in 25 genes in a child with CRMO from consanguineous parents. Of these genes, *FBLIM1* stood out as it was the most differentially expressed gene in a mouse microarray experiment using the *cmo* mouse model of CRMO. We sequenced *FBLIM1* in 96 CRMO subjects and found a second proband with a novel frameshift mutation in exon 6 and a rare regulatory variant in the third intron. Complementary enhancer and repressor activities of a 1-kb region around the regulatory mutation were validated by luciferase assay in SaOS2 cells and Thp1 macrophages, as was the effect of the mutation on regulatory activity. We conclude that mutations in *FBLIM1* are implicated in CRMO and autoinflammatory disease and our finding suggest that CRMO is a disorder of imbalanced bone remodeling.

1289T

Design of diagnostic algorithm for Primary Immune Deficiencies and validation of NGS panel: From traditional immunophenotyping and polymorphism studies to the next generation sequencing. *A. Bisgin[1], M. Yilmaz[2].* 1) Department of Medical Genetics, Cukurova University Faculty of Medicine, Adana, Turkey; 2) Department of Pediatric Allergy & Immunology, Çukurova University Faculty of Medicine, Adana, Turkey.

Objective: Primary immune deficiency disorders (PIDs) are a group of diseases with profound defects in immune cells. The traditional diagnostics have evolved from clinical evaluation, flow cytometry, western-blotting and Sanger-sequencing focus on small groups of genes. However, this is not sufficient to confirm the suspicion of certain PIDs. Our innovative approach to diagnostics outlines the algorithm for PIDs and the clinical utility of immunophenotyping with a custom-designed gene panel. **Methods:** We have designed a diagnostic algorithm based on polymorphism and flow cytometry studies. Metaanalysis were used to select the genes associated with PIDs and spesific immune phenotypes for NGS panel development. Then, the test panels for severe-combined-immune-deficiency (SCID), common-variable-immune-deficiency (CVID) and chronic-granulomatous-disease (CGD) via next-generation-sequencing by MiSeqSystem (Illumina) were customized. In-silico analysis for mutations was carried out using SIFT, Polyphen2 and Mutation-Taster. Clinical evaluation and primary consideration for testing was done by experienced medical geneticist in immunogenetics. **Results:** NGS identified more than 15000 sequence variants. The causative mutation was identified in 52% of PIDs (78% efficiency on CGD) with 20% of novel mutations detected. In the validation study, NGS and Sanger-sequencing were 100% concordant for the variants across all genes selected from polymorphism studies. Based on these results, this new algorithm including immune phenotyping and NGS for PIDs were started to use for clinical assessments. **Conclusion:** This study provides a thorough validation of diagnositc algorithm and indicates that still the traditional researches can be used to collect significant information related to design of most current diagnostics. The benefits of such testings are for diagnosis, prevention including the prenatal and pre-implantation diagnosis, prognosis, treatment, and research.

1290F

Genome-wide association study identifies a sequence variant in *HORMAD2* associated with tonsillectomy. *B. Feenstra[1], P.M. Bager[1], X. Liu[1], H. Hjalgrim[1], E.A. Nohr[2], D.M. Hougaard[3], F. Geller[1], M. Melbye[1,4,5].* 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Research Unit of Gynaecology & Obstetrics, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark; 3) Danish Centre for Neonatal Screening, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, Denmark; 4) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 5) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

Inflammation of the tonsils is a normal response to infection, but some individuals experience recurrent, severe tonsillitis and massive hypertrophy of the tonsils in which case surgical removal of the tonsils may be considered. To identify genetic variants underlying an exaggerated tonsillar immune response, we used tonsillectomy information from Danish health registers and carried out a genome-wide association study comprising 1,464 patients and 12,019 controls of Northwestern European ancestry, with replication in an independent sample set of 1,567 patients and 1,367 controls. The variant rs2412971, intronic in *HORMAD2* at chromosome 22q12, was significantly associated with tonsillectomy (OR = 1.22; P = 1.48 × 10^{-9}) and is highly correlated with SNPs previously found to be associated with IgA nephropathy, Crohn's disease (CD) and early-onset inflammatory bowel disease (IBD). The risk allele for tonsillectomy corresponded to increased risk of IgA nephropathy and decreased risk of CD and IBD. We further performed lookup analyses of the top SNP for outcomes related to tonsillectomy in the combined discovery and replication sample and found that rs2412971 was associated with acute tonsillitis (OR = 1.19; P = 7.82 × 10^{-4}), chronic disease of the tonsils (OR = 1.19; P = 2.32 × 10^{-6}) and appendectomy (OR = 1.18; P = 1.13 × 10^{-3}). Further functional investigation is required to illuminate whether the molecular mechanisms underlying the genetic associations at this locus involve general lymphoid hyper-reaction throughout the mucosal-associated lymphoid tissue (MALT) system.

1291W

A cancer-related mucin 1 splice site variant regulates IgG antibody response against polyomaviruses. *C. Hammer[1,2,3], P. Scepanovic[1,2], T. Waterboer[4], M. Pawlita[4], H. Ehrenreich[3], J. Fellay[1,2].* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen, Germany; 4) Division of Molecular Diagnostics of Oncogenic Infections, Infections and Cancer Program, German Cancer Research Center, Heidelberg, Germany.

Polyomaviruses have been shown to play a role in human diseases, mainly in immunocompromised individuals. However, several members of this family could not yet be reliably associated with specific pathologies. One way to obtain more insight into possible mechanisms is to identify human genes that play a role in the immune response to infection. The magnitude of the human antibody response to viral antigens is highly variable, and a significant part of this variation is heritable and can be attributed to differences in the host genome. To explore the impact of human genetic variation on humoral immunity against polyomaviruses, we performed genome-wide association studies of immunoglobulin G (IgG) response to KIPyV, WUPyV, HPyV6, and HPyV7, in 2,363 immunocompetent adults. Seroprevalence rates for all viruses were between 64% and 96%. Compatible with previous results for a wide range of viruses, significant associations were found in the HLA class II locus on chromosome 6, implicating different *HLA-DRB1* haplotypes in the regulation of anti-polyomavirus IgG levels. In addition, we discovered a genome-wide significant association signal on chromosome 1, located in the mucin 1 (*MUC1*) gene. The T allele of rs4072037 was found to be strongly associated with increased IgG response to polyomavirus infection (P = 6.6 x 10^{-13}). The same variant has been been implicated before in gastric cancer and *H.pylori* positive gastritis. Using Oxford Nanopore MinION full-length sequencing of *MUC1* short isoforms, amplified from gastric tissue RNA obtained from the Genotype-Tissue Expression Project (GTEx), we confirmed that rs4072037 affected splice site selection in intron 1, and showed that it had no further impact on additional splicing patterns. Functional follow-up analyses investigating the potential link between *MUC1* and polyomaviruses are currently ongoing.

1292T

QTL mapping of cytokine responses to LPS in peripheral blood leukocytes (PBLs). *M. Stein[1], C. Hrusch[2], C. Igartua[1], A. Sperling[2], C. Ober[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, Section of Pulmonary and Critical Care Medicine, and the Committee on Immunology, University of Chicago, Chicago, IL.

Inter-individual variation in gene expression has been extensively studied across tissues, populations, and treatments, but relatively few studies have explored genome-wide variation on protein levels or response to treatments. We undertook this study to identify genetic variants that influence cytokine levels after PBL stimulation with lipopolysaccharide (LPS), a component of the Gram-negative bacterial cell wall, and whose exposure levels are associated with asthma risk. Whole blood samples were collected from 107 South Dakota Hutterites. Cells were incubated with media+LPS in a closed system (TruCulture®, Myriad RBM) for 30 hours. 25 cytokines were measured in the supernatants in duplicate using the Milliplex Th17 Magnetic Bead Panel (EMD Millipore); 17 cytokines had measurements within the range of detection in >70% individuals and were retained for analysis. We performed a genome-wide cytokine QTL study with ~5M variants (MAF>5%) imputed from Hutterite whole genome sequences using a linear mixed model (GEMMA). We included kinship as a random effect to correct for relatedness, and sex and age as fixed effects. Only interleukin 10 (IL-10), IL-12 (p70) and IL-23 responses were associated with SNPs at a Bonferroni-corrected threshold ($P=3.9 \times 10^{-7}$, after pruning for LD). SNPs on chromosomes 3p and 22q were associated with IL-10 response, on chromosome 3q with IL-12 response, and SNPs on chromosomes 4q and 9q were associated with IL-23 response. Notably, an intergenic SNP on chromosome 9 that was associated with IL-23 response (rs4476017, $P=3.5 \times 10^{-7}$) is located in a region with histone marks consistent with enhancer activity in peripheral blood neutrophils and monocytes (Epigenome Roadmap). One other SNP associated with IL-23 response on chromosome 4 (rs75627478, $P=1.76 \times 10^{-7}$) is in an intron of the glycine receptor alpha 3 (*GLRA3*) gene and in LD ($r^2=0.66$ in the Hutterites) with a SNP that is also in an enhancer region in peripheral blood neutrophils and monocytes. These results suggest involvement of regulatory regions in cytokine response to LPS, and identifies potential candidate SNPs for immune-mediated diseases. Funded by R01 HL085197.

1293F

Evidence for a heritable response to infection for immunological traits associated with Chagas disease. *N.B. Blackburn[1,2], R. Corrêa-Oliveira[3], V.P. Diego[1], J.M. Peralta[1], J. Blangero[1], J.L. VandeBerg[1], S. Williams-Blangero[1].* 1) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX; 2) Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia; 3) Centro de Pesquisas Rene Rachou, FIOCRUZ, Belo Horizonte, MG, Brazil.

Chagas disease (CD) is a neglected tropical disease caused by infection with *Trypanosoma cruzi*, a parasite transmitted by triatomine bugs. Despite successful vector control programs in areas where *T. cruzi* infection and CD are endemic, CD is still a leading cause of heart disease in Latin America, and WHO estimates that 8-10 million worldwide are infected with *T. cruzi*. Individuals who are seropositive for *T. cruzi* infection display significant variation in progression of CD. Some individuals remain asymptomatic throughout their lives, while others develop cardiac disease or gastrointestinal involvement. Research from our group previously has demonstrated that seropositivity is heritable in a large family cohort from the Posse region of Goiás, Brazil. Chagas disease is endemic to the rural Posse region of Brazil, with approximately 50% of the adult population being seropositivity for *T. cruzi*. Our current work in this cohort involves the assessment of six immunological traits (the cytokines IL-2, IL-4, IL-6, IL-10, TNFα, and IFNγ, measured in serum) and their relationship to seropositivity. The focal population consists of 1417 study participants (50.5% male). 83% of the study participants are part of a seven generation pedigree of 2572 individuals with 65% of the assayed individuals being seropositive for *T. cruzi*. Our hypothesis was that by modelling seropositivity as an 'environment' we would be able to detect genotype-by-environment interactions, indicating a heritable response to *T. cruzi* infection. The pedigree design enabled us to observe the same genotypes in relatives who were and were not seropositive (i.e., in each 'environment'). Analyses were conducted in SOLAR. Cytokine measurements were residualized with age, sex (and their interactions), and seropositivity as covariates, and inverse normalized to prevent non-normal distribution errors. All traits showed significant heritabilities; they ranged from 13% for IFNγ ($P=7.02 \times 10^{-4}$) to 27% for IL-2 ($P=2.5 \times 10^{-12}$), indicating a genetic basis to variation in these traits across the cohort. Further, IL-4 and IFNγ exhibited significant genotype-by-environment interactions. IL-4 displayed evidence of different variances in each environment ($P=0.03$), while different genes influence IFNγ levels in each environment ($P=0.04$). Our findings demonstrate that cytokine response to infection is itself heritable in the Posse Family Study population.

1294W

Genome wide association study for identification of genetic suscep-tibility loci for meningeal tuberculosis. *E. Png[1,9], T.T.T. Nguyen[2,3,9], Z. Li[4], K.S. Sim[4], D.B. Nguyen[5], T.T.H. Chau[6], T.H.M Nguyen[6], M. Caws[7], C.C. Khor[4], D. Heemskerk[2,3], G. Thwaites[2,3], M.L. Hibberd[1,9], S.J. Dunstan[8,9].* 1) Infectious Diseases, Genome Institute of Singapore, Singapore, Singapore; 2) Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 764 Vo Van Kiet, District 5, Ho Chi Minh City, Vietnam; 3) Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, OX37LJ, United Kingdom; 4) Human Genetics, Genome Institute of Singapore, Singapore; 5) Pham Ngoc Thach Hospital for Tuberulosis and Lung Disease, District 5, Ho Chi Minh City, Vietnam; 6) Hospital for Tropical Diseases, 764 Vo Van Kiet, District 5, Ho Chi Minh City, Vietnam; 7) Liverpool School of Tropical Medicine, Liverpool, UK; 8) Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Australia, 3010; 9) Joint authors, who equally contributed to this work.

   Tuberculosis (TB) remains a leading cause of infection-associated mortality, with close to 10 million new cases and 2 million deaths annually. Although *Mycobacterium tuberculosis* (MTB) is thought to have infected around a third of the world's population, only 10% of those infected develop active disease during their lifetime. Despite apparently similar exposures to MTB, there are substantial variation in progression to TB disease. Furthermore, while a majority of active TB leads to pulmonary disease (PTB), in a minority of cases, extra-pulmonary TB can develop in other organs, with tuberculous meningitis (TBM) the most severe form, occurring in approximately 1% of all TB cases. Although PTB is largely curable, TBM is known to have poor outcomes and kills or severely disables about half of the people affected. The underlying mechanisms leading to TBM are largely unknown, but their elucidation will aid the development of more effective interventions. Recent genome wide association studies (GWAS) have convincingly shown SNPs associated with PTB, this has encouraged us to conduct GWAS for genes relevant to TBM disease, since TBM and PTB may have distinctive mechanisms. Following imputation from the Illumina Human660W-Quad and Infinium OmniExpress-24 arrays, we obtained genotypes for 4 million SNPs that passed genotype cluster plot inspection or imputation call threshold ≥ 0.9 and quality control filters; for call rate ≥ 95%, minor allele frequency ≥ 5%, Hardy Weinberg Equilibrium (in controls) P ≥ 1 x 10[-7], in 441 TBM and 2000 population controls from Vietnam. We observed minimal genome-wide inflation of the single-SNP association test statistics (λGC= 1.03), and found an excess of 80 significant *P*-values (*P* ≤ 1 x 10[-5]) at the tail end of the distribution, suggesting that at least some of these significant *P*-values could represent true associations. We next plan to replicate these associations by genotyping these variants in an independent sample set from Vietnam as well as other populations.

1295T

Host genetic determinants of HIV-1 broadly neutralizing antibody induc-tion. *C.W. Thorball[1,2], C. Hammer[1,2], P. Rusert[3], R.D. Kouyos[3,4], M. Schanz[3], M. Huber[3], H.F. Günthard[3,4], A. Trkola[3], J. Fellay[1,2], the Swiss HIV Cohort Study.* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Institute of Medical Virology, University of Zurich, Zurich, Switzerland; 4) Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland.

   The induction of broadly neutralizing antibodies (bnAbs) is considered essential in the development of an effective vaccine against HIV-1 due to their potent activity against divergent HIV-1 strains. During the course of natural infection, 10-25% of all individuals infected with HIV-1 produce antibodies with limited neutralization breadth, while only 1% of all infected individuals become capable of producing highly potent antibodies with a broad, elite neutralization activity. But why these highly potent bnAbs are only successfully induced in a fraction of HIV-1 infected individuals remains unknown. To identify host genetic variants associated with the ability to produce HIV-1 broadly neutralizing antibodies, we performed whole-exome sequencing (WES) of 63 patients with elite neutralization activity and 273 patients with no or weak cross neutralization activity. Neutralization breadth in patients was identified in a stringent 8-virus panel neutralization screen of 4484 HIV-1 positive plasma samples with matched disease progression in the Swiss HIV Cohort Study (SHCS) and the Zurich Primary HIV Infection Study (ZPHI). Variant calling was performed using GATK (v3.5) following the GATK Best Practices. In order to further improve data quality, the dataset was imputed with PBWT and the Haplotype Reference Consortium reference panel (r1.1). Only variants with INFO score above 0.8 and patients of Caucasian ancestry were retained for further analyses. Variants associated with the ability to produce highly potent bnAbs were identified using the recently developed logistic mixed model association test GM-MAT. Significant genome-wide associations (p < 5x10e-8) were identified for 9 variants, despite the limited power of this study, in *HLA-DRB1* (rs35445101), *PABPC3* (rs11619791, rs9511646), *EPG5* (rs78940552, rs73436059, rs73436053, rs3744999, rs56159632) and *NUP85* (rs193149375). Analysis of GTEx data showed that in whole blood, rs35445101 is an eQTL affecting in particular the other HLA class II genes *HLA-DQA2* (p = 2.8e-53) and *HLA-DRB6* (p = 1.4e-44) indicating that the contribution of the HLA class II locus to the induction of bnAbs might go beyond *HLA-DRB1*. These results, although preliminary, identify multiple host genetic factors influencing the ability of HIV-1 patients to successfully produce highly potent bnAbs.

1296F

Association analysis identified susceptible alleles to tuberculosis based on imputed HLA class I and II alleles. *L. Toyo-oka[1,2,3], S. Mahasirimongkol[1,3], H. Yanai[2,3,4], S-S. Khor[2], T. Mushiroda[3,5], N. Wichukchinda[1,3], S. Wattanapokay-akit[1,3], N. Yamada[6], N. Smittipat[3,7], T. Juthayothin[7], P. Palittapongarnpim[3,8], S. Nedsuwan[3,9], P. Kantipong[9], A. Takahashi[10], M. Kubo[11], P. Sawanpanyalert[12], K. Tokunaga[2,3].* 1) Medical Genetic Center, Medical Life Sciences Institute, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Tokyo, Japan; 3) AMED/JICA, SATREPS; 4) Fukujuji Hospital, Japan Anti-Tuberculosis Association (JATA), Kiyose, Japan; 5) Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Research Institute of Tuberculosis, Japan Association of Tuberculosis, JATA, Kiyose, Japan; 7) Tuberculosis Research Laboratory, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathum Thani, Thailand; 8) Department of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand; 9) Chiangrai Prachanukroh Hospital, Ministry of Public Health, Chiang Rai, Thailand; 10) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 11) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 12) Food and Drug Administration, Ministry of Public Health, Nonthaburi, Thailand.

Introduction Tuberculosis (TB [MIM 607948]) is chronic respiratory infection with high burden to global health. Human leukocyte antigen (HLA) is significantly associated with TB in Icelandic population, where *HLA-DQA*03:01* and *HLA-DRB1*12:02* were associated with TB. *HLA* class I was detected to show significant association with TB caused by specific lineage of the pathogen of TB, *Mycobacterium tuberculosis* (MTB). This study aims to further investigate associations between HLA class I and II alleles imputed from GWAS data with TB in Thais. Overall, 1,554 TB cases and 834 controls were selected and GWAS data was obtained by Illumina Human610-Quad BeadChip (724 cases and 834 controls) and HumanOmniExpress BeadChip (830 cases). From these TB cases, 951 MTB isolates were classified by spoligotyping based on lineages of MTB into Beijing, EAI and non-Beijing non-EAI. Imputation of *HLA-A, -B, -C, -DPB1, -DQA1, -DQB1* and *-DRB1* were conducted by HIBAG. After quality control processes and outliers exclusion based on principal component analysis, 1,792 samples were included for the association analysis. Two cases groups were defined as discovery and replication set, and the controls were utilized as common control. HLA alleles with frequency higher than 5% were subjected to Fisher's exact test and *P*-values (*Pc*) was empirical significant level corrected by Bonferroni correction. Results We could not detect association between HLA class I and TB in this group of samples. However, *HLA-DRB1*14:01* showed significant protective association in discovery set (*Pc*: 1.04E-02 and odds ratio (OR): 0.4 (0.2-0.73) with confirmation in replication set (*Pc*: 1.93E-04 and OR: 0.27(0.12-0.54)) in positivity model. The protective effect is stronger in TB cases with Beijing lineage comparing with TB cases with the other lineages. Disussion This result suggested that HLA class II was more critical than HLA class I associated with TB and consideration for lineage information was important to reveal pathogenesis of TB. HLA class II alleles interacting with lineage based variations might be important in host-pathogen interactions in TB.

1297W

Associations between mitochondrial variants and asthma in a multi-ethnic cohort. *K.M. Magnaye[1], C. Igartua[1], J. Hartiala[2], Z. Chen[2], H. Allayee[2], F.D. Gilliland[2], C. Breton[2], D.L. Nicolae[1,3], C. Ober[1], EVE Consortium.* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Preventive Medicine, University of Southern California, Los Angeles, CA; 3) Medicine and Statistics, University of Chicago, Chicago, IL.

Asthma is a common chronic inflammatory disease of the airways, affecting over 230 million people worldwide. One of the strongest risk factors for childhood onset asthma is maternal asthma. One possible explanation for this maternal risk may be the presence of asthma-associated variation in the maternally-inherited mitochondrial genome. To test this hypothesis, we performed meta-analyses of mitochondrial single variants, as well as with gene- and complex-based analyses, on asthma risk in participants in the EVE Consortium on the Genetics of Asthma, which consists of 3,611 asthmatics and 3,623 non-asthmatic controls of European, African American, Latino and African Caribbean ancestries from 7 studies. Subjects were genotyped using the Illumina Infinium HumanExome BeadChip array, which yielded 160 mitochondrial markers after quality control checks; 41 variants had MAF >1%. To evaluate the effects of individual variants with MAF >1% on asthma risk, we used a linear mixed model (GEMMA), correcting for sex and the first 10 autosomal principal components (PCs) within each study, and combined the effects across studies within each ethnic group in meta-analyses. For the meta-analyses, we required each variant to be present and with a MAF >1% in at least two groups. After adjusting for the number of variants tested, one missense mutation in each of the *ND3* and *ATP6* genes was significant in the Latinos (OR =1.04, p = 3.63x10[-5]; OR = 1.04, p = 8.68x10[-5]) and one missense mutation in the *CYB* gene was significant in the European Americans (OR = 1.04, p = 6.20x10[-4]). We next performed gene-based and complex-based tests of all 160 variants, using SNP-set Kernel Association Test (SKAT), also adjusting for sex and the first 10 autosomal PCs, and combined the results across groups using MetaSKAT. Only the D-loop regulatory region, defined by three variants, was associated with asthma in the African Caribbean sample after correcting for 15 tests (SKAT p = 2.81 x 10[-3]). No mitochondrial complexes were significantly associated with asthma in any of the meta-analyses. Our results suggest that associations with mitochondrial variants and genes are ethnic specific. Moreover, this study did not reveal any large-effect functional mitochondrial variants or genes in the etiology of asthma or support the hypothesis that the risks associated with maternal asthma are due to genetic variation inherited in the mitochondrial genome. Supported by HL101651. .

1298T

An inframe deletion in *NUDT15* is a common cause of thiopurine-induced myelosuppression. *J.W. Harrison[1], G. Heap[1], G. Jones[1], C. Bewshea[1], P. Irving[2], C. Mowat[3], J. Simmons[4], M. Parkes[5], F. Cummings[6], S. Sebastian[7], R. Chaudhry[8], R. Russell[9], C.W. Lees[10], T. Creed[11], R. Fedorak[12], C. Hawkey[13], M.J. Weedon[1], T. Ahmad[1].* 1) RILD Building, University of Exeter Medical School, Exeter, Devon, United Kingdom; 2) Guy's and St Thomas' Hospital; 3) Ninewells Hospital & Medical School, Dundee; 4) Royal Berkshire NHS Foundation Trust; 5) Addenbrookes NHS Foundation Trust; 6) Southampton NHS Foundation Trust; 7) Hull and East Yorkshire Hospitals NHS Trust; 8) Watford General Hospital Trust; 9) Royal Hospital for Sick Children, Glasgow; 10) Western General Hospital, Glasgow; 11) Bristol Royal Infirmary; 12) University of Alberta Edmonton, Alberta; 13) Nottingham Digestive Diseases Centre.

The thiopurines azathioprine and mercaptopurine are common medications used in the treatment of inflammatory bowel disease and other autoimmune disorders. Despite their widespread use 40% of patients fail to tolerate these medications. Previously identified variants in the TPMT gene have been shown to strongly predispose individuals to this serious condition. However, despite widespread adoption of TPMT screening, up to 10% of patients still experience severe myelosuppression necessitating drug withdrawal. Common variants in the *NUDT15* gene have recently been reported to strongly predispose individuals in Asian populations to thiopurine-induced myelosuppression (e.g. Japanese populations: R139C OR = 5.82). The reported variants are, however, rare within European populations (MAF= 0.0046). Here we performed exome sequencing in 290 cases of thiopurine-induced myelosuppression and 767 thiopurine tolerant controls all of Caucasian European-descent to identify novel variants associated with myelosuppression. We performed single variant genome-wide association studies for 843,893 variants and identified a novel 6-base pair inframe deletion in the NUDT15 gene which was present in 14 cases and only 1 control (OR=38.8, P= 1.226×10^{-07}, 5.08-1631.3, 95% confidence interval, MAF=0.014). We confirmed sequence accuracy by Sanger sequencing. These data suggest this novel variant is strongly associated with the development of this potentially life-threatening reaction to thiopurine treatment. In conclusion we have identified a novel variant in *NUDT15*, a gene already shown to be important in the metabolism of thiopurines. This novel variant is strongly associated with the development of thiopurine-induced myelosuppression. *NUDT15* variants are as important as *TPMT* variants in determining whether a patient develops myelosuppression during thiopurine treatment in European populations. The addition of this and other NUDT15 variants to established genetic screening will further improve the personalising of thiopurine treatment and thus increasing its effectiveness in European populations.

1299F

Pharmacogenetics to inform global drug development: Utilizing data from Western countries to predict PK and tolerability in Eastern countries. *L. Parham[1], Z. Xue[1], S. Ford[2], Y. Lou[3], D. Margolis[4], A. Hughes[1], W. Spreen[4], P. Patel[4].* 1) Genomic Medicine, PAREXEL, Durham, NC; 2) Quantitative Clinical Development, PAREXEL, Durham, NC; 3) Biostatistics, PAREXEL, Durham, NC; 4) ViiV Healthcare, Research Triangle Park, NC.

In global drug development, extensive pharmacokinetic (PK) data from early phase studies may be limited to Western countries. The ability to extrapolate what PK profiles might be expected when expanding into other racial populations and ethnicities is of great value. Utilizing available subject data and knowledge of drug metabolism/disposition pathways, along with pharmacogenetics (PGx), can inform expectations of PK effects due to regional genetic differences. Many genes implicated in drug metabolism have known genetic polymorphisms that vary by ancestry which can impact enzyme activity, and may lead to clinically important differences in exposure. UGT1A1 is responsible for the glucuronidation of multiple medicines and known alleles result in decreased UGT1A1 enzyme activity. Carriage of known alleles varies by ancestry group with *28 being common among all populations, *37 common among those with African ancestry, and *6 being commonly observed in East Asians. The impact of reduced UGT1A1 activity varies by medicine. Cabotegravir (CAB) is an integrase strand transfer inhibitor in development for the treatment and prevention of HIV-1 infection and is primarily metabolized by UGT1A1. An analysis was conducted utilizing 347 subjects who received CAB across 4 phase I and 2 phase II clinical studies to evaluate the genetic effects of *UGT1A1* genotypes on oral PK parameters, including C_τ (plasma concentration at end of the dosing interval). The majority of subjects available were of European, or African American/African heritage; less than one percent were self-reported as Asian. Genetically predicted UGT1A1 activity (determined by carriage of *28, *37 and/or *6) was significantly associated with CAB C_τ (p=4.89×10^{-11}). Comparing subjects with a low vs. normal genetically predicted UGT1A1 activity, a 1.5-fold increase in mean C_τ was observed. Increased CAB exposure in carriers of UGT1A1 reduced function polymorphisms is not clinically significant based on the cumulative safety data available and no dose adjustment in these individuals is required. Three subjects carried *6, the reduced function UGT1A1 allele that is commonly observed in Asians. While sample sizes were limited, CAB C_τ values observed for these 3 subjects were within the range of others in the same predicted activity classes due to other functional alleles. This PGx analysis provided useful information to support accelerated clinical development timelines in Eastern countries.

1300W

HLA imputation from whole genome sequencing data and association with disease severity and treatment response to Golimumab in rheumatoid arthritis. *K.A. Standish[1,2], C.C. Huang[3], S. Lamberth[3], K.A. Campbell[3], M. Curran[3], N.J. Schork[1,2].* 1) Human Biology, J. Craig Venter Institute, La Jolla, CA; 2) University of California, San Diego; 3) Janssen R&D, LLC, Springhouse, PA.

The highly polymorphic human leukocyte antigen (HLA) region on chromosome 6 encodes the major histocompatibility complex (MHC) proteins. These cell-surface proteins are responsible for antigen presentation to the adaptive immune system and play critical roles in host defense, tissue compatibility, and immune regulation. Polymorphisms in this region have been linked to several autoimmune diseases; for example, the HLA-DRB1 "shared epitope" harbors several polymorphisms strongly associated with susceptibility to rheumatoid arthritis. Using whole-genome sequencing (WGS) data from a cohort of moderate-to-severe RA patients in a phase 3 clinical study, we explored different computational HLA imputation methods and built a robust pipeline to determine HLA type and primary protein sequence with high confidence. We then profiled different HLA-DRB1 haplotypes by classic 4-digit type (e.g. 04:01) as well as individual and combinations of amino-acids at various polymorphic residues (positions 11, 70 - 74). Using traditional response measurements before and after treatment, as well as with a more sophisticated approach employing the use of longitudinal trial design, we recapitulated a previous observation that haplotypes associated with more severe disease states are also associated with greater treatment responses. This trend was observed for the collapsed haplotypes defined by residues 11, 71, and 74, as well as positions 70 to 74 (shared epitope) in HLA-DRB1. However, no significant association was observed between any particular haplotype and disease severity or treatment response. Our approach of using WGS data to assign patients HLA haplotypes at varying levels of resolution and linking them to clinical phenotypes provides added utility to sequencing data in an era when sequencing is increasingly becoming available to general public.

1301T

Genetic risk prediction and subtyping for generalized vitiligo. *S.A. Santorico[1,2,3], S. Paul[1], D. Yorgov[1], Y. Jin[2,4], T. Ferrara[2], R.A. Spritz[2,4].* 1) Mathematical and Statistical Sciences, University of Colorado, Denver, USA; 2) Human Medical Genetics and Genomics Program, University of Colorado School of Medicine, Aurora, Colorado, USA; 3) Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado, Aurora, Colorado, USA; 4) Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA.

Generalized vitiligo (GV) is an autoimmune disease in which white patches of skin and hair result from destruction of melanocytes. GV patients have ~20% risk of other autoimmune diseases, pointing to shared genetic risk factors and perhaps environmental triggers. In previous work, we discovered and replicated 46 autosomal GV loci via meta-analysis of three genome-wide association studies, with independent replication, consisting of a total of 4,680 vitiligo cases and 39,586 controls of European (EUR) ancestry with genotypes for 8,801,562 variants. Loci mostly encode immune regulators and apoptotic regulators, some of which are also associated with other autoimmune diseases, as well as several melanocyte regulators. Working towards our long-term goal of accurately targeting preventive measures and ensuring the use of the most effective interventions and treatment approaches, we have created a "polygenic risk score" for vitiligo, optimizing the set of variants for inclusion, and compared the predictive ability to a "major loci risk score" based on our 46 replicated autosomal GV loci. The polygenic risk score is built from estimated logistic regression coefficients, after adjustment for ancestry. Variants were included in risk scores for levels of significance, α, in {5×10^{-8}, 0.001, 0.002, ..., 0.009, 0.01, 0.02, ..., 0.09, 0.1} after LD clumping with an index variant P value of 0.1. Performance was assessed by 10-fold cross-validation. Over this range of significance levels, the maximal area under the curve (AUC) of 0.769 occurred at an inclusion α of 0.008; however, this was lower than the AUC for the major loci risk score of 0.776. As a preliminary analysis to explore disease subtypes, recursive partitioning was applied to the 46 replicated autosomal GV loci. Six loci were found to best separate cases and controls into subgroups. These loci included *HLA-A*, *PTPN22*, *CD44*, *GZMB*, *TYR*, and *MC1R*. Additional methods of clustering are being considered and will be compared to the results of recursive partitioning. Long term these will be correlated with vitiligo sub-phenotypes, occurrence/non-occurrence of other autoimmune diseases, age of vitiligo onset, and self-reported treatment histories. Work will be presented through the lens of our long-term goal: optimized prediction of vitiligo risks and subtypes to facilitate clinical application of optimal therapies based on genetic subtyping of disease.

1302F

Novel genetic variants associated with asthma exacerbations in African American youth. *P. Goddard[1], M.J. White[1], O. Risse-Adams[1,2], J. Adams[1], M.G. Conteras[1,3], D. Hu[1], C. Eng[1], A. Davis[4], E. Brigino-Buenaventura[5], M.A. LeNoir[6], M. Seibold[7], N. Thakur[1], J. Galanter[1], S.S. Oh[1], K. Bibbins-Domingo[1], M. Pino-Yanes[8,9], E. Burchard[1,10].* 1) Department of Medicine, University of California, San Francisco, San Francisco, CA; 2) Lowell Science Research Program, Lowell High School, San Francisco, CA, USA; 3) SF BUILD, San Francisco State University, San Francisco, CA, USA; 4) Children's Hospital and Research Center Oakland, Oakland, CA, USA; 5) Department of Allergy and Immunology, Kaiser Permanente Vallejo Medical Center, Vallejo, CA, USA; 6) Bay Area Pediatrics, Oakland, CA, USA; 7) Integrated Center for Genes, Environment, and Health; Department of Pediatrics, National Jewish Health, Denver, Colorado, USA; 8) Research Unit, Hospital Universitario N.S. de Candelaria, Tenerife, Spain; 9) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; 10) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA.

There are major disparities in asthma prevalence, mortality, and morbidity among racial/ethnic groups in the U.S. African Americans carry a heavier disease burden than their European and Mexican American counterparts. Asthma exacerbations, characterized by oral steroid use, and/or hospital visits, account for a significant portion of the cost and morbidity associated with asthma in the U.S. There have been limited studies exploring the genetic variation underlying susceptibility to asthma exacerbations in minority populations. It is vital that the genetic variation underlying susceptibility to asthma exacerbations in African American youth with asthma be evaluated in order to identify novel treatments and therapeutic interventions for this "at risk" population. The main aim of this study is to identify genetic variants that are associated with asthma exacerbations in a pediatric cohort of African American asthmatics.Our study population consists of asthma cases from the Study of African Americans, Asthma, Genes, and Environments (SAGE II). An asthma exacerbation was defined as the recent occurrence of at least one of the following events: use of oral corticosteroids, a hospital visit or an ER visit for an asthma related ailment. Exacerbation incidence was determined by self-reported responses to a questionnaire. LAT1 Axiom array was used for genotyping. 812 individuals and 519,236 SNPs were analyzed following quality control. Logistic regression analysis was performed to evaluate the association between SNP genotype and asthma exacerbation occurrence with adjustments for body mass index (BMI), age, gender, and global African ancestry.Logistic regression analysis revealed several novel SNPs associated with acute asthma exacerbations in African Americans. Our most significant result was rs67267432, located on chromosome 2 (OR: 1.984, p-value = 7.56×10^{-7}). Of note, another of our suggestive results on chromosome 17, rs4789054 (OR: 1.678, p-value = 1.11×10^{-5}) is located proximal to *DNAI2* (Dynein Intermediate Chain 2) encoding a lung ciliary protein. *DNAI2* is regulated by the interferon viral response system and has been implicated in other respiratory diseases. Our results failed to replicate variants previously associated with asthma exacerbations in populations of European-origin, suggesting that these may be population-specific effects. Further research is underway to evaluate our findings in the context of air pollution exposure and pathway enrichment.

1303W

Eosinophilic esophagitis risk variant at 2p23 dampens IL-13-induced calpain-14 promoter activity in a STAT6-dependent manner. *L. Kottyan[1,2], D. Miller[1], A. Maddox[1,2], A. Lynch[1], J. Lee[1], C. Schroeder[1], M. Brusilovsky[2], M. Weirauch[1,3], M. Rothenberg[2].* 1) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Eosinophilic esophagitis (EoE) is a chronic, food-driven, esophageal, inflammatory allergic disease characterized by marked mucosal eosinophil accumulation. We have recently found that, in addition to genetic risk loci for allergic sensitization, EoE susceptibility is linked to a tissue-specific genetic factor(s) at 2p23, encoding the *CAPN14* gene. In our initial studies, we showed that *CAPN14* is dynamically up-regulated as a function of EoE disease activity as well as after exposure of epithelial cells to IL-13, a critical regulator of esophageal inflammation in EoE. Patients with EoE and the 2p23 risk haplotype express decreased *CAPN14.* Furthermore, shRNA-mediated knockdown of *CAPN14* disrupts the regulation of IL-13-mediated esophageal epithelial cell inflammation. Together, these data underscore calpain-14 as a regulator of esophageal inflammation in EoE. This genetic linkage has been replicated at genome-wide significance in multiple cohorts, as well as in a recent independent study, adding credence to the importance of the 2p23 genetic association. In the current study, we performed a replication and fine-mapping study of the 2p23 locus in an additional cohort of subjects with and without EoE. Using this independent cohort, we further replicated this locus (rs76562819 $p_{meta} < 10^{-14}$, Odds Ratio=1.98) and performed fine mapping using Bayesian and frequentist strategies, identifying five genetic variants most likely to be causal for increased EoE risk. We used DNA affinity precipitation analysis and electromobility shift assays to identify proteins that differentially bound specific variants, and confirmed STAT6 binding to the three putative binding sites in the promoter and first intron. Using luciferase reporter assays, we demonstrated that the risk variant of rs76562819 was sufficient to reduce promoter activity of *CAPN14* by 50% in IL-13 treated esophageal epithelial cells - similar to the allelic decrease seen in patient biopsies. Each of the three STAT6 elements were required for the 10-fold increase in IL-13 induced promoter activity and for the 50% reduction in genotype-dependent expression. Our work establishes a candidate molecular mechanism for EoE disease etiology in which the risk variant rs76562819 at 2p23 dampens IL-13-induced calpain-14 promoter activity in a STAT6-dependent manner.

1304T

Association of *STAT4 rs7574865* polymorphism with susceptibility to systemic lupus erythematosus. *V.E. Maso[1], F.B. Barbosa[1], C.E.V. Wiezel[1], M. Simioni[2], F.R. Torres[2], M.C. Molck[2], E.A. Donadi[3], V.L. Gil-da-Silva-Lopes[2], A.L. Simões[1].* 1) Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Campinas, SP, Brazil; 3) Division of Clinical Immunology, Department of Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil.

   Systemic lupus erythematosus (SLE [MIM: 152700]) is a multisystem autoimmune disease characterized by autoantibody production and immune complex deposition in several tissues. Beyond the immunological factors related to the loss of self-tolerance, genetic factors also strongly contributed to the SLE etiology. Regarding single nucleotide polymorphisms (SNPs) implicated in SLE etiology, the *rs7574865* SNP located in the *STAT4* (MIM: 600558) gene has been observed to have a consistent association with SLE and other autoimmune diseases. Considering that ancestral composition of populations can modify the results of association tests, the study of SNPs in admixed populations can produce different even conflicting results compared with those reported. The aim of this study was to compare allelic frequencies of the *STAT4 rs7574865* G/T polymorphism between SLE Brazilian patients and healthy individuals. A total of 139 unrelated SLE patients and 156 healthy controls were genotyped by TaqMan-based quantitative PCR using the StepOnePlus instrument (Applied Biosystems). SNP allelic/genotypic frequencies and Hardy-Weinberg equilibrium (HWE) were calculated by Plink v. 1.07. Statistical analysis was performed using R v. 3.2.3. No deviation from the HWE were found in patient (p = 0.85) and control (p = 0.81) groups. The *rs7574865* T allele and TT genotype were significantly increased in SLE in comparison to the healthy controls (p = 0.0001, odds ratio [OR] = 2.0, 95% confidence interval [CI] = 1.4—3.0 and p = 0.0051, OR = 3.7, 95% CI = 1.3—11.7, respectively). G allele and GG genotype were found as a protective factor against susceptibility to SLE (p = 0.0001, OR = 0.5, 95% CI = 0.3—0.7 and p = 0.0007, OR = 0.4, 95% CI = 0.3—0.7, respectively). These findings demonstrated that the association of *STAT4 rs7574865* SNP with SLE were effectively replicated in the Brazilian population. Our data highlighted the role of *STAT4* in SLE pathogenesis regardless of the studied population. Support: FAPESP (2013/17062-9, 2011/23794-7), CNPq (312547/2009-9, 304455/2012-1), and FAEPA.

1305F

Genetic variants of SMADs affect susceptibility to ulcerative colitis through their alternative expressions in Japanese patients. *S. Taniguchi[1], S. Suzuki[1], T. Inamine[1], A. Yamashita[1], S. Fukuda[1], S. Kondo[1], H. Isomoto[2], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch, Nagasaki, Nagasaki, Japan; 2) Division of Medicine and Clinical Science, Department of Multidisciplinary Internal Medicine, Tottori University Faculty of Medicine, Yonago, Japan.

   PURPOSE Inflammatory bowel diseases, comprised of ulcerative colitis (UC) and Crohn's disease, are attributed to inappropriate inflammatory response in the intestinal epithelia of patients. TGF-b/SMAD signals play a key role in differentiation of naive CD4+ T cells to Th17 cells or regulatory T (Treg) cells. In order to identify genetic determinants of UC, we investigated an association between susceptibility to UC and polymorphisms of SMADs in the Japanese population.   METHODS The study subjects consisted of 108 patients with UC and 199 unrelated healthy control subjects. A total of 21 tag single nucleotide polymorphisms (SNPs) in four genes (*SMAD2, SMAD3, SMAD4,* and *SMAD7*), which are involved in the TGF-b/SMAD signaling pathway, was genotyped by PCR-restriction fragment length polymorphism, PCR-direct DNA sequencing, or PCR-high resolution melting curve analysis. The frequencies of alleles and genotypes were compared between UC patients and control subjects by chi-square test or Fisher's exact test in three inheritance models: the allele, the minor allele dominant, and the minor allele recessive models. In addition, in order to investigate whether the risk allele of rs7229678 in *SMAD4* affects its gene expression, the DNA sequence including this SNP site was integrated into pGL3-promoter plasmid and reporter gene assay was performed in Jurkat and Caco-2 cell lines.   RESULTS Five SNPs of *SMAD2*, one SNP of *SMAD3*, two SNPs of *SMAD4*, and one SNP of *SMAD7* showed a close and significant association with susceptibility to UC. Moreover, the reporter gene assay demonstrated that the transcriptional activity of the risk allele of rs7229678 in *SMAD4* was lower than that of the non-risk allele in both Jurkat and Caco-2 cell lines.   CONCLUSION Genetic variants of SMAD signaling molecules in the TGF-b/SMAD pathway may disturb the signal transduction and alter the balance of differentiation to Th17 or Treg cells, thereby leading to dysregulation of immune response and eventually resulting in the development of UC.

1306W

Candidate genes predisposing to primary biliary cirrhosis/cholangitis. *G. Macintyre[1,2], D.A. Bugbee[2], P. Nuin[1], P.M.K. Gordon[4], B. Eksteen[5], E. Carpenter[2], J. Reeve[3], G. Wong[2], A. Mason[2].* 1) Medical Genetics, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada; 2) Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada; 3) Laboratory Medicine and Pathology, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada; 4) Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada; 5) Department of Medicine, University of Calgary, Calgary, Alberta, Canada.

Primary Biliary Cirrhosis/Cholangitis (PBC) is a rare liver disease (prevalence 1:2,500 to 1:100,000) that is more common in women, particularly in middle-age. However, several cases of females with early age of onset (<22 years) have now been recognized with more severe and progressive disease. PBC is characterized by bile duct destruction and is considered an autoimmune disease as 95% of patients make anti-mitochondrial antibodies (AMA) that are highly specific for diagnosis. Our previous studies indicate that a genetic predisposition and specific beta-retroviral infection may result in PBC. We have performed a whole exome analysis on a single family: affected mother and daughter, healthy father and healthy daughter. Both affected mother and daughter exhibited severe symptoms, the daughter in her teens, and each required a liver transplant. Both mother and daughter have developed recurrent PBC in the allograft and the daughter has required anti-retroviral therapy to prevent progressive disease. A preliminary analysis of variants displaying a dominant mode of inheritance implicates several rare variants including a deletion in the angiomotin gene (*AMOT*) as a possible candidate for PBC. AMOT is important in angiogenesis, hepatic epithelial cell proliferation and migration, and is implicated in viral infection. We also analyzed known GWAS variants from previous studies of rare liver diseases and have identified many that are homozygous and follow a recessive mode of transmission. These include several genes including fucosyltransferase-2 (*FUT2*), and genes encoding toll-like receptors (*TLR2, TLR3*). The candidate genes provide support for the gene-plus-infection hypothesis and implicate possible alterations in key immune pathways in PBC in this family.   .

1307T

Interplay of genetic and environmental factors, N-glycosylation of plasma immunoglobulin G (IgG) and immune traits in autoimmune thyroid diseases (AITD). *T. Martin[1], M. Simurina[2], M. Mangino[1,3], M. Rydlewska[4], M. Roederer[5], M. Ząbczyńska[4], I. Ugrina[2,6], I. Trbojević-Akmačić[2], A. Burri[7,8], A. Hubalewska-Dydejczyk[9], A. Lityńska[4], F. Nestle[3,10], M. Tadijanović[11], M. Trofimiuk-Müldner[9], R. Turek-Jabrocka[9], J. Walsh[12,13], E. Pochec[4], S. Wilson[1,12,13], T. Spector[1], G. Lauc[2,6].* 1) King's College of London, London, United Kingdom; 2) Genos, Glycoscience Research Laboratory, Zagreb, Crotia; 3) NIHR Biomedial Research Centre at Guy's and St. Thomas's NHS Foundation Trust, London, SE1 9RT, UK; 4) Department of Glycoconjugate Biochemistry, Institute of Zoology, Jagiellonian University, Krakow, Poland; 5) ImmunoTechnology Section, Vaccine Research Center, NIAID, NIH, Bethesda, MD 20892, USA; 6) Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia; 7) Health and Rehabilitation Research Institute, Auckland University of Technology, Auckland, New Zealand; 8) Waitemata Pain Service, Department of Anaesthesia and Perioperative Medicine, North Shore Hospital, Auckland, New Zeeland; 9) Endocrinology Clinic, University Hospital, Krakow, Poland; 10) Cutaneous Medicine Unit, St John's Institute of Dermatology, King's College, London, SE1 9RT, United Kingdom; 11) General Hospital "Dr. Josip Benčević", Slavonski Brod, Croatia; 12) Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; 13) School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia.

Autoimmune thyroid disease (AITD) is one of the most common autoimmune diseases, affecting about 5% of Europeans, with a heritability of 75%. AITD is caused by abnormal thyroid autoantibodies (e.g. thyroperoxidase antibody: TPOAb) and T cells that damage the thyroid cells. Variation in the levels of immune traits, immunoglobulins (Ig) and glycans in the blood are also strongly heritable and can play key roles in AITD, driven by underlying genetic variation. Our aims were (1) to assess the association of plasma IgG glycome composition and immune cell traits with AITD or TPOAb levels (2) to explore the interaction of IgG glycans with immune traits and (3) to explore the role of genetic and environmental factors on these potential interactions. AITD was assessed clinically. TPOAb was measured using chemiluminescent immunoassays and glycans using hydrophilic interaction chromatography-UPLC. Immunophenotyping was done by flow-cytometry. Genome-wide scans on different omics for AITD status and interaction networks were conducted in the TwinsUK cohort (4,473 twins for 76 IgG glycans, 487 twins for 23,485 immune traits). Glycan findings were replicated in 2 independent cohorts and validated by lectin analysis on a subset (11 Hashimoto's disease samples and 13 healthy donors). We identified 10 glycans significantly associated with AITD in the TwinsUK samples, of which 7 replicated. These observed associations indicated a depletion of core fucose on IgG N-glycans in AITD that was then validated. No significant association between any blood immune trait and AITD could be detected. Tobacco smoking appeared to enhance IgG core afucosylation and to alter the associations of glycans with natural killer (NK) cells and B IgG cells. However, no shared SNPs between these parameters could be found. Our study shows for the first time a depletion of core fucose in the IgG glycome in AITD patients with a potential functionally relevant interaction with NK cells independent of age and sex. IgG core afucosylation in AITD seems unlikely to be directly triggered by common genetic variation. Smoking however appears to enhance the core afucosylation in AITD and modify the associations between IgG glycans and NK cells as well as B IgG cells. Further studies on the glycosylation of TPOAb and thyroglobulin antibodies and their interactions with immune traits and thyroid cells will help elucidate their roles in thyroid disease and to determine their potentials in next-generation therapeutics.

1308F

GWAS of Danish Hirschsprung disease using pathology confirmed cases detects novel associations and makes case/control prediction with high accuracy. *J. Fadista[1], M. Lund[1], F. Geller[1], P. Nandakumar[2], S. Chatterjee[2], L. Carstensen[1], J. Grauholm[3], D. Hougaard[3], A. Chakravarti[2], M. Melbye[1,4,5], B. Feenstra[1].* 1) Dept Epidemiology Research, Statens Serum Institut, Copenhagen, Copenhagen, Denmark; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 3) Danish Centre for Neonatal Screening, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, Denmark; 4) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 5) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

   Hirschsprung disease (HSCR) is a congenital disorder with a population incidence of ~1/5000 live births, defined by an absence of enteric ganglia along a variable length of the colon. The condition results from failure in the migration, survival, proliferation or differentiation of ENCCs (enteric neural crest-derived cells) in the intestinal tract. Recent HSCR GWAS have found common associated variants at *RET*, *SEMA3* and *NRG1* but they explain only a small proportion of HSCR heritability. We performed a GWAS with 170 cases identified from the National Patient Registry (ICD-8:75139 or ICD-10:Q43.1) and from the Pathology Registry (SNOMED: S62590) and 4,717 controls from other projects, all genotyped on the Illumina HumanOmniExpressExome array. The analysis was based on 6.2 million variants imputed from the Haplotype Reference Consortium panel (info score ≥ 0.9; MAF > 1%), using logistic regression on imputed SNP dosages under an additive genetic model. In addition to known variants at the *RET* and *SEMA3* loci, we found four new loci with variants at genome-wide significance ($P<5\times10^{-8}$) or close to it ($P<1\times10^{-7}$) in our discovery analysis. Doing *in silico* replication in an independent cohort of 189 probands and 484 controls replicated one of these new loci, with the closest RefSeq gene being *PCDH10*, a neuronal receptor that regulates cell migration and is mostly expressed in fetal brain. We show that, in the mouse, *Pcdh10* is also expressed in the gut during gut development. The top SNP at this locus has a risk allele frequency of 25%, with a combined discovery and replication OR=1.67 ($P=7.6\times10^{-9}$) in comparison with OR=5.1 ($P=6.4\times10^{-77}$) for the known RET locus. Using the GCTA method on all genotyped SNPs, we estimated the heritability of HSCR to be 88%, close to classical family-studies-based estimates. Moreover, by using LASSO regression on all variants with $P<0.01$, we were able to assign case/control labels with 94% accuracy using 191 LASSO SNPs. Furthermore, we detected an enrichment of regulatory regions in neural progenitor cells (MAF-matched 10K permutations & Bonferroni corrected for 127 ENCODE/Roadmap tissues). In conclusion, we detected one novel locus associated with HSCR and were able to predict case/control status with high accuracy based on a subset of less than 200 SNPs.

1309W

EHR data illuminates patient subtypes in obstructive lung diseases yielding new insights for genetic discovery. *N. Banerjee[1], C. Van Hout[1], B. Cajes[1], N. Abul-Husn[1], N. Karra[1], H. Kirchner[2], J. Overton[1], J. Reid[1], I. Borecki[1], F. Dewey[1], A. Shuldiner[1], A. Baras[1], O. Gottesman[1], S. Bruse[1], Geisinger-Regeneron DiscovEHR collaboration.* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Geisinger Health System, Danville, NY.

   Asthma and chronic obstructive pulmonary disease (COPD) are heterogeneous diseases that include several shared and distinct clinical and pathophysiological features. The intrinsic heterogeneity and shared features of these diseases complicates their clinical diagnosis and the downstream classification of patients into case-control cohorts for the purposes of genetic discovery. We hypothesized that data-driven classification methods using multiple dimensions of information - diagnosis codes, laboratory tests and other quantitative measures - could generate more granular and more homogeneous classifications of asthma and COPD patients that would be more suited for genetic discovery than using diagnosis codes alone. We focused on 6494 COPD patients and 7997 asthma patients from the Regeneron Genetics Center-Geisinger Health System DiscovEHR collaboration. All patients have exome-sequence data in additional to rich longitudinal electronic health records. Initial assignment to asthma or COPD was defined by ICD-9 codes. EHR-derived phenotypes used for further classification included all ICD-9 codes and quantitative measures such as blood test results and lung function testing. Through principal component analysis, we observed that even though there was substantial overlap between asthma and COPD patients, there were distinct subgroups. To investigate further, we used information gain analysis to identify feature pairs that are important in each disease and utilized an ensemble learning method to classify between COPD and asthma patients. We identified relevant traits such as BMI, monocytes, leukocytes and neutrophils that contributed significantly toward classifying the two groups with an area under the ROC curve of ≥ 0.85. This distinction is independent of the contribution of known classifiers like age and smoking status. In preliminary exome-wide association analyses, utilizing cluster members as cases and non-cluster members as controls, we observed suggestive evidence of association with multiple variants in the NWD1 gene (most significant variant $p=4.23\times10^{-7}$), with a cluster comprised of predominantly COPD-diagnosed patients (n=494). This gene has been previously implicated by a published GWAS of bronchodilator response in COPD patients. Our findings highlight the potential of leveraging a broader landscape of quantitative and binary phenotypes to refine the classification of obstructive airway diseases for the purposes of genetic discovery.

1310T

Whole exome sequencing in atypical hemolytic uremic syndrome patients suggests involvement of cell adhesion and signaling genes.

F. Bu[1,2], M. Lemaire[3], K. Frees[1], N. Ghiringhelli Borsa[1], C. Nishimura[1], C.M. Nester[1,7], R.P. Lifton[4], V. Fremeaux-Bacchi[5,6], R.J.H. Smith[1,7]. 1) Molecular Oto-laryngology and Renal Research Laboratories, University of Iowa, Iowa City, IA, USA; 2) Medical Genetics Center, Southwest Hospital, Chongqing, China; 3) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; 5) Laboratoire d'Immunologie Biologique, Hôpital Européen Georges Pompidou, AP-HP, Paris, France; 6) INSERM UMRS 1138, Cordeliers Research Center, Complement and Diseases Team, Paris, France; 7) Rare Renal Disease Clinic, Departments of Pediatrics and Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA, USA.

   Introduction: Atypical hemolytic uremic syndrome (aHUS) is an ultra-rare disease characterized by hemolytic anemia, thrombocytopenia, and renal failure. More than half of aHUS patients carry mutations in complement genes that predispose to uncontrolled activation of the alternative pathway of complement at glomerular endothelial surface. The penetrance of mutations in genes known to be associated with aHUS is extremely variable, ranging from 12.5% to 100%. An underlying genetic association is lacking for the other half of patients. We hypothesize that other genes/pathways contribute to aHUS pathogenesis in patients without molecular diagnosis. **Patients & Methods:** We completed whole-exome sequencing (WES) in 106 aHUS patients negative for complement gene mutations, DGKE and Factor H-autoantibodies. We excluded 13 patients because they failed the relatedness or population stratification tests. In the remaining 93 patients, GATK Variant Quality Score Recalibration (VQSR) was used to filter variants with a 99.5% sensitivity threshold for SNVs and 99.0% for indels. The top 0.5% of exonically variable genes (n=94) and the top 0.2% of variable exons were excluded. Remaining variants were annotated and filtered on the basis of missing genotype rate, minor allele frequency and predicted functional consequence. Gene burden was tested using Sequence Kernel Association Test (SKAT) implemented in Ingenuity Variant Analysis (IVA). Genes with P<0.05 were included in pathway analyses using WebGestalt. The control group is comprised of WES data generated locally from 300 adults without kidney disease. **Results:** In total, WES data from 93 aHUS patients and 300 controls yielded 18,322 ultra-rare/novel variants that passed all filtering steps. Pathway analysis was performed with the 120 genes that had p-values < 0.05 in a gene burden analysis. Of these, 13 genes (P<0.01) are bona fide candidates for functional studies. Significant gene enrichment was observed in cell adhesion and signaling pathways, including E-cadherin and N-cadherin signaling events (Hypergeometric, FDR adjusted p-values of 0.0032 and 0.0076, respectively). **Summary:** These preliminary findings show enrichment for ultra-rare/novel variants in cell adhesion and signaling pathways in this aHUS cohort. These data suggest that a novel pathophysiologic path to aHUS is driven by mutated adhesion proteins expressed at the surface of endothelial cells.

1311F

Rare variants in 5α-reductase-2 and GWAS susceptibility loci are associated with elevated androgen levels in polycystic ovary syndrome.

M. Dapas[1], R. Sisk[1], R.S. Legro[2], M. Urbanek[1], A. Dunaif[1], M.G. Hayes[1,3,4]. 1) Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611; 2) Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, PA 17033; 3) Department of Anthropology, Northwestern University, Evanston, IL 60208; 4) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611.

   Polycystic ovary syndrome (PCOS) is a common endocrine disorder affecting ~7% of premenopausal women worldwide. Analogous to other complex diseases, common genetic susceptibility variants for PCOS account for little of the observed heritability. We tested the hypothesis that this deficit in heritability was due to rare genetic variants with larger biological effects. Whole genome sequencing was performed on 64 two-generation families with one or more daughters with PCOS. Variants were filtered according to allele frequency (MAF≤2%), call quality, consistency with Mendelian inheritance, and predicted deleteriousness. Associations between sets of rare variants and PCOS and its quantitative hormonal traits were assessed with burden tests, grouping variants at the gene-level (including 3' UTR and 7.5kb upstream of 5' TSS) and in sliding windows across the genome, accounting for relatedness, and adjusting for age and BMI. After correcting for multiple testing (P_{adj}), significant associations with testosterone (T) levels were found for rare variants in *SRD5A2* ($P=7.16 \times 10^{-7}$, $P_{adj}=0.012$), *CAMK4* ($P=5.98 \times 10^{-8}$, $P_{adj}=0.0074$), and an intergenic region in 9p23 ($P=2.29 \times 10^{-7}$, $P_{adj}=0.035$). Significant associations were also found with dehydroepiandrosterone (DHEAS) levels for rare variants in *PTCH1* ($P=1.11 \times 10^{-7}$, $P_{adj}=0.017$) and *CAMKMT* ($P=1.51 \times 10^{-7}$, $P_{adj}=0.016$). These genes are both in linkage disequilibrium (LD) with the previously identified PCOS risk alleles in or near *C9orf3* and *THADA*, respectively. No significant associations were found with PCOS diagnosis or with fasting insulin, luteinizing hormone, or sex hormone binding globulin levels. *SRD5A2* encodes 5α-Reductase-2, which modulates target tissue androgen action by converting T to its more potent metabolite, dihydrotestosterone. Increases in 5α-Reductase activity have been found in PCOS, and functional *SRD5A2* SNPs have been associated with PCOS. Further, we recently reported increases in global 5α-Reductase activity in infant daughters of affected women, suggesting that these changes play a primary role in disease pathogenesis. The rare variants associated with DHEAS levels in *PTCH1* and *CAMKMT*, which encode Patched-1 and Calmodulin-Lysine N-Methyltransferase, respectively, are in LD with two nearby PCOS GWAS susceptibility loci, suggesting that they are causal variants contributing to signals at these loci. Collectively, our findings suggest that rare variants contribute to elevated androgen levels in PCOS.

1312W

Effects of *NRG1* variants in Hirschsprung patients. *D.R. Gunadi[1], M. Sunardi[2], N.Y.P. Budi[2], A.S. Kalim[2], K. Iskandar[3], A. Makhmudi[1].* 1) Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Universitas Gadjah Mada/Dr. Sardjito Hospital, Yogyakarta, Indonesia; 2) Molecular Biology Laboratory, Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia; 3) Department of Child Health, Faculty of Medicine, Universitas Gadjah Mada/UGM Hospital, Yogyakarta, Indonesia.

   Purpose: Hirschsprung disease (HSCR) is a neurocristopathy characterized by absence of intramural ganglion cells along variable lengths of the intestines in neonates. The *NRG1* gene, encoding neuregulin 1, is implicated in the development of the enteric nervous system (ENS), and seems to contribute by both common and rare variants. In this study, we analyzed the *NRG1* common and rare variants in Indonesian HSCR patients. **Methods:** We diagnosed a HSCR in our hospital based on the clinical manifestation, contrast enema, and/or histopathology findings. The Hematoxillin & Eosin staining and/or S100 immunohistochemistry were utilized for the histopathology diagnosis of HSCR. Sixty isolated HSCR patients and 124 controls were ascertained for *NRG1* common variant study. The two genetic markers, rs7835688 and rs16879552, within intron 1 of *NRG1* were examined using TaqMan Genotyping Assays in genomic DNA for association studies. To identify a rare variant of *NRG1* in forty HSCR patients, Sanger sequencing method was performed using a BigDye Terminator Cycle Sequencing Kit. **Results:** *NRG1* rs7835688 was associated with HSCR by case–control analysis (OR = 2.0, 95% CI = 1.3-3.3; p = 4.3 × 10^{-3}), whereas rs16879552 demonstrated no association (OR = 1.6, 95% CI = 0.9-2.7; p = 0.097). Further, the transmission disequilibrium tests (TDT) showed that neither of the two *NRG1* common variants was significant, with the transmission rates were at ~62% for both variants. We identified a missense mutation, c.397G>C, in exon 7, which led to a substitution of valine with leucine in the neuregulin 1 (p.V133L). Further, the analysis of *NRG1* coding sequences in this study was not finished yet due to time constraint. **Conclusions:** *NRG1* rs7835688 variant is a common susceptibility factor for HSCR in Indonesia. The *NRG1* rare variant might also have a significant impact in Indonesian HSCR.

1313T

Impact of *Semaphorin 3* variants in patients with Hirschsprung disease. *K. Iskandar[1], N.Y.P. Budi[2], A.S. Kalim[2], M. Sunardi[2], A. Makhmudi[3], . Gunadi[3].* 1) Department of Child Health, Faculty of Medicine, Universitas Gadjah Mada/UGM Hospital, Yogyakarta, Indonesia; 2) Molecular Biology Laboratory, Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia; 3) Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Universitas Gadjah Mada/Dr. Sardjito Hospital, Yogyakarta, Indonesia.

   Purpose: Recently, common and rare variants within a locus on 7q21.11 containing the *Semaphorin* (*SEMA*) *3A, SEMA3C,* and *SEMA3D* genes were reported to be associated with Hirschsprung disease (HSCR). Here, we investigated three genetic markers, rs1583147, rs12707682, and rs11766001, at this locus and *SEMA3D* rare variants to determine their potential contributions to the susceptibility of Indonesian HSCR patients. **Methods:** Sixty isolated HSCR patients and 118 ethnicity-matched controls were ascertained for *SEMA3* common variant study. The three genetic markers were examined using TaqMan Genotyping Assays in genomic DNA for association studies. To identify a rare variant of *SEMA3D* in forty HSCR patients, Sanger sequencing method was performed using a BigDye Terminator Cycle Sequencing Kit. **Results:** The risk allele frequencies of *SEMA3* rs12707682 (allele C) and rs1583147 (allele T) is higher in cases, 53% and 23%, than in controls, at 42% and 13%, respectively. However, these frequency differences were not statistically significant with p-value of 0.06 and 0.023, respectively. These findings were consistent with transmission disequilibrium test (TDT) results with p-values of 0.041 and 0.11 for rs12707682 and rs1583147, respectively. Furthermore, the frequency of *SEMA3* rs11766001 risk allele in HSCR cases and controls were 1.7% and 0.8%, respectively. We were unable to find any rare variant in the first five exons of the *SEMA3D* gene in Indonesian HSCR patients. **Conclusions:** *SEMA3* rs12707682 and rs1583147 variants are not common risk factors for HSCR in Indonesia. The rarity of the *SEMA3* rs11766001 polymorphism in Indonesian population might be due to a founder effect. Furthermore, the effects of *SEMA3D* rare variant in Indonesian HSCR could not be determined since the sequencing analysis was not completed yet due to time limit. .

1314F

Copy number gains encompassing genes *SLC2A14* and *OR4K1* are significantly associated with spirometric measures of Chronic Obstructive Pulmonary Disease (COPD). *F. Begum[1], L. Jiang[1], I. Ruczinski[2], M. Cho[3], M. Parker[3], J. Crapo[4], E. Silverman[3], T. Beaty[1], COPDGene Investigators.* 1) Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Maryland, USA; 3) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 4) Department of Medicine, National Jewish Health, Denver, USA.

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death in the United States. Besides cigarette smoking, a major environmental factor, it is crucial to understand the genetic underpinning of COPD, as multiple genetic factors have been identified from large-scale genetic studies of COPD. Spirometry is the defining pulmonary function test (PFT) for diagnosing COPD, especially FEV1 (forced expiratory volume in 1s) and the ratio of FEV1 to FVC (forced vital capacity). Previous large genome-wide association studies (GWAS) have identified several genes and pathways associated with quantitative measures of pulmonary function. A genome-wide sequencing study by Govindan et al. (2012) validated structural variants, which may also contribute to lung adenocarcinoma. Begum et al. (2016) reported a hemizygous deletion at 5q35.2 among African Americans (AA) associated with reduced total lung capacity using COPDGene study. In this study we extended our search for association between spirometric measures with both amplification and deletions. We used PennCNV to estimate CNVs on 9,076 cigarette smokers with and without COPD from the COPDGene study exploiting Illumina's Omni-Express genome-wide marker data. Among these 9,076 subjects, 2,889 subjects are AAs and the remainder are Non-Hispanic Whites. In a race-stratified analysis, we tested for association between estimated CNV components (defined as disjoint intervals of copy number regions) and multiple spirometric measures. We observed a large region of polymorphic CNVs (mostly amplifications) spanning the ~71kb region encompassing the *SLC2A14* gene on chromosome 12, and ~19kb region of chr14q11.2 involving the *OR4K1* gene achieved genome-wide significance (p-value=9.38E-06 & p-value=2.33E-05) in tests of association with multiple spirometric measures (FEV1 percent predicted, FVC, pre-bronchodilator FEV1 and FVC, and FEV6) among AAs. The *SLC2A14* gene is a glucose transporter gene involved with cell proliferation and is included in multiple cancer pathways. Moreover, it has been previously reported in the literature that copy number gain in the chr12p13.31 region showed overexpression of 13 novel asbestos-related miRNAs among lung cancer patients. *OR4K1* is an olfactory receptor. We are currently following up on these association signals using data from other studies including the Atherosclerosis Risk in Communities (ARIC) study and the lung carcinoma data from The Cancer Genome Atlas (TCGA). .

1315W

Disease associations of common and rare calcium sensing receptor variants in the 50K DiscovEHR cohort. *G. Breitwieser, R. Metpally, S. Krishnamurthy, D. Smelser, J. Robishaw, D. Carey, Geisinger-Regeneron DiscovEHR collaboration.* Weis Center for Research, Geisinger Clinic, Danville, PA.

Calcium sensing receptor (*CASR*) variants contribute to serum calcium regulation. Unfortunately, our current understanding of the impact of *CASR* genetic variants is restricted to narrowly focused studies on patients having severe calcium perturbations. The overall impact of more modest alterations in serum calcium on disease prevalence is unknown. Our goals in this study were to determine the prevalence of *CASR* common and rare variants in an unselected population, and assess the impact on disease. <u>Methods</u>. We analyzed whole exome sequences from 51,289 individuals in the DiscovEHR cohort who consented to participate in the MyCode Community Health Initiative. Common *CASR* tri-locus haplotypes were tested for association with serum calcium levels. *CASR* haplotype association with disease phenotypes was determined by Phenome-wide Association Study (PheWAS), comparing the haplotype with highest serum calcium with the reference haplotype. Rare coding variants in *CASR* were functionally classified as benign, loss or gain of function using bioinformatics plus serum calcium measures from the electronic health record. Optimized sequence kernel association tests (SKAT-o) were used to detect rare variant disease associations. <u>Results</u>. Of the common tri-locus haplotypes, S986/R990/E1011 was associated with the highest serum calcium. The SRE/SRE haplotype (present in 1.4% of cohort) predisposed to cardio-renal diseases. A total of 2116 individuals (1/24) carried rare (MAF<1%) *CASR* variants. Fourteen rare *CASR* variants (1/2100 individuals) significantly altered serum calcium in the heterozygous state (2 gain of function, 4 strong loss of function and 8 mild loss of function variants). Rare variant tests showed significant association of loss of function variants with cardiovascular diseases, sleep disturbances, epilepsy, and skin and breast cancer phenotypes. <u>Conclusions</u>. Rare and common *CASR* variants influence total serum calcium levels and represent a significant disease risk factor. Supported by funds from GHS and REGN.

1316T

GWAS of thyroid function markers identifies novel loci and highlights connections to metabolic phenotypes and thyroid disease. *A. Teumer[1,2], L. Chaker[3,4], M. Medici[3], A. Köttgen[5], H. Völzke[1,2], R. Peeters[3] on behalf of the CHARGE thyroid function working group.* 1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) DZHK - German Center for Cardiovascular Research, partner site Greifswald, Greifswald, Germany; 3) Department of Internal Medicine and Rotterdam Thyroid Center, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; 4) Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; 5) Division of Genetic Epidemiology, Medical Center - University of Freiburg, Germany.

Overt and subclinical thyroid dysfunction are associated with several diseases. Circulating concentrations of the thyroid-stimulating hormone (TSH) and free thyroxin (FT4) have a strong heritable component, but previously identified loci only explain 19% and 10% of their estimated genetic variance, respectively. To find novel loci associated with thyroid function, we performed sex-stratified 1000Genomes imputation-based genome-wide association studies (GWAS) in euthyroid subjects of European-ancestry from 23 population-based cohorts (n=54,288 for TSH and n=49,269 for FT4). Additionally, we conducted GWAS to reveal susceptibility loci for hypo- and hyperthyroidism defined by increased (n=3,340 cases) and decreased (n=1,840 cases) TSH levels, respectively. After the discovery stage, we revealed 24 and 22 novel genome-wide significant loci for TSH and FT4, respectively. The explained genetic variance increased to 33% for TSH and 22% for FT4. All 30 previously associated loci with thyroid hormones were replicated. Among the novel loci associated with TSH were the genes encoding for the TSH receptor and thyroglobulin, and the thyroid hormone degrading enzyme *DIO2* associated with FT4. Although all loci associated with hypo- or hyperthyroidism were also associated with TSH, only some overlapped with both hypo- and hyperthyroidism (*PDE8B, PDE10A, FOXE1, CAPZB*). No significant gene-by-sex interactions were found. Genetic risk scores (GRS) built from the significant loci for TSH and FT4 were associated with goiter and body height. The GRS for TSH was also associated with TPO-antibody positivity, Graves' disease, hypo- and hyperthyroidism, type 2 diabetes, kidney function and schizophrenia, whereas the GRS for FT4 was additionally associated with body mass index. The most significant pathways were related to regulation of phosphate metabolism and body length for the TSH, and to the amount of body fat for the FT4 associated loci. Replication of the new associations among ~17,000 independent samples will be completed soon. The results of this GWAS highlight the close connection of thyroid function to metabolic phenotypes and suggest a relation of thyroid function and schizophrenia. They provide a basis for functional follow-up studies and assessment of the causality of associations between thyroid function and other diseases using Mendelian randomization. Together, this will facilitate an improved understanding of the systemic effects of thyroid hormones.

1317F

SERPINA1 **Z variant heterozygosity increases risk for lung and liver disease in a large clinical care cohort.** *S. Bruse[1], N. Abul-Husn[1], J. Leader[2], H. Kirchner[2], C. Van Hout[1], J. Overton[1], J. Reid[1], A. Baras[1], M. Murray[2], A. Shuldiner[1], F. Dewey[1], O. Gottesman[1], Geisinger-Regeneron DiscovEHR collaboration.* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

The *SERPINA1* ZZ genotype is the most common cause of alpha-1-antitrypsin (AAT) deficiency, with associated increased risk of chronic obstructive pulmonary disease (COPD) and liver disease. While heterozygosity for the *SERPINA1* Z variant (principally the MZ genotype) is suspected to confer disease risk, this has not been definitively established. We leveraged the rich source of electronic health record (EHR) and exome sequencing data from the Regeneron Genetics Center–Geisinger Health System DiscovEHR collaboration to determine the association of MZ genotype with lung and liver disease in a clinical care cohort. In 49,176 sequenced adults of European ancestry, we examined the association of MZ genotype with EHR-extracted measures of AAT (n=1,360), alanine aminotransferase (ALT; n=40,505), aspartate aminotransferase (AST; n=39,904), alkaline phosphatase (ALP; n=39,533), and spirometry (n=9,126). MZ genotype was also tested for association with alcoholic (n=180) and non-alcoholic (n=3,316) liver disease, asthma (n=7,129), COPD (n=5,858), and COPD-specific diagnoses of emphysema (n=1,422) and chronic bronchitis (n=2,267), as defined by ICD9 diagnosis codes. The common S variant (MS, SS, and SZ genotype carriers; n=3,482) was excluded from these analyses. We identified 1618 MZ genotype carriers in the cohort (3.3%). MZ genotype was associated with a 33% reduction in AAT ($p=1.2 \times 10^{-80}$), and with increased ALT (9.8%; $p=2.0 \times 10^{-19}$), AST (6.3%; $p=2.1 \times 10^{-19}$), and ALP (7.3%; $p=1.1 \times 10^{-21}$) levels. There was no association with spirometry. In case/control analyses, MZ genotype was associated with alcoholic and non-alcoholic liver disease (odds ratio [OR] 2.6, p=0.0006; OR 1.26, p=0.03, respectively), COPD (OR 1.27, p=0.009), and emphysema (OR 1.42, p=0.02). When restricting analyses to COPD (n=1,858) and emphysema (n=668) patients with spirometry-confirmed airway obstruction, MZ remained significantly associated with disease (OR 1.47, p=0.004; OR=1.77, p=0.003 for COPD and emphysema respectively). There was no association with asthma or chronic bronchitis. Thus, in a large clinical care cohort, *SERPINA1* MZ genotype was significantly associated with increased liver enzyme levels, and with increased risk of COPD, emphysema, and liver disease. To our knowledge, this is the first study to clearly demonstrate the association of Z heterozygosity in *SERPINA1* with clinical disease risk, which has important implications given the high population frequency of the Z variant.

1318W

Whole exome sequencing and linkage analysis identified *TTK* as a candidate gene for pulmonary nontuberculous mycobacterial lung disease. *F. Chen[1], E.P. Szymanski[2], K.N. Olivier[3], X. Liu[4], H. Tettelin[4], S.M. Holland[2], P. Duggal[1].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Laboratory of Clinical Infectious Diseases, NIAID, NIH, Bethesda, MD; 3) Cardiovascular and Pulmonary Branch, NHLBI, NIH, Bethesda, MD; 4) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD.

   Background: Pulmonary nontuberculous mycobacterial (PNTM) lung disease is caused by nontuberculous mycobacteria (NTM) ubiquitous in the environment, but only a fraction of immunocompetent individuals develop the disease. A distinct group of individuals affected by PNTM are white postmenopausal women, with a tall and lean body habitus and higher rates of scoliosis, pectus excavatum, mitral valve prolapse, and mutations in the *CFTR* gene. These clinical features and the familial clustering of the disease suggest an underlying genetic mechanism. **Methods and Results:** To map the genes associated with PNTM, whole-exome sequencing (WES) was conducted in 10 PNTM families and 57 sporadic cases recruited at the NIH Clinical Center during 2001-2013. Eighty-five percent of PNTM cases in this cohort were women of European ancestry. We carried out a variant-level and a gene-level parametric linkage analysis on nine PNTM families (16 affected and 20 unaffected) under a recessive disease model. We performed multipoint linkage analysis in MERLIN with 4,328 independent (pairwise $r^2 < 0.10$) genome-wide common variants, and identified a 20 cM region on chromosome 6q12-6q16 (HLOD=3.9) with 100% penetrance and a risk allele frequency of 5%. To determine the gene(s) underlying the linkage signal on chromosome 6, all sequenced variants in each gene region were collapsed to construct haplotypes using the collapsed haplotype pattern (CHP) method implemented in SEQLinkage, and genes were then tested as regional markers in a single-point linkage analysis in MERLIN. In this gene-level linkage analysis, the TTK protein kinase (*TTK* [MIM 604092]) gene on chromosome 6q14.1 was found to be most significant (HLOD=3.38). Evidence of suggestive linkage was also found in the same linkage region on chromosome 6 for gene trophoblast glycoprotein (*TPBG* [MIM 190920], HLOD=3.08), origin recognition complex subunit 3 (*ORC3* [MIM 604972], HLOD = 3.10), and ankyrin repeat domain 6 (*ANKRD6* [MIM 610583], HLOD = 3.02). **Conclusion:** *TTK* encodes a protein kinase that is essential for mitotic checkpoints and the DNA damage response. Dysfunction in DNA repair is a common finding in patients with chronic obstructive pulmonary disease (COPD) and tuberculosis (TB). Inhibition of *TTK* expression has been associated with compromised DNA repair and cell survival, which may contribute to the progression from NTM infection to the occurrence of pulmonary disease.

1319T

Identifying modifying genes to explain the variation in severity of fragile X-associated primary ovarian insufficiency. *C.E. Trevino, E.G. Allen, P. Jin, H.R. Johnston, D.J. Cutler, M.E. Zwick, S.T. Warren, S.L. Sherman.* Human Genetics, Emory University, Atlanta, GA.

   Primary ovarian insufficiency (POI) is a condition characterized by symptoms of early menopause and infertility that affects 1% of the general population, making it a leading cause of infertility in women. Patients with this condition also display elevated follicle stimulating hormone, low anti-Müllerian hormone levels and early estrogen deficiency. Women who carry the *FMR1* premutation, a CGG repeat expansion in the range of 55-200 repeats in the 5'UTR of the X-linked *FMR1* gene, are at a significantly increased risk for POI: about 20% of carriers are diagnosed with fragile X-associated POI (FXPOI). Although the risk for FXPOI is significant, not all carriers experience the disorder, and among those who do, the range of severity is broad. We hypothesize that there are other genetic variants, in addition to premutation repeat number, that modify the age of onset and severity of FXPOI. To test this hypothesis, we are currently conducting a whole genome sequencing (WGS) study among women with the premutation, comparing those who had FXPOI diagnosed before age 35 with those who had menopause after age 50. Currently, we have sequence data on 50 in each group and variants from the WGS are being prioritized using a well-established pipeline. From these data, we will take the top candidate genes and screen them using *Drosophila* models as a high-throughput, whole organism functional screen. Models will include the KO of the candidate gene and the KO on a premutation background (90 CGG repeats). The premutation *Drosophila* model has been shown to have reduced fecundity compared with the control flies. Using both models will allow us to determine whether there are modifying effects of the candidate gene specifically with the premutation. We will present results from the WGS and from the *Drosophila* models based on the top 5 candidate genes. The overall goal of this project is to elucidate novel molecular mechanisms for FXPOI to facilitate early risk assessment and diagnosis, allowing more informed choices about family planning and fertility options. In addition, the identified genes may also play a role in idiopathic POI and lead to a new understanding of possible treatment targets.

1320F

Allelic imbalance in regulation of *ANRIL* through chromatin interaction at 9p21 endometriosis risk locus. *H. Nakaoka, A. Gurumurthy, T. Hayano, S. Ahmadloo, W. Omer, C. Hata, T. Kimura, K. Hosomichi, I. Inoue.* Division of Human Genetics, National Institute of Genetics, Mishima, Shizuoka, Japan.

A large number of variants associated with human complex diseases have been discovered by genome-wide association studies (GWASs). These discoveries have been anticipated to be translated into the definitive understanding of disease pathogeneses; however, functional characterization of the disease-associated SNPs remains a formidable challenge. Here we explored regulatory mechanism of a SNP on chromosome 9p21 associated with endometriosis by leveraging "allele-specific" functional genomic approaches. By scrutinizing linkage disequilibrium structure and DNase I hypersensitive sites across the risk locus, we prioritized rs17761446 as a candidate causal variant. Chromosome conformation capture followed by high-throughput sequencing revealed that the protective G allele of rs17761446 exerted stronger chromatin interaction with *ANRIL* promoter. We demonstrated that the protective allele exhibited preferential binding affinities to TCF7L2 and EP300 by bioinformatics and chromatin immunoprecipitation (ChIP) analyses. ChIP assays for histone H3 lysine 27 acetylation and RNA polymerase II reinforced the enhancer activity of the SNP site. The allele specific expression analysis for eutopic endometrial tissues and endometrial carcinoma cell lines showed that rs17761446 was a *cis*-regulatory variant where G allele was associated with increased *ANRIL* expression. Our work illuminates the allelic imbalances in a series of transcriptional regulation from factor binding to gene expression mediated by chromatin interaction underlie the molecular mechanism of 9p21 endometriosis risk locus. Functional genomics on common disease will unlock functional aspect of genotype-phenotype correlations in the post-GWAS stage.

1321W

Genome-wide association study identified new susceptible genetic variants in HLA class I region for hepatitis B virus-related hepatocellular carcinoma. *H. Sawai[1], N. Nishida[1,2], M. Sugiyama[2], S-S. Khor[1], M. Mizokami[2], K. Tokunaga[1].* 1) Department of Human Genetics, The University of Tokyo, Tokyo, Japan; 2) Genome Medical Sciences Project, National Center for Global Health and Medicine, Chiba, Japan.

Recent genome-wide association studies (GWAS) of chronic hepatitis B virus (HBV) carriers with or without hepatocellular carcinoma (HCC) in China reported that SNPs in *KIF1B*, *HLA-DQA1/DRB1*, *GRIK1*, *STAT4* and *HLA-DQ* were associated with HBV-related HCC susceptibility. In this study, we performed a GWAS using 473 Japanese HBV-positive HCC patients and 516 HBV carriers including chronic hepatitis (CH) and asymptomatic carrier (ASC) individuals to identify new host genetic factors associated with HBV-derived HCC in Japanese and other East Asian populations. There were 110 SNPs with P values < 10^{-4} with 65 and four SNPs located within the HLA class I and II regions, respectively. In a subsequent replication analysis, three SNPs in the HLA class I region were genotyped in four independent population-based replication sets (Japanese, Korean, Hong Kong Chinese, and Thai). The meta-analysis with the combined Japanese samples and three other replication sample sets confirmed the association for the SNPs (SNP1: OR = 1.63, P = 6.17×10^{-11}; SNP2: OR = 1.70, P = 4.79×10^{-13}; and SNP3: OR = 1.70, P = 4.61×10^{-9}). We performed two-field HLA genotype imputation for five HLA loci (*HLA-A, B, DRB1, DQB1* and *DPB1*) using GWAS SNP typing data to investigate the association of HLA alleles. A subsequent HLA allele association test revealed that a particular HLA class I allele was significantly associated with disease progression to HCC (OR = 1.97, P = 4.58×10^{-4}). Conditioning analysis of each of the three SNPs on the HLA class I region abolished the association of the HLA allele (P > 0.05). However, conditioning the HLA allele could not eliminate the association of the three SNPs (SNP1: OR = 1.67, P = 6.53×10^{-4}; SNP2: OR = 1.72, P = 2.49×10^{-4}; and SNP3: OR = 1.48, P = 1.02×10^{-2}), suggesting that additional genetic factors may exist in the HLA class I region.

1322T

Developmental genes modify the pathological picture of Biliary Atresia but are not the predominant cause of the disease. *G. Cheng[1], M.M. Garcia-Barceo[1], P.H.Y. Chung[1], E.W.K. Tang[1], J.M.T. So[1], E.K.W. Chan[2], D.N. Ngo[5], N.S. Tran[4], P.A.H. Nguyen[3], S. Cherney[6], P.S. Sham[6], J.M. Nicholls[7], P.K.H. Tam[1].* 1) Surgery, University of Hong Kong, Hong Kong, Hong Kong, Hong Kong; 2) Surgery, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Hong Kong; 3) Hepatology, the National Hospital of Pediatrics, Hanoi, Vietnam; 4) Surgery, the National Hospital of Pediatrics, Hanoi, Vietnam; 5) Genetics, the National Hospital of Pediatrics, Hanoi, Vietnam; 6) Psychiatry, University of Hong Kong, Hong Kong, Hong Kong, Hong Kong; 7) Pathology, University of Hong Kong, Hong Kong, Hong Kong, Hong Kong.

Biliary atresia (BA) is rare but is the most common indication for pediatric liver transplantation (LT) worldwide. The symptoms appear shortly after birth, a few patients have extrahepatic developmental anomalies, and some have ductal plate malformation (DPMs), altogether suggesting that developmental deregulation underlies BA. Yet the developmental theory of BA is still controversial especially with possible disease heterogeneity. We aim to test whether the developmental deregulation causes non-syndromic BA. Liver DNAs of 22 BA patients were screened for mutations in developmental genes by whole exome sequencing and the liver histopathology was analysed. Immunostaining was performed on 4 biopsies taken at the initial Kasai operation (KO) and18 at LT. Biliary proliferation that is typical of BA was reported at KO for the 18 patients, yet at LT, ductopenia occurred in 9 patients, 4 carrying ≥1 damaging mutations in *JAG1-NOTCH2* associated with Alagille disease. The remaining 9 patients showed signs of secondary biliary cirrhosis with profound biliary proliferation at LT with 4 carrying mutations in cystic biliary disease causal genes but no mutations in *JAG1-NOTCH2*. Nonetheless, no DPM was detected in the 22 livers either at Kasai or LT, leaving the effect of the developmental mutations on the pathogenic stage of BA elusive, despite the link with the BA pathological progress after KO. Further, the mutation burden of developmental genes indeed is heavy in the healthy population (N=616). The burden test showed that the IL4R pathway but not developmental gene/pathways are associated with BA (FDR *q*<0.05). We then explored if there was any connection between the developmental genes and the BA-associated genes without disease or biological hypothesis, for which we retrieved 103 genes through a genome-wide gene-based SNP association test (181 cases vs 431 controls). We examined the connectivity between the 103 candidates tagged by SNPs and the developmental genes (N=8, along with 184 genes with CMC *p*<0.05 in Exome screening). The SNP-tagged genes and the mutant genes were connected (*p*=0.031) in the interactome, while both are enriched with proinflammatory pathways. Moreover, their network fits into a core/periphery structure where immunity modules dominate the core and developmental genes fall at the periphery. This suggests that developmental genes may modify the BA manifestation but the occurrence of the disease is predominantly driven by immunity deregulation.

1323F

Characterization of *ELMO1* as a target for kidney disease using genetic and gene expression data. *D.M. Waterworth[1], L. Li[2], J. Yao[3], E. Hu[4], M. O'Donoghue[5], H. White[6], L. Wallentin[7], D. Rajpal[3].* 1) Genetics, GlaxoSmithKline, King of Prussia, PA; 2) Genomic Medicine Group, PAREXEL, Durham, NC; 3) Computational Biology, GlaxoSmithKline, King of Prussia, PA; 4) Heart Failure Discovery Performance unit, GlaxoSmithKline, King of Prussia, PA; 5) TIMI study group, Cardiovascular Division, Brigham and Women's Hospital, Boston, MA; 6) Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand; 7) Department of Medical Sciences, Cardiology & Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden.

ELMO1 (engulfment and cell motility 1) encodes a member of the engulfment and cell motility protein family. These proteins interact with dedicator of cytokinesis proteins to promote phagocytosis and cell migration. The *ELMO1* gene locus has been associated with diabetic nephropathy in the GoKinD analysis but was not robustly replicated in the Genetics of Nephropathy (GENIE) consortium. Mouse studies have demonstrated that high Elmo1 expression aggravates and low Elmo1 expression prevents diabetic nephropathy. It has been hypothesized that *ELMO1* may have a role in the advanced stages of diabetic nephropathy, perhaps contributing to renal function decline, rather than its initiation. In order to see if we could obtain further evidence on the relationship between *ELMO1* and renal function as well as failure, we investigated CKD status and incident serious renal failure events (n=293) in two large cardiovascular outcome trials (STABILITY and SOLID-TIMI 52, total n=21310). These trials had a high proportion of diabetics (37.8 and 34.6% respectively) and CKD II-IV (73 and 61% respectively). No marked associations with baseline eGFR or CKD criteria were observed after bonferroni correction for the number of variants within the gene, but an association with incident renal events was observed with a low frequency variant rs73689696 (MAF 0.02) OR 2.9 (1.9-4.4), p=7.4e-7. This association was directionally consistent across both trials and each arm of the trial (IP vs placebo) and remained significant after bonferroni correction adjusting for variants analyzed in the gene (p=7.22e-5). Similar to GoKinD, the associated region is at the 3' end of the *ELMO1* gene, but rs73689696 is not in high LD with the reported variants. The renal events in these trials were predominantly acute renal events. Additionally, investigation of multiple human renal gene expression studies suggested that *ELMO1* gene expression is dys-regulated in the glomeruli of hypertensive nephropathy and IgA nephropathy as well as in the tubuli of minimal change disease, which causes nephrotic syndrome. The combined evidence suggests that ELMO1 is a promising target for a range of kidney diseases, but in particular for more advanced diabetic nephropathies and acute renal events in coronary heart disease patients.

1324W

Rare anti-Müllerian hormone (*AMH*) variants with functional impact *in vitro* **identified in women with polycystic ovary syndrome.** *L.K. Gorsic[1], G. Kosova[1], B. Werstein[1], R. Sisk[1], R.S. Legro[4], M.G. Hayes[1,2,3], J.M. Teixeira[5], A. Dunaif[1], M. Urbanek[1,2].* 1) Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Department of Anthropology, Northwestern University, Evanston, IL; 4) Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, PA; 5) Department of Obstetrics, Gynecology and Reproductive Biology, College of Human Medicine, Michigan State University, Grand Rapids, MI.

Polycystic ovary syndrome (PCOS) is a common endocrine condition characterized by disordered gonadotropin secretion, increased androgen production and polycystic ovarian morphology (PCO). It is the leading cause of anovulatory infertility. As with other complex traits, common disease-susceptibility variants account for a small percentage of PCOS heritability. We performed whole genome sequencing in 57 PCOS cases and 12 reproductively normal control women to test the hypothesis that rare variants contribute to this deficit in heritability. Implementing a candidate gene approach, we detected likely to be deleterious variants in *AMH*, which encodes anti-Müllerian hormone (AMH), a key regulator of follicular recruitment and development and is secreted by granulosa cells of preantral and small antral follicles. Since circulating AMH levels correlate with antral follicle counts, AMH has been proposed to be a marker for follicle number. PCO is characterized by an increase in ovarian follicles and consistent with this, circulating AMH levels are typically elevated in PCOS. However, AMH also inhibits androgen production in the theca cells of the ovary. Given its biological relevance to PCOS, we performed targeted re-sequencing of the *AMH* gene in an additional 644 PCOS women and 153 reproductively normal women of comparable age. We identified 24 rare (MAF<0.01) variants within *AMH* dispersed across all five exons. All women were heterozygous for the variants. Four variants occurred in PCOS and control women, two variants occurred solely in control women and 18 variants were specific to women with PCOS. Fifteen of the 18 PCOS-specific variants were singletons (found in one subject), while three variants occurred in multiple affected women (V12G in 6; T99S in 2; P352S in 3). One woman with PCOS had two missense variants, R194H and A385V, mapping to the same copy of the gene. Dual luciferase reporter assays assessed the impact of AMH variants on AMH signaling. Seventeen of the 18 PCOS-specific variants showed significantly reduced AMH signaling. No variants observed in control women had impaired signaling activity. In this comprehensive genetic screen of the *AMH* gene, we demonstrate that a subset of women with PCOS have functional *AMH* variants. This study is the first to identify and functionally validate rare genetic variants associated with a common PCOS phenotype and suggests a previously unrecognized mechanism for AMH's role in PCOS: decreased AMH bioactivity.

1325T

Treatment of hepatitis C with interferon may lead to reduced risk of liver fibrosis via downregulation of *MERTK* by allele-specific binding of IRF1. *C. Wadelius, G. Pan, H. Nord, E. Wallen Arzt, M. Cavalli.* Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.

Background: Hepatitis C virus infection can result in the development of liver fibrosis and may eventually progress into cirrhosis and hepatocellular carcinoma. Genome-wide association studies of reduced risk of fibrosis have shown association to rs4374383 and 72 SNPs in high LD but the functional variant has not been detected. **Methods and results:** We utilized a bioinformatics approach to identify a non-coding regulatory variant, rs6726639, with genome-wide significant differential transcription factor binding between the alleles which could likely be the driver of the association, as it is in high LD with rs4374383. ChIP-seq showed significantly more reads to the A than the C allele consistent with a stronger band for the A allele in electrophoresis mobility shift assay (EMSA). In luciferase assays the A allele showed significantly lower signal indicating that a repressor may bind. The A allele of rs4374383 is on the same haplotype as the A allele of rs6726639 and is associated to reduced risk of liver fibrosis. Interferon alpha (INFA) is used to treat hepatitis C to prevent fibrosis and it up-regulates interferon regulatory factor 1 (*IRF1*). Motif analysis showed that binding sites for the IRF family overlaps rs6726639 and that this variant could disrupt the binding of IRF1. EMSA supershift experiments with an antibody against IRF1 showed decreased binding at this site. **Conclusions:** Treatment of hepatitis C with INFA results in increased IRF1 levels and our data suggest that this leads to an allele-specific down-regulation of *MERTK* mediated by allelic down-regulation of the regulatory element containing the functional rs6726639. This variant also shows the hallmarks for being the driver of the GWAS for reduced risk of liver fibrosis at *MERTK*.

1326F

Association of polymorphisms Glu298Asp and 4b/a of *eNOS* gene with renal function parameters in Mexican patients with Fabry disease. *A. Marín-Medina[1,2], J.J. Gómez-Ramos[3], G. Rodríguez-Pinto[1], L.E. Figuera-Villanueva[1,2].* 1) Genetica , Centro de Investigación Biomedica de Occidente (CIBO) IMSS , Guadalajara , Jalisco, Mexico; 2) Doctorado en Genetica Humana-CUCS Universidad de Guadalajara; 3) Departamento de Urgencias, HGZ 89 IMSS Guadalajara, Jalisco,México.

Fabry disease (FD) is an X-linked inherited lysosomal disease that courses with renal failure in an important percentage of the affected individuals. On the other hand, *eNOS* gene encodes to the endothelial nitric oxide synthase that plays an important role in glomerular hemodynamics. This gene has 2 main polymorphisms (Glu298Asp and the 4b/a) that have been studied in many diseases including those with cardiovascular and renal alterations. Considering the lack of information, we investigated if there was association of these genetic variants with renal function parameters in Mexican patients with FD and renal impairment. We included 15 patients with FD and renal alterations and associated these *eNOS* polymorphisms with the renal function parameters (urea, creatinine and glomerular filtration rate or GFR) taken from the clinical records at the first visit to the nephrology service. The alleles Asp298 and 4a of *eNOS* gene were significantly associated with increased levels of urea, creatinine and decreased GFR in the patients included that behaved in a co-dominant fashion. Our results coincide with other reports showing an association of these polymorphisms with kidney disease and others involved in the the nitric oxide pathway and suggest that these variants affect the severity of nephropathy in patients with FD.

1327W

Whole genome sequencing in severe chronic obstructive pulmonary disease. *M.H. Cho[1], D. Prokopenko[3], H. Fier[3], D. Qiao[1], C. Lange[4], T.H. Beaty[2], J.D. Crapo[5], E.K. Silverman[1], COPDGene Investigators, NHLBI TOPMed Investigators.* 1) BWH, Boston, MA; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) University of Bonn, Bonn; 4) Harvard School of Public Health, Boston, MA; 5) National Jewish Health, Denver, CO.

Background: Genome-wide association studies have identified common variants associated with chronic obstructive pulmonary disease (COPD). Rare variants can also increase COPD susceptibility. Whole-genome sequencing (WGS) offers several advantages over exome sequencing, including improvements in calling of coding regions and interrogation of impactful variants in non-coding regions. We hypothesize that WGS in subjects with severe COPD and smoking controls with preserved pulmonary function will allow us to identify novel genetic determinants for COPD. Methods: As part of the NHLBI Trans☐Omics for Precision Medicine (TOPMed) program, we submitted DNA samples for sequencing from 600 severely affected white and 330 African-American COPD cases with a mean forced expiratory volume in 1 second (FEV_1) of 32% and 35% predicted, respectively. Current and ex-smoker controls (500 white and 570 AA) had FEV_1 > 85%. Samples were sequenced at > 30x coverage using PCR-free libraries, with centralized mapping and variant calling by the University of Michigan Informatics Research Core. We adjusted for population substructure using the Jaccard method and grouped variants using a region-based approach. We performed preliminary single-variant and gene-based analysis both genome-wide and in candidate regions. Results: We performed an initial analysis in 689 subjects. In this small sample size, we identified 15q25 ($p = 9 \times 10^{-6}$) in whites and *THSD4* ($p = 2 \times 10^{-5}$) in African-Americans among the top SNPs in genome-wide single variant analyses. In a region-based assessment, top results included the 15q25 region ($p = 4 \times 10^{-7}$). Conclusions: Whole-genome sequencing can identify large numbers of potentially functional and deleterious variants, and will serve as an important resource for identifying causal variants at known and novel loci for COPD. Future plans include association tests in the larger sample for affection status and for secondary phenotypes including imaging and transcriptomics data. Funding: This work was supported by NHLBI R01 HL084323, P01 HL083069, P01 HL105339 and R01 HL089856 (E.K.S.); R01 HL113264 (M.H.C. and E.K.S.), and R01 HL089897 (J.D.C.). The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, GlaxoSmithKline, Siemens and Sunovion.

1328T

Genetics of surfactant protein D serum levels and chronic obstructive pulmonary disease.

M. Obeidat[1], N. Fishbane[1], N. Hansel[2], Y. Bossé[3], P. Joubert[3], K. Hao[4], D. Nickle[5], D. Postma[6], W. Timens[7], N. Rafaels[8], R. Mathias[9], I. Ruczinski[10], T. Beaty[11], K. Barnes[8], P. Paré[1], D. Sin[1]. 1) Centre for Heart Lung Innovation, University of British Columbia, Vancouver, BC, Canada; 2) Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA; 3) 3 Department of Molecular Medicine, Laval University, Québec, Canada; 4) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, USA; 5) Merck Research Laboratories, Boston, MA, USA; 6) University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC research institute, Groningen, The Netherlands; 7) University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, GRIAC research institute, Groningen, The Netherlands; 8) Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO, USA; 9) Division of Genetic Epidemiology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA; 10) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 11) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

Background: Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide. Surfactant protein D (SPD) is lung-specific protein that is involved in the innate immune response and clearance of pathogens. SPD can be detected in peripheral blood and serum/plasma concentrations have been shown to negatively correlate with lung function and health status in COPD. We hypothesized that SPD serum protein and lung mRNA levels are under genetic control in COPD subjects. **Methods**: The Lung Health Study (LHS) is a longitudinal multicenter study of mild-to-moderate COPD subjects, who were followed over 11 years. In year 5, SPD serum levels were measured in 4,061 individuals who also had whole genome genotypes. A genome-wide association study (GWAS) for SPD levels was undertaken using 1000 genome imputed genotypes, assuming an additive genetic model, and adjusting for age, sex, BMI, and genetic principal components. The SPD-associated single nucleotide polymorphisms (SNPs) were tested for lung expression quantitative trait loci (eQTLs) in the lung eQTL study (n=1,111) and additionally tested for association with lung function in LHS subjects and in a number of large publically available lung function GWAS data sets including the UKBiLEVE study (n=15,000) and the SpiroMeta consortium (n=38,000). **Results**: There were 4 loci associated with SPD levels. The strongest association was on chr10 near the SPD encoding gene *SFTPD* (P=1.1E-81) and SNPs in this locus were lung eQTLs for *SFTPD* and interestingly for two additional genes; annexin A11 (*ANXA11*) and tetraspanin 14 (*TSPAN14*). The second SPD locus included the major histocompatibility complex region on chr6 (P=9.33E-55). The third included SNPs spanning the ATPase secretory pathway Ca2+ transporting 2 (*ATP2C2*) gene region on chr16 (P= 1.52E-16) and were also lung eQTLs for *ATP2C2*. The fourth locus included SNPs which reside in, and are lung eQTLs for the complement component 4 binding protein alpha (*C4BPA*) gene on chr1 (P= 3.75E-09). The sentinel SNPs on chr 1, 6, and 10 loci showed association (P<0.05) with lung function in LHS and in publically available lung function GWAS data from the UKBiLEVE study and the SpiroMeta consortium. **Conclusion**: Integrating genome-wide protein and mRNA genetic determinants of SPD and lung function phenotypes provided significant insights into SPD biology and supported the role of SPD in the pathogenesis of COPD. SPD represents a promising biomarker for COPD.

1329F

Novel genetic variants for cartilage thickness and hip osteoarthritis.

J. van Meurs[1], M. Castano-Betancourt[2], D. Evans[1], C. Boer[3], Y. Ramos[4], S. Metrustry[4], Y. Liu[5], W. den Hollander[4], J. van Rooij[1], V. Kraus[6], M. Yau[7], B. Mitchell[8], A. Hofman[9,10], M. Doherty[11], S. Doherty[11], W. Zhang[11], R. Kraaij[1], F. Rivadeneira[1], E. Barrett-Conner[12], R. Maciewicz[13], N. Arden[14], R. Nelissen[15], M. Kloppenburg[16], J. Jordan[17], M. Nevitt[18], F. Lafeber[19], N. Lane[17], I. Meulenbelt[3], A. Valdes[11]. 1) Internal Medicine, Erasmus Medical Center, Rotterdam, Netherlands; 2) California Pacific Medical Center Research Institute, San Francisco, CA, USA; 3) Department of Medical Statistics and Bioinformatics, Section Molecular Epidemiology. Leiden University Medical Center, Leiden, The Netherlands; 4) Department of Twins Research and Genetic Epidemiology Unit, King's College London, London, UK; 5) Thurston Arthritis Research Center, Department of Medicine, and Department of Orthopedics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 6) Duke Molecular Physiology Institute and Division of Rheumatology. Duke University School of Medicine, Durham, NC, USA; 7) Departments of Medicine and Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD, USA; 8) Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD 21201, USA; 9) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 10) Department of Epidemiology, Harvard T.H. School of Public Health, Boston, MA, USA; 11) School of Medicine, University of Nottingham, Nottingham, UK; 12) Department of Family Medicine and Public Health UCSD Medical Center, San Diego, USA; 13) Respiratory, Inflammation, Autoimmunity Innovative Medicines, AstraZeneca AB, Mölndal, Sweden; 14) Nuffield Department of Orthopaedics, Rheumatology and musculoskeletal sciences, University of Oxford, UK; 15) Department of Orthopedics, Leiden University Medical Center, Leiden The Netherlands; 16) Department of Rheumatology and Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 17) Thurston Arthritis Research Center, The University of North Carolina, Chapel Hill, USA; 18) University of California at San Francisco, San Francisco, CA; 19) University Medical Center Utrecht, the Netherlands.

Osteoarthritis is one of the most frequent and disabling diseases of the elderly. Only few genetic variants have been identified for osteoarthritis, which is partly due to large phenotype heterogeneity. To reduce heterogeneity, we here examined cartilage thickness, one of the structural components of joint health. We conducted a genome-wide association study of minimal joint space width (mJSW), a proxy for cartilage thickness, in a discovery set of 13,013 participants from five different cohorts and replication in 8,227 individuals from seven independent cohorts. We identified five genome-wide significant (GWS, P≤5·0×10⁻⁸) SNPs annotated to four distinct loci. In addition, we found two additional loci that were significantly replicated, but results of combined meta-analysis fell just below the genome wide significance threshold. The four novel associated genetic loci were located in/near *TGFA* (rs2862851), *PIK3R1* (rs10471753), *SLBP/FGFR3* (rs2236995), and *TREH/DDX6* (rs496547), while the other two (*DOT1L* and *SUPT3H/RUNX2*) were previously identified. A systematic prioritization for underlying causal genes was performed using diverse lines of evidence. Exome sequencing data (n=2,050 individuals) indicated that there were no rare exonic variants that could explain the identified associations. In addition, *TGFA, FGFR3* and *PIK3R1* were differentially expressed in OA cartilage lesions versus non-lesioned cartilage in the same individuals. In conclusion, we identified four novel loci (*TGFA, PIK3R1, FGFR3 and TREH*) and confirmed two loci known to be associated with cartilage thickness. The identified associations were not caused by rare exonic variants. This is the first report linking *TGFA* to human OA, which may serve as a new target for future therapies.

1330W

Mutation of the planar cell polarity gene *VANGL1* **in adolescent idiopathic scoliosis.** *L.A. Larsen[1], M.R. Andersen[1], M. Farooq[1], K. Koefoed[1], K.W. Kjaer[1], A. Simony[2], S.T. Christensen[3].* 1) Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; 2) Center for Spine Surgery & Research Middelfart, Region of Southern Denmark, Denmark; 3) Department of Biology, University of Copenhagen, Denmark.

Adolescent idiopathic scoliosis (AIS) is a spinal deformity which occurs in 1-3% of the population. The cause of AIS is often unknown, but genetic factors are important in the etiology. We performed mutation screening in 157 unrelated patients with moderate to severe AIS and identified two rare missense mutations in *VANGL1*, encoding a receptor involved in WNT/planar cell polarity (PCP) signaling. The mutations, p.I136N and p.F440V, are predicted to be damaging and to affect evolutionary conserved amino acid residues of VANGL1. Functional analysis in Madin-Darby Canine Kidney cells showed that both mutations abolish the normal translocation of VANGL1 to the cell membrane. Our data suggest a possible involvement of *VANGL1* in the aetiology of AIS and further analysis of the role of WNT/PCP signalling in AIS is warranted.

1331T

The rare mutations of *MYH3* **gene is an important etiology of congenital scoliosis.** *W. Chen[1,2], Y. Zuo[2], J. Liu[3], Z. Liu[2], S. Liu[2], G. Qiu[2,4], N. Wu[2,4,5], Z. Wu[1,4,6].* 1) Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, Beijing, China; 2) Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 3) Breast Surgical Oncology, Cancer Hospital of Chinese Academy of Medical Sciences, Beijing 100021, China; 4) Beijing Key Laboratory for Genetic Research of Skeletal; 5) Medical Research Center of Orthopaedics, Chinese Academy of Medical Sciences, Beijing China; 6) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China.

Introduction: Congenital scoliosis (CS) is a rare and severe congenital disease with an estimated prevalence of 0.5 to 1 in 1000 live births. It is defined as a lateral curvature of the spine exceeding 10 degrees caused by congenital vertebral malformation (CVM). Etiology of CS is very complex. Previously, we have elucidated the genetic basis of a significant portion of sporadic CS by a model of compound inheritance (Wu, et al 2015 NEJM 372: 341-350). *MYH3* encodes muscle embryonic myosin heavy chain, which is associated with Distal Arthrogryposis (DA) type 2A, type 2B and type 8. In DA type 8, *MYH3* mutations were reported to be associated with vertebral fusion. In this study, we report that the mutations of *MYH3* gene may contribute to the genetic etiology for CS. **Methods:** We enrolled patients diagnosed as CS in Peking Union Medical College Hospital (PUMCH) between October 2010 and November 2015. The CS diagnosis were confirmed by radiological imaging. The patients with known syndromes were excluded. We performed whole exome sequencing (WES) in 6 trios with additional 316 patients target sequenced by an in silicon designed target panel. **Results:** We identified one de novo *MYH3* mutation (c.841G>A,p.E281K) in one patient by WES. This variant did not exist in ExAC, 1000 Genome and ESP6500 database. The prediction of Polyphen2 and SIFT were probably damaging or deleterious,respectively (the Polyphen2 score is 1, the SIFT score is 0,the GERP score is 5.11 and the CADD score is 26). Notably, the patient presents as CS, unsegmentation in Thoracic vertebrae and distal arthrogryposis. Furthermore, we identified 6 patients with *MYH3* missense mutations (c.265G>A,p.E89K; c.1400A>C,p.E467A; c.1802C>T,p.P601L; c.3449G>A,p.R1150Q; c.3526C>T,p.R1176C; c.5134G>A,p.E1712K) in 316 patients highly conserved domain (6/316, 1.9%). **Conclusion**: Up to our knowledge, our findings indicated that *MYH3* might affect vertebral development. Our data showed that about 2% CS patients carried highly pathogenic *MYH3* mutations,firstly indicating *MYH3* might be another important gene causing CS.

1332F

RNA expression elucidates roles of osteoporosis candidate genes in skeletal fragility. *E. Quillen[1], T.L. Bredbenner[2], J.P. Glenn[1], V. Mattern[1], A.E. Nichols[2], D.E. Moravits[2], L.M. Havill[3], L.A. Cox[1], D.P. Nicolella[2].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Musculoskeletal Biomechanics, Southwest Research Institute, San Antonio, TX; 3) Southwest National Primate Research Center, San Antonio, TX.

Skeletal fragility is a major risk factor for osteoporotic fracture but is incompletely indexed by the most common clinical indicator, bone mineral density. The structural integrity of bone is a complex phenotype with many contributing genetic, environmental, and lifestyle factors. Using the baboon as a naturally occurring model of osteoporosis (OP), we identify 31 differentially expressed transcripts associated with bone fragility traits and biomarkers in 100 adult animals (50 female) with a wide age range. No animals were sacrificed for this study; all died naturally or were euthanized for other reasons. Both femurs were scanned by microCT, digitally reconstructed, and total variation reduced into a statistical shape model (SSM). Fracture strength was determined in a non-habitual loading direction equivalent to a side-ways fall. Cortical bone from the femur mid shaft was mechanically tested to determine tensile material properties. Osteopontin (OPN) concentrations were measured in the serum. Total RNA was extracted from white blood cells collected at necropsy and mRNA libraries generated using the KAPA Stranded mRNA-Seq kit for sequencing on the HiSeq2500. Reads were aligned to the *Papio anubis* reference genome using the STAR aligner and normalized to total read counts per sample. Read counts were scaled across samples and adjusted for age and sex effects. For 909 transcripts derived from known OP candidate genes selected from the HuGE Navigator, normalized counts were assessed for association with bone health traits using a variance component model in SOLAR. After Bonferroni correction, 31 were associated with one or more structural traits, mechanical strength, or biomarkers. Thirteen transcripts were associated with overall femur size while 7 were associated with OPN including *GPNMB* which regulates bone turnover in a mouse model. Matrix metalloproteinase *MMP1* and insulin-regulating *PCSK1* were associated with SSM factors. *TTC23*, a locational candidate gene from previous OP meta-analyses, was associated with tensile strength. *CYP27B1*, a contributor to age-related decline in bone mass in a rat model, was associated with bone strength in the fall simulation. These findings illuminate potential mechanisms for previously identified genetic variants while clarifying the role of these genes in the complex interactions contributing to bone fragility. This work was funded by NIH/NIAMS R01 AR060341-03 and the William & Ella Owens Medical Research Foundation.

1333W

Genome-wide association study of DNA methylation identifies a novel locus associated with bone mineral density. *J.A. Morris[1], P.-C. Tsai[2], Y.-H. Hsu[3], R. Joehanes[3], J. Zheng[4], K. Trajanoska[5], M. Soerensen[6], V. Forgetta[1], M. Frost[6], K. Christensen[6], L. Christiansen[6], F. Rivadeneira[5], T.D. Spector[2], J.H. Tobias[4], D.M. Evans[4], D.P. Kiel[3], J.B. Richards[1,2], J.T. Bell[2].* 1) Department of Human Genetics, Lady Davis Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Twin Research & Genetic Epidemiology, King's College London, London, UK; 3) Institute for Aging Research, Hebrew SeniorLife, Department of Medicine, BIDMC, & Harvard Medical School, Boston, Massachusetts, USA; 4) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 5) Departments of Internal Medicine and Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands; 6) The Danish Twin Registry, Institute of Public Health, Endocrine Research Unit, KMEB, University of Southern Denmark, Odense, Denmark.

Purpose Osteoporosis is a common, complex disease characterized by increased bone fragility, often resulting in fracture. Bone mineral density (BMD) is measured to diagnose osteoporosis and estimate fracture risk. Genetic and environmental determinants of osteoporosis risk may converge through the epigenome, providing a tool to better understand osteoporosis pathophysiology. As the epigenetics of BMD have been largely unexplored in humans, we undertook an epigenome wide association study (EWAS) of BMD. **Methods** We undertook an international, large-scale BMD EWAS using the Illumina Infinium HumanMethylation450 array to measure site-specific DNA methylation across four cohorts in up to 4,827 individuals of European descent. Using these data, we aimed to identify differentially methylated probes (DMPs) associated with dual energy X-ray absorptiometry derived femoral neck and lumbar spine BMD. DNA methylation was measured in whole-blood tissue as epigenetic status is partially shared across tissues, immune cells may influence BMD, and some bone cells arise from monocyte-macrophage precursors. We performed sex-combined and stratified analyses, controlling for age, weight, smoking status, sex (for combined analyses), estimated white blood cell proportions, and random effects for relatedness and batch effects. We defined statistical significance of DMPs as being associated with BMD at a 5% false-discovery rate (FDR). **Results** We identified one DMP significantly associated with femoral neck BMD in females-only ($P = 7.9x10^{-11}$; $N = 3,444$) and pooled ($P = 3.0x10^{-8}$; $N = 4,827$) analyses. The DMP maps to the liver carboxylesterase 1 (*CES1*) gene 5' UTR. This association postulates a novel locus for bone biology as no genetic associations with BMD and *CES1* have been reported to date. The association is not mediated by genetic variation mapping to the DMP, as the association remains significant after conditional analyses on SNPs at the locus. **Conclusion** Undertaking the most comprehensive genome-wide analysis to-date for the role of whole-blood methylation on BMD variation, we identified a novel epigenetic locus for bone biology mapping to *CES1*. Further follow-up experiments to understand the function of DNA methylation and *CES1* with respect to BMD are underway. Interrogation of the function of this region will help to understand how environmental and genetic effects converge upon the epigenome to regulate BMD.

1334T

Clubfoot-associated *FSTL5* implicates the FSTL5/YAP axis in embryonic limb development. *J. Rios[1,2,3,4], A. Khanshour[1], L. Fraga[5], A. Diamond[6], V. De Mello[5], J. Kozlitina[2], M. Wilson[6], J. Hecht[7], B. Richards[8], M. Dobbs[9], H. Wackerhage[5], N. Vargesson[5], C. Gurnett[10].* 1) Research Department, Texas Scottish Rite Hospital for Children, Dallas, TX; 2) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX; 3) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; 4) Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX; 5) School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, United Kingdom; 6) Department of Anatomy, University of Otago, Dunedin, New Zealand; 7) McGovern Medical School, University of Texas Health, Houston, TX; 8) Department of Orthopaedic Surgery, Texas Scottish Rite Hospital for Children, Dallas, TX; 9) Department of Orthopaedic Surgery, School of Medicine, Washington University, St Louis, MO; 10) Department of Neurology, School of Medicine, Washington University, St Louis, MO.

Congenital clubfoot, or talipes equinovarus (TEV), is a common structural birth defect occurring twice as frequently in males than females. TEV presents with bilateral or unilateral involvement at similar frequencies and is associated with maternal smoking during pregnancy. TEV may develop secondary to other musculoskeletal disease (syndromic TEV) or as an isolated idiopathic TEV (iTEV). The transcription factors *PITX1* and *TBX4* were previously implicated in TEV and are known to regulate embryonic limb specification and development, though pathogenic variants occur rarely in TEV patients and neither has been associated with TEV in population studies. We performed a genome-wide association study (GWAS) following imputation of 81.2 million variants in 399 non-Hispanic white (NHW) iTEV patients and 7,820 NHW controls. After quality control filtering, a total 8.6 million variants were included in the association analysis. GWAS analysis identified a significantly associated locus (rs76973778, P_{GWAS}=1.92×10^{-8}, OR=0.348) within intron 3 of the *FSTL5* gene, encoding follistatin-like 5. A highly correlated SNP (rs76811724, r^2=0.96, P_{GWAS}=2.06×10^{-7}) was tested for replication in an independent cohort of 335 and 2,738 NHW cases and controls, respectively. Following replication, the association of rs76811724 improved to $P_{COMBINED}$=2.99×10^{-9}. Though largely uncharacterized, the FSTL5 protein is similar to follistatin-like 1, an antagonist of bone morphogenetic protein that is required for proper hindlimb skeletal development in mice. Expression profiling of developing embryonic limb buds showed *Fstl5* expression increased throughout the time points investigated in both bat (CS15-17) and chicken (S15-30) embryos as well as mouse (E12-13.5), albeit at lower levels. *Fstl5* expression was detected in the distal segments of developing embryonic chick wings by whole-mount *in situ* hybridization, suggesting *Fstl5* may be expressed later in limb development and may regulate distal limb formation. FSTL5 was recently shown to negatively regulate *YAP1* and may also regulate expression of *VCAN*, which encodes versican, a regulator of TGF-beta signaling known to be required for proper joint development in the distal hind limbs of both mouse and chicken embryos. Using genome-wide association analysis, we identified the *FSTL5* gene as the first locus associated with iTEV, suggesting a role for the FSTL5/YAP/Hippo axis in regulating later stages of embryonic limb development.

1335F

Association of 3q13.32 variants with hip trochanter and intertrochanter bone mineral density identified by a genome-wide association study. *L. Zhang[1,2], Z.G. Xie[3], W.Z. Hu[1,2], L.B. Li[3], S. Ran[4], Y. Lin[4], R. Hai[5], H. Shen[6], Q. Tian[6], Y.H. Zhang[2,7], S.F. Lei[1,2], C.J. Papasian[8], H.W. Deng[6], Y.F. Pei[2,7].* 1) Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University, Suzhou, Jiangsu, China; 2) Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Suzhou, Jiangsu, China; 3) The Second Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China; 4) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai, P. R. China; 5) Inner Mongolia Autonomous Region People's Hospital, Hohhot, Inner Mongolia, P. R. China; 6) Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA; 7) Department of Epidemiology and Health Statistics, School of Public Health, Soochow University, Suzhou, Jiangsu, China; 8) Department of Basic Medical Science, University of Missouri-Kansas City, Kansas City, MO, USA.

Hip trochanter (TRO) and intertrochanter (INT) sub-regions have important clinical relevance to subtrochanteric and intertrochanteric fractures, but have rarely been studied by genome-wide association studies (GWASs). Aiming to identify genomic loci associated with BMD variation at TRO and INT regions, we performed a GWAS utilizing the Framingham heart study (FHS, *N*=6,912) as discovery sample, and utilized the Women's health initiative (WHI) African-American sub-sample (*N*=845), WHI Hispanic sub-sample (*N*=446), and Omaha osteoporosis study (*N*=963), for replication. Combining the evidence from both the discovery and the replication samples, we identified one novel locus around the associated variations at chromosomal region 3q13.32 (*rs1949542*, discovery p=6.16×10^{-8}, replication p=2.86×10^{-4} for INT-BMD; discovery p=1.35×10^{-7}, replication p=4.16×10^{-4} for TRO-BMD, closest gene *RP11-384F7.1*). We also replicated two loci at chromosomal regions 3p21 (*rs148725943*, discovery p=6.61x10^{-7}, replication p=5.22×10^{-4} for TRO-BMD, closest gene *CTNNB1*) and 8q24 (*rs7839059*, discovery p=2.28×10^{-7}, replication p=1.55×10^{-3} for TRO-BMD, closest gene *TNFRSF11B*) that were reported previously. We demonstrated that the effects at both 3q13.32 and 3p21 were specific to the TRO, but not to the femoral neck and spine. In contrast, the effect at 8q24 was common to all the sites. Our findings provide useful insights that enhance our understanding of bone development, osteoporosis, and fracture pathogenesis.

1336W

A genome-wide association study identifies novel genetic loci for pediatric atopic dermatitis in European and African American populations. *B. Almoguera[1], L. Vazquez[1], J. Connolly[1], A. Cederquist[1], E. Ryan[1], M. Khan[1], F. Mentch[1], P.L. Peissig[2], J.G. Linneman[2], M. Brilliant[2], P. Sleiman[1], H. Hakonarson[1].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA.

   Introduction: Atopic dermatitis (AD) is a common chronic inflammatory skin disease with a major genetic component. To date, 31 genetic loci have been identified through GWAS but despite AD being more prevalent in African American populations, this group has been under-represented in these studies. Here we present a GWAS with the largest African American pediatric population studied to date where we have identified a variant in chromosome 1 significantly associated with AD. **Methods**: Cases and controls were defined by a phenotyping algorithm that mines electronic medical record (EMR) data for diagnostic (i.e. ICD9 codes) and medication information. The algorithm was run and validated internally at the Center for Applied Genomics (CAG), with external validation at an independent site from the electronic Medical Records and Genomics network (eMERGE). After imputation of non-observed genotypes with SHAPEIT and IMPUTE2, a GWAS was performed in European and African American populations separately using SNPtest, including sex and 10 principal components as covariates, followed by meta-analysis with METAL. **Results**: Validation of the phenotyping algorithm yielded positive predictive values of 92%-96%. A total of 9,075 subjects were included in the analysis: 4,014 African American (1,415 cases) and 5,061 individuals of European ancestry (231 cases). Meta-analysis of the GWAS results identified rs6665826 in locus 1q24.3 reaching genome-wide significance ($p=2.40 \times 10^{-8}$). This variant is located 159kb downstream of *TNFSF18*, which encodes the member 18 of the tumor necrosis factor superfamily, and has a *cis*-regulatory effect over *FASLG* (Fas ligand) in whole blood. The region is a known risk factor for Crohn's disease. Two additional loci with suggestive significance were found: variant rs7540799 ($p=1.14 \times 10^{-7}$) in the gene *RORC* (RAR-related orphan receptor C) also in chromosome 1, associated with intestinal bowel disease (IBD) and rs79790432 in chromosome 4 ($p=1.82 \times 10^{-7}$), upstream the member 1 RasGEF domain family gene (*RASGEF1B*) and expressed in skin. **Conclusion**: We have identified a novel loci for AD in children of European and African American ancestry with intriguing function and previously associated with IBD, which is consistent with the known genetic overlap of this condition with other inflammatory and autoimmune diseases. Replication of these results in an independent sample is currently in progress.

1337T

A genome-wide analysis of severe cutaneous adverse drug reactions to anti-epileptic drugs. *M. McCormack[1], G. Wright[2], U. Amstutz[3], N. Shear[4], A. Gamberdella[5], G. Sills[6], Z. Rener-Premec[7], H. Lerche[8], W. Kunz[9], C. Depondt[10], F. Zara[11], P. Striano[12], N. Delanty[1], A. Ingason[13], S. Diehl[14], R. Krause[14], S. Sisodiya[15], M. Hayden[2], C. Ross[16], G.L. Cavalleri[1], B. Carleton[17], EpiPGX Consortium, Canadian Pharmacogenomics Network for Drug Safety.* 1) Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland; 2) Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada; 3) University Institute of Clinical Chemistry, Inselspital Bern University Hospital, University of Bern, Bern, Switzerland; 4) Dermatology and Toxicology, Department of Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; 5) Department of Medical Sciences, Institute of Neurology, University Magna Graecia, Catanzaro, Italy; 6) Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK; 7) Department of Pediatric Neurology, University Children's Hospital, Ljubljana, Slovenia; 8) Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; 9) Division of Neurochemistry, Department of Epileptology, University Bonn Medical Center, Bonn, Germany; 10) Department of Neurology, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium; 11) Department of Head and Neck Diseases, "G. Gaslini" Institute, Genova, Italy; 12) Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, "G. Gaslini" Institute, Genova, Italy; 13) DeCode Genetics, Reykjavik, Iceland; 14) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; 15) Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, National Hospital for Neurology and Neurosurgery, London, UK; 16) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, Canada; 17) Division of Translational Therapeutics, Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada.

   Anti-epileptic drugs (AEDs) are widely used as a treatment for epilepsy. However, they commonly cause cutaneous adverse drug reactions (cADRs) ranging from mild rash to potentially fatal hypersensitivity syndrome (HSS) and Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) that have an annual incidence of 0.4 to 1.2 individuals per million. Alleles in the major histocompatibility locus, *HLA-A*31:01* and *HLA-B*15:02*, and the cytochrome P450 allele, *CYP2C9*3*, have diagnostic value in predicting cADR risk in various ethnic backgrounds, yet these markers do not have complete prognostic sensitivity and specificity, and risk factors in patients experiencing cADRs but not carrying these markers remain elusive. In order to address this need, we have collected detailed point-of-care phenotypic and genomic information for epilepsy patients via two large consortiums across Europe and North America, providing a framework to study these cADRs. To date, we have assembled 80 cases of severe cADRs (n=41 HSS and n=39 SJS/TEN) associated with carbamazepine, oxcarbazepine, lamotrigine and phenytoin and 198 matched drug-tolerant controls of European descent. In addition, we have collected a further 18 cases of other ancestral backgrounds. We are currently performing a genome-wide assessment of genetic variation employing both high throughput genotyping arrays and exome sequencing. These approaches will enable us to analyse our cohort for association of severe cADRs with common and rare variation. Preliminary analyses show replication of the *HLA-A*31:01*/carbamazepine and *CYP2C9*3*/phenytoin cADR associations, while the *HLA-B*15:02* allele was not observed in our European patients, which is in line with previous studies. In conclusion, the present study represents one of the largest collections of severe cADRs caused by AEDs in European ancestral patients assembled to date. Identification of genomic markers for AED-induced severe cADRs will facilitate more personalized prescribing and is expected to reduce the overall rates of AED discontinuation due to ADRs as well as improve the safety of anticonvulsant therapies. .

1338F

Pleiotropic effects of noncoding variants near *EFEMP1*, *WT1*, *ADAMTS6*, **and** *EBF2* **on the risk of abdominal hernias support an etiological role for collagen and elastin homeostasis in hernia susceptibility.** *E. Jorgenson[1], N. Makki[2,3], K.K. Thai[1], D.C. Chen[4], W. Eckalbar[2,3], A. Avins[1], N. Ahituv[2,3].* 1) Division of Research, Kaiser Permanente, Oakland, CA; 2) Department of Bio-engineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA, 94158, USA; 4) Lichtenstein Amid Hernia Clinic, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095, USA.

Abdominal hernias are some of the most frequently diagnosed conditions in clinical practice, with more than twenty million hernia repair surgeries performed annually around the world. While family studies suggest that genetic factors play an important role in hernia susceptibility, studies assessing specific genetic risk factors have been limited. Because the risks associated with hernias, specifically, bowel incarceration with a substantial risk of mortality, must be balanced against the risks associated with their treatment, which include chronic pain (6%) and hernia recurrence (10%), there is a clear need for a better understanding of hernia etiology and improved treatment options. We conducted the first large-scale genetic study of hernia risk, identifying noncoding variants at four novel genetic loci underlying the risk of inguinal hernia—the most common type of hernia. We showed that four genes in these loci (EFEMP1, WT1, EBF2, and ADAMTS6) are expressed in mouse connective tissue. We investigated whether risk SNPs in these loci were associated with other abdominal hernia subtypes, specifically, umbilical, femoral, and ventral hernias. We observed significant associations between the EFEMP1 and WT1 loci and each of these subtypes, including a genome-wide significant association between EFEMP1 and ventral hernia (p = 7.9*10-10). Pathway analysis suggested that WT1 and EFEMP1 might influence connective tissue maintenance/homeostasis through their action on extracellular matrix enzymes, including matrix metalloproteinases that degrade collagen and elastin fibers. These findings indicate that these loci affect the risk of hernias across anatomical sites and suggest their mechanism of action may underlie hernia susceptibility more generally.

1339W

Association between single nucleotide polymorphisms of *BMP5* **gene and risk of knee osteoarthritis.** *A. Sharma, R.N. Srivastava, S.R. Srivastava, S. Raj.* orthopaedic surgery, king george medical university, Lucknow, Uttar Pradesh, India, India.

Background: Osteoarthritis (OA) is a chronic degenerative disorder probably affected by both genetics and environmental causes. Bone morphogenetic proteins (BMPs) are bone-derived factors that can induce new bone formation. Previous study identified single nucleotide Polymorphisms (SNP□s) that altered transcriptional activity of the BMP5 promoter, which implies that these polymorphism of BMP5 gene is involved in OA susceptibility. In this case control study, we investigated the correlation of rs1470527 and rs9382564 SNP of BMP5 gene with susceptibility to knee OA. Methods: A total of 499 cases that confirmed radiographic knee OA and 458 age and sex matched healthy controls were enrolled. Venous blood samples were obtained from all cases as well as controls for genetic analysis. Restriction fragment length polymorphism was performed for SNP analysis. Results: The genotype distribution for rs1470527and rs9382564 SNP was significantly different between the cases and controls (P<0.0001). Within both the SNPs of BMP5 gene, genotype CT and TT was found to be significantly (p<0.0001) associated with knee OA as compared with the CC genotype. In addition when alleles were compared, T allele of both the studied SNP was observed to be significantly associated with knee OA (P<0.0001). Further in relation with clinical symptom of OA, we observed signification association of TT genotype with both visual analogue scale (VAS) (p<0.0001) and WOMAC (Western Ontario and McMaster Universities) score (p<0.05). Conclusion: Our results indicate significant association of rs1470527 and rs9382564 polymorphism of BMP5 gene with knee OA.

1340T

Two novel hidradenitis suppurativa risk loci identified by genome-wide association study. *M. Liu, J. Degner, A. Grundstad, H. Teixeira, D. Williams, W. Davis, J. Waring.* Abbvie Inc, North Chicago, IL.

   Hidradenitis suppurativa (HS, also known as acne inversa) is an autosomal chronic inflammatory inherited skin disease affecting approximately 1% of the population. Genes involved in HS susceptibility have yet to be determined. A total of 359 DNA samples from HS subjects enrolled in Pioneer I and II (NCT01468207 and NCT01468233) were combined with 188 healthy controls in order to conduct a case/control GWAS of HS risk. Genotyping was performed using the Illumina HumanOmniExpressExome v1.2B bead-chip and imputation made to the 1KG Phase III reference panel using BEAGLE 4.1, resulting in approximately 31.6 million candidate SNPs for association. We identified two genome-wide significant (p<5x10^{-8}) regions in Chromosomes 4 and 13. The assoiated LD block in Chr13 covers *SLC25A15* and some intergenic space near *ELF1*. Variants in the region were previously reported to associate with other autoimmune diseases such as Crohn's and Systemic Lupus Erythematosus; for these other diseases, *ELF1* was proposed as the causal gene because of its involvement in T-cell specific gene regulation, including regulation of *CD4*. However, public eQTL databases revealed that the putative SNPs associated with HS are significant eQTLs with *WBP4* in skin tissue and with *SLC25A15* in muscles. Interestingly, *SLC25A15* is a key enzyme that catalyzes the urea cycle for transporting ornithine which has been shown to enhance wound healing. Furthermore, an assessment of several studies in psoriasis showed that *SLC25A15* is more elevated in lesion tissue than non-lesional tissue in psoriasis patients. Potential casual variants in the associated Chr4 LD-block include *PPEF2, ASAHL* and *CXCL9/10/11,* with the later all being chemokines important in psoriasis disease etiology and whose levels in serum has been shown to increase during wound healing. Moreover, *CXCL9/10* expression is 3 to 4 fold greater in skin samples from lesions as compared to normal individuals or from non-lesion tissue in HS patients. *PPEF2* is involved in epidermal barrier function while *ASAHL* inhibitors have been proposed as therapeutic agents in the treatment of inflammatory disease such as allergic dermatitis and psoriasis. Ongoing functional assays will help us to dissect the causal pathway by which these variants lead to differences in HS susceptibility. * Liu M and Degner JF contributed equally to this work.

1341F

Lack of predictive deep phenotyping variables in *PIK3CA* mosaicism. *S. Polubothu[1and2], R. Knox[3], L. Al-Olabi[1], V. Parker[3], R. Semple[3], V. Kinsler[1and2].* 1) Genetics and Genomic Medicine, UCL Institute of Child Health, London, United Kingdom; 2) Paediatric Dermatology, Great Ormond St Hospital for Children, London; 3) WT-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom.

   PIK3CA Related Overgrowth Spectrum (PROS) is an umbrella term describing the known and emerging clinical entities associated with post-zygotic *PIK3CA* mutations. More recently *PIK3CA* mutations have been described in venous malformations without overgrowth. Many individuals present with overlapping features of syndromes previously thought to be clinically distinct. Accurate phenotypic and genotypic description of the clinical spectrum is necessary to improve our understanding of the condition, and to direct individuals with *PIK3CA* mutations to international trials of MTOR inhibitors. We undertook targeted next generation sequencing of 60 known overgrowth genes on tissue samples from a cohort of 85 children with a diagnosis of vascular malformation and/or overgrowth, on whom genotyping results are so far available for 43 (all will be available by the time of the ASHG 2016). Sequencing had a sensitivity of 1% for fresh and 5% for paraffin-embedded samples, using an average read depth of 1000X. Multiple logistic regression was used to model *PIK3CA* mutation status on the basis of deep phenotyping, including clinical, radiological, histopathological, haematological and outcome variables. Median age at presentation was 1.2 years, median follow up 7.0 years. 42% of patients had overgrowth of body and/or limb, 7% had macrocephaly, 57% had a visible superficial component to the vascular malformation, 11% had epidermal naevi, 49% had abnormal clotting parameters, 95% had positive pS6 (MTOR) immunohistochemistry, 59% required surgery and 41% sclerotherapy for pain or overgrowth. *PIK3CA* mutations were detected in affected tissue of 17/43 (41%), and absent from blood. The mutant allele frequency was 1.2-28%. No phenotypic variables were statistically predictive of *PIK3CA* mosaicism. Our findings confirm the broad phenotypic spectrum reported with post-zygotic mutations in *PIK3CA*. Thus far we have not identified any phenotypic predictors of genotype, although a larger group is required to stratify by affected protein domain. Currently we will continue to offer sensitive genotyping to this whole phenotypic group to best direct the choice of novel targeted therapeutic strategies.

1342W

Rare variant analysis of human neurodegenerative disorders guided by high-throughput yeast screens for alpha-synuclein toxicity. *D. Zielinski[1], V. Khurana[2], J. Sandoe[2], S. Lindquist[2], Y. Erlich[1].* 1) New York Genome Center, New York, NY; 2) Whitehead Institute for Biomedical Research, Cambridge, MA.

Rare risk variants pose a major challenge for association studies of complex human conditions. Hypothesis free searches require an overwhelming number of individuals to reach the statistical power needed to associate these variants. We took a radically different approach for the search of rare variants in neurodegenerative diseases. In our previous studies, we established an alpha-synuclein yeast model that recapitulates the molecular pathologies of Parkinson's disease. Using this model, we conducted high-throughput screens to search for modifiers of alpha-synuclein toxicity in yeast. We conducted an exploratory sequencing study targeting the human homologs of these genes as well as known Parkinson's risk genes in more than 500 familial and sporadic synucleinopathy patients. Despite the small number of samples, the relatively small set of 430 genes allowed us to concentrate the statistical power on more likely targets. As a control, we compared the amount of rare variants to the ExAC resource and a cohort of aged individuals with no CNS pathology. This process generated nearly 300 variants that passed an FDR of 5%. After the exploratory stage, we rescreened the same exact variants in a yeast model and patient-derive neurons to obtain an orthogonal validation of their pathogenicity. We envision that our cross-species approaches will continue to yield important insights for complex diseases with evolutionarily conserved biological mechanisms, and perhaps help fulfill therapeutic promises in the post-genomics era.

1343T

Analysis of whole-genome-sequence in an extended pedigree with autism: Identification of a candidate gene, with replication. *N. Chapman[1], R. Bernier[2], D. Bilder[3], T. Darlington[3], J. Morgan[3], H. Coon[3], E. Wijsman[3].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 3) Psychiatry, University of Utah, Salt Lake City, UT.

Autism is a complex disorder, with high heterogeneity that inhibits replication. Under heterogeneity, large multiplex families, while rare, are a powerful tool for discovery of susceptibility genes. We report the discovery in a large family of an excellent candidate gene for autism, with replication in a 2nd family. We carried out gene-discovery in Family A (47 individuals, 5 generations). Dense SNP data and whole exome sequence were available on 44 and 5 individuals, respectively. Linkage analysis identified a 2.1Mb region of interest on chr 22. Of 11 affected individuals, 8 share a haplotype in this region. There are also 11 unaffected individuals carrying the risk haplotype (RH), including 3 female siblings of affecteds. The only rare exome variant on the RH is synonymous and unlikely to be a risk allele (Hum Hered 74:153, 2012). We also obtained whole genome sequence (WGS) on two trios in Family A where the RH was transmitted to an affected child. We filtered for variants that were transmitted from non-founder parents to the affected child, with a frequency of ≤1% in the 1000 Genomes Project Europeans. 3 of the resulting 16 variants are rare (0.1% frequency) and are close to or intronic to the same gene, which is an excellent candidate for autism. For replication, we examined WGS on 88 individuals (57 affecteds and 31 unaffected siblings) in 13 large families with autism. Dense SNP panels were available for these and many family members. After quality filtering, we found that 10 of the 16 rare variants on the RH were represented in the replication sample. 5 of these could be rejected as autism risk variants, either because they appeared solely in unaffected individuals, or because they did not segregate with autism. The remaining 5 replicated variants include the same 3 implicated in Family A. In the replication set all 5 appear together in an affected male and his unaffected sister, who has epilepsy and symptoms of ADHD. Analysis of SNP data in these individuals and 6 additional nuclear family members showed that the 5 variants are transmitted from the grandfather to the father and then to all 4 children, as an intact haplotype. The grandfather has an inflated score on the Social Responsiveness Scale, a measure of autistic social impairment. A rare haplotype with variants in a novel autism candidate gene segregates with autism in 2 large multiplex families. Details of the variants and the gene will be shared at the meeting.

1344F

Induced pluripotent stem cell models of autism spectrum disorder associated with early brain overgrowth. *C. Fu[1], H. Belinson[3], S.P. Smieszek[2], L.A.D. Bury[1], M.X. Zhang[1], J.I. Haines[2], A. Wynshaw Boris[1].* 1) Genetics and Genomics, Case western reserve university, Cleveland, OH; 2) Epidemiology and Biostatistics, Case western reserve university, Cleveland, OH; 3) Department of Pediatrics, School of Medicine, University of California San Francisco.

Autism Spectrum Disorder (ASD) is a highly heterogeneous neurodevelopmental disorder, characterized by deficits in social interaction, verbal communication and repetitive behavior. The molecular mechanisms for ASD are poorly understood. We have been investigating the hypothesis that a subset of patients (~25-30%) with ASD display early brain overgrowth. Our lab has recently produced two relevant models that we believe model important aspects of early brain overgrowth in ASD. First, we produced mouse models deficient for *Dvl1* and *Dvl3* (*Dvl1-/-3+/-* mutant) that display adult social behavior abnormalities associated with transient embryonic brain enlargement during deep layer cortical formation (Belinson et al. 2016). Second, we generated human induced pluripotent stem cell (iPSC) models by reprogramming fibroblasts obtained from 8 ASD individuals with early brain overgrowth and 5 non-ASD controls with normal brain size. Neuronal progenitor cells (NPCs) derived from ASD iPSC lines displayed enhanced proliferation compared to control NPCs (Marchetto, Belinson et al. 2016). In both models, these aberrant ASD phenotypes were caused by down-regulation of Beta-catenin activity and its direct target BRN2. Additionally, treatment of GSK3β inhibitor to both mice embryo and human NPCs can rescue the cellular and behavioral phenotype (Belinson et al. 2016, Marchetto, Belinson et al. 2016). To investigate the primary and secondary targets of Beta-catenin and BRN2, we are performing ChIPSeq using antibodies for Beta-catenin and BRN2 for each of human NPC lines, as well as NPCs from wild-type and *Dvl1-/-3+/-* mutant mice strain, in both GSK3β inhibitor treated and untreated condition. Significant RNASeq and ChIPSeq results are used as input for Gene Ontology and pathway analysis. Initial results of RNASeq and ChIPSeq for BRN2 for 1 ASD NPC line and 1 control line in both GSK3β inhibitor treated and untreated condition demonstrated that 20 RNAs displayed differential expression levels and differential BRN2 binding (and corrected by GSK3β inhibitor treatment). These 20 genes were participants in 13 significant pathways, including the Wnt signaling pathway, cell cycle and synaptic vesicle cycle. We are currently performing similar analyses on the other ASD and control lines. Taken together, the current results suggest our approach is promising for uncovering molecular mechanisms for brain overgrowth and ASD.

1345W

Rare inherited and *de novo* CNVs reveal complex contributions to ASD. *V. Leppa[1,2], S.N. Kravitz[1], C.L. Martin[3], J. Andrieux[4], C. Le Caignec[5,6], D. Martin-Coignard[7], C. DyBuncio[1], S.J. Sanders[8], J.K. Lowe[1,2], R.M. Cantor[9], D.H. Geschwind[1,2,9].* 1) Center for Autism Research and Treatment, and Program in Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, CA, USA; 2) Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, CA, USA; 3) Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA, USA; 4) Institut de Génétique Médical, Hospital Jeanne de Flandre, CHRU de Lille, Lille, France; 5) CHU Nantes, Service de Génétique Médicale, Nantes, France; 6) INSERM, UMR-957, Laboratoire de Physiopathologie de la Résorption Osseuse et Thérapie des Tumeurs Osseuses Primitives, Nantes, France; 7) Service de Génétique, CH Le Mans, France; 8) Department of Psychiatry, University of California San Francisco, CA, USA; 9) Department of Human Genetics, University of California Los Angeles, CA, USA.

Rare mutations contribute significantly to autism spectrum disorder (ASD) and the importance of large, rare *de novo* copy number variants (CNV) and *de novo* protein disrupting single nucleotide variants (SNV) is well established in simplex families in the Simons Simplex Collection (SSC) with no close relatives with ASD. The SSC families are likely to be enriched for sporadic ASD, potentially overemphasizing the importance of *de novo* variation. We analyzed 1464 families from the Autism Genetic Resource Exchange (AGRE) to assess the impact of inherited and *de novo* CNVs on ASD risk in a primarily multiplex cohort. The overall burden of rare, large CNVs is higher in ASD cases compared to their unaffected siblings (OR 1.7, p 0.01) in AGRE. However, the rate of large, rare *de novo* CNVs is not significantly associated with ASD in AGRE and is significantly lower in AGRE cases compared to cases in the SSC (p 0.01). We examined CNVs overlapping previously established ASD associated genes and risk-CNVs in AGRE. A significant overall CNV burden was discovered in cases compared to unaffected siblings, although no individual gene was associated with ASD. We also assessed 49 CNVs that were observed in previously identified ASD risk loci, comprising 24 inherited, 19 *de novo*, and 6 CNVs of unknown inheritance. In 21 (71%) of 30 families harboring a previously identified ASD CNV, such as 15q11.2-13.1 duplication, at least one additional affected sibling did not share the risk CNV with their other affected sibling(s), although 5/21 families had inherited risk CNVs. We conclude that the genetic risk in multiplex families is complex, multifactorial and clearly differs from SSC families, warranting the collection of larger multiplex ASD cohorts. Finally, we find evidence for a novel *de novo* CNV locus at 2q24 that includes the genes *NR4A2* and *GPD2*, of which, former is a highly constrained gene and the latter harbors several likely deleterious variants in the general population. There are no *NR4A2* overlapping CNVs in any non-ASD cohorts, including the DGV, and siblings form AGRE and SSC (p = 0.02). We found in total seven individuals with 2q24.1 deletion: 2 in AGRE, and 4 in DECIPHER and 1 in ClinGen clinical databases. Patients with a deletion in 2q24 share distinct clinical phenotypes, such as ASD, developmental delay and delay in language development, which implies a novel form of syndromic ASD. .

1346T

Genetic determinants of treatment outcome and social impairments in autism spectrum disorder. *K. Tammimies[1,2], V. Nicolaou[1], S. Stamouli[1], N. Choque Olsson[2], S. Berggren[1,2], S. Bölte[1,2].* 1) Center of Neurodevelopmental Disorders (KIND), Pediatric Neuropsychiatry Unit, Department of Women's and Children's Health, Karolinska Institutet , Stockholm , Sweden; 2) Child and Adolescent Psychiatry Stockholm, Center for Psychiatry Research, Stockholm County Council, Stockholm, Sweden.

Deficits in social communication and interaction are core symptoms in autism spectrum disorder (ASD). ASD is currently incurable, but behavioral and other interventions, such as social skills group training (SSGT), can be used to prevent worse outcomes and comorbidity. However, an individual's response to these treatments can be highly variable and the specific factors influencing the outcome are mostly unknown. Limited information exists as to whether rare genetic variants, such as large copy number variants (CNVs), can be used to predict changes in social impairments in ASD and treatment outcome. This study aims to investigate the genetic predictors associated with treatment outcome of SSGT and analyze genetic profiles connected with distinct longitudinal changes in social impairments. For this, a cohort of 302 ASD individuals with common comorbid conditions such as attention-deficit hyperactivity disorder and IQ >70 from the randomized control trial (RCT) for SSGT KONTAKT (clinicaltrials.gov Identifier: NCT01854346) was used. The primary outcome measure used was the change in the parent rated social responsiveness scale (SRS) scores from pre- to post-treatment and after three month follow-up. We are now conducting high resolution genomic analyses in this cohort including CNV profiling using the Affymetrix Cytoscan HD® platform and whole exome/genome sequencing. First, we have assessed whether the presence of large rare CNVs (>500 kb) affected the change in SRS during the training period using linear regression model. Our preliminary results revealed that large rare CNVs were associated with worse outcome following SSGT (β= -14.6, p=0.02) explaining 10% of the variation in the treatment outcome among the participants. The majority of these CNVs were well-known risk loci for neurodevelopmental disorders. For example, we identified a participant with 7q11.23 duplication syndrome, in which social anxiety is known characteristic, whose social impairments increased significantly during the training. We are now expanding these analyses to include the standard care group, selecting individuals for sequencing based on their SRS profile and performing pathway enrichment analyses from the genomic findings. The results from this study can pave the way to understand how genetic information can be used to understand the social impairment profiles in ASD and better tailor the limited resources in treatments for affected individuals.

1347F

***PLXNA3* variant rs5945430 is associated with severe clinical course in male multiple sclerosis patients from Kuwait.** *M. Hatem[1], M. Qureshi[1], R. Alroughani[2,3], F. Al-Mulla[4], R. Al-temaimi[5].* 1) 7th year Medical student , Faculty of Medicine , Kuwait University; 2) Division of Neurology, Department of Medicine, Amiri Hospital, Kuwait; 3) Neurology Clinic, Department of Medicine, Dasman Diabetes Institute, Kuwait; 4) Molecular Pathology Unit, Department of Pathology, Faculty of Medicine, Kuwait University; 5) Human Genetics Unit, Department of Pathology, Faculty of Medicine, Kuwait University.

ABSTRACT: Statement of Purpose: Multiple Sclerosis (MS) is an inflammatory autoimmune demyelinating disease affecting the central nervous system. MS prevalence and clinical course are significantly affected by gender. MS clinical course has been shown to be more aggressive in MS males. Our objective was to identify the genetic factors influencing this phenomenon. **Methods Used:** Eight male and 19 female Kuwaiti MS patients were recruited at Dasman Diabetes institute's neurology clinic. Patient selection was based on having a high MS severity score (MSSS >5). DNA was extracted from blood using Qiagen DNA mini kit. Exome sequencing was performed on an Illumina HiSeq2000 platform using TruSeq v4 chemistry with target coverage of 50X. Patient exome sequences were analyzed using Tute Genomic s application. An additional 62 males and 120 female MS patients were genotyped to confirm exome analysis finding using Taqman genotyping assay for *PLXNA3* variant rs5945430. **Summary of Results:** Female MS exomes were compared to male MS exomes and a genetic variant was discovered in *PLXNA3* gene on chromosome X resulting in a non-synonymous amino acid change (c.2589C>G, p.D863E, rs5945430) (p=3.6300e-5). Male MS patients hemizygous for allele G had significantly higher MS severity scores (MSSS) and increased expanded disability status scale (EDSS) when compared to allele C male MS patients (p=0.005, p=0.02; respectively). Allele G associated with an aggressive MS clinical course in male MS patients (p=0.01; Relative risk (RR) 2.5; 95%CI: 1.07-5.8). No significant association was found for homozygous GG females and MS clinical variables. Our results suggest that a genetic variant in *PLXNA3* (p.D863E), a protein involved in axon path-finding; is associated with MS aggressive clinical course specifically in male patients. We propose *PLXNA3* as a novel sex influenced predictive marker for male MS patients' clinical course.   Keywords from list: Autoimmune disorder (8), Brain/nervous system (14), Exome sequencing (59), Neurodegeneration (127),.

1348W

The pharmacogenomics of serious adverse drug reactions associated with disease-modifying therapies for multiple sclerosis: Work in progress. *K. Kowalec[1], E. Kingwell[1], R. Carruthers[1], A. Traboulsee[1], G.E.B. Wright[1], B.I. Drögemöller[1], F. Aminkeng[1], S. Bernatsky[2], C.J.D. Ross[1], B.C. Carleton[1], H. Tremlett[1].* 1) University of British Columbia, Vancouver, British Columbia, Canada; 2) McGill University, Montréal, Québec, Canada.

   Background: Multiple sclerosis (MS) is a chronic degenerative disease of the central nervous system. Ten different disease-modifying therapies (DMT) have become available over the past two decades for the treatment of relapsing MS. Some recently licensed include fingolimod, dimethyl fumarate (DMF), teriflunomide and alemtuzumab and are associated with an increased risk of numerous adverse drug reactions (ADRs). DMF and fingolimod have been associated with progressive multifocal leukoencephalopathy (a serious brain infection); individuals exposed to teriflunomide experience an increased risk of hepatotoxicity, with alemtuzumab increasing the risk for autoimmune thyroiditis and immune-mediated thrombocytopenia. These serious ADRs are currently unpredictable using current clinical and demographic information. **Statement of purpose**: To identify genomic variation associated with the occurrence of serious ADRs in MS, with the ultimate goal of preventing these reactions by either avoiding therapies in specific patients with a high risk of drug-induced harm or by critically informing complex treatment discussions between health care providers and patients. **Study design:** This case-control study will recruit from specialist MS clinics across Canada. Cases are defined as MS patients who have experienced a serious ADR as follows: DMF-associated lymphopenia; DMF-, teriflunomide- or fingolimod-associated hepatotoxicity; or alemtuzumab-associated autoimmune thyroiditis or immune thrombocytopenia. Controls are MS patients who have been exposed to the relevant DMT for an equivalent period of time (e.g. +/- 1 year), with no biochemical/clinical evidence of the ADR and will be matched to cases based on follow-up time. Patient's saliva-derived DNA will be genotyped on a genome–wide panel of genes (1.7M SNPs) involved in drug biotransformation and those associated with DMT ADRs (if any) to identify variants associated with ADR susceptibility. Data will be analyzed using logistic regression to compare genotype frequencies between cases and controls with adjustment for relevant clinical factors and statistical significance set at the Bonferroni-corrected alpha level. **Status**: Recruitment will begin in June 2016, with a total of 200 patients to be enrolled for the first ADR (DMF-lymphopenia). **Conclusion**: Few published studies have attempted to use pharmacogenomic methods to predict and prevent ADRs to MS DMTs. We will present this innovative study as a work in progress.

1349T

Analysis of sporadic inclusion body myositis using high density arrays for immune related genes confirms the human leukocyte antigen as the most associated region and suggests a role for chemokine receptor 5. *J.A. Lamb[1], S. Rothwell[2], R.G. Cooper[3], I.E. Lundberg[4], P.K. Gregersen[5], M.K. Herbert[6,7], G.J.M. Pruijn[6], J. Bowes[8], M.G. Hanna[9], P.M. Machado[9,10], J.B. Lilleker[8,11], M.F. Seldin[12], H. Platt[1], A.T. Lee[5], W.E.R. Ollier[1], H. Chinoy[8], MYOGEN Inclusion Body Myositis Investigators.* 1) Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, United Kingdom; 2) Centre for Musculoskeletal Research, University of Manchester, Manchester, UK; 3) MRC/ARUK Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK; 4) Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Karolinska Institutet, Stockholm, SE; 5) Center for Genomics & Human Genetics, Feinstein Institute for Medical Research, Manhasset, NY, US; 6) Department of Biomolecular Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, NL; 7) Department of Haematology/Oncology, Beth Israel Deaconess Medical Center (BIDMC), Boston, MA, US; 8) National Institute of Health Research Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, University of Manchester, UK; 9) MRC Centre for Neuromuscular Diseases, University College London, London, UK; 10) Centre for Rheumatology Research, University College London, London, UK; 11) Greater Manchester Neurosciences Unit, Salford Royal NHS Foundation Trust, Salford, UK; 12) Department of Biochemistry and Molecular Medicine, University of California Davis, Davis, CA, US.

   Objective: Sporadic inclusion body myositis (IBM) is characterised by a combination of inflammatory and degenerative changes affecting muscle. While the primary cause of IBM is unknown, genetic factors may influence disease susceptibility. We have conducted the largest genetic association study to date in IBM, investigating immune-related genes using the Immunochip.Methods: 252 Caucasian IBM cases fulfilling Griggs/ENMC/MRC inclusion criteria were recruited from 11 countries through the Myositis Genetics Consortium (MYOGEN), and compared with 1,008 ethnically matched controls. Classical HLA alleles and amino acids were imputed using SNP2HLA. Linear regression analyses were used to investigate the relationships between HLA alleles and age of onset of IBM.Results: The results confirmed the HLA as the most strongly associated region (p=3.58x10[-33]) in IBM. Three non-HLA regions reached suggestive significance, including chr3(p21.21), an established risk locus for autoimmune disease, where a frameshift mutation in CCR5 is thought to be the causal variant. HLA imputation identified three independent associations with HLA-DRB1 alleles, although the strongest association was with amino acid positions 26 and 11 of the HLA-DRB1 molecule. No association with anti-cN-1A positive status was found independent of HLA-DRB1*03:01. There was no association of HLA genotypes with age of onset of IBM.Interpretation: This is the largest, most comprehensive genetic association study to date in IBM. The data confirm that the HLA is the most strongly associated region and identifies novel amino acid associations that may explain the risk in this locus. A novel suggestive association within the chr3(p21.21) locus suggests a role for CCR5.

1350F

Causal networks of Alzheimer's disease on a large RNA-seq dataset. *N.D. Beckmann[1], M. Wang[1], R. Chang[1], V. Haroutunian[2], B. Zhang[1], E.E. Schadt[1].* 1) Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY.

   Alzheimer's Disease (AD) is the leading cause of dementia worldwide, with more than 28 million cases globally, and this number projected to double in the next 20 years. AD is characterized by progressive loss of cognitive function, affecting ones ability to think, reason and remember. The hallmarks of AD are senile plaques in the brain, composed of accumulated Amyloid Beta and neu-rofibrillary tangles (hyperphosphorylated Tau protein). While several genetic components of AD have already been identified and many biological pathways associated to the disease, the lack of causal genes, pathways/networks, and mechanisms of the disease continue to elude the scientific community.   Here, we study a large group of post-mortem samples (n>300) with the complete range of neuropathologies (from controls to definite AD) across 4 brain regions from the AMP-AD consortium. Using whole exome sequencing (WES) for each sample and RNA-seq from each brain region, we defined a set of differen-tially expressed and differentially connected (between cases and controls) genes that were enriched in coexpression subnetworks (modules) that were associated with AD. These modules were then assigned, when possible, to specific brain-associated cell types. We then constructed probabilistic causal network models of the genes in these modules for cases and controls across the 4 brain regions. These networks organize the genetic, expression, and clinical data of this large dataset in ways that enable us to identify important key drivers of these regulatory networks. We demonstrate that the key driver genes we identify and their surrounding sub-networks provide novel insights into the mechanisms underlying AD risk and pathogenesis.

1351W

Whole genome sequencing in an Amish Parkinson Disease pedigree reveals large allele frequency differences between noncoding variants in individuals with PD versus those successfully aged. *K. Nuytemans[1,2], L. Maldonado[1,2], K. John-Williams[1,2], A. Mehta[1,2], G.W. Beecham[1,2], E.R. Martin[1,2], J.L. Haines[3], W.K. Scott[1,2], J.M. Vance[1,2].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Miami Udall Center, University of Miami, Miami, FL; 3) Case Western Reserve, Cleveland, OH.

   Previous genetic analyses in a large multi-case family with Parkinson Disease (PD) in the Amish community in Ohio and Indiana has supported a complex nature of the disease in this isolated group. However, as a population isolate, it should have an enrichment of rare variants (RV) that contribute to PD. We focused our analysis on a 0.5 MB region on chromosome 5 that we have shown previously to have the strongest evidence of sharing between PD cases in this family. This region also overlaps with a linkage peak previously reported in non-Amish families [Scott et al 2001]. After whole exome sequenc-ing in 32 cases failed to identify obvious rare coding variant(s) responsible for PD in this region, we sequenced the whole genome of 28 affected individuals from 10 branches of the family, as well as 21 successfully aged (SA; lacking significant cognitive or physical impariment, >80yo) Amish individuals from the same community. A total of 1880 noncoding variants were identified in the chr5 region, of which 181 variants were shared between cases in branches identified to contribute to the chr5 signal in the genetic analyses. Twenty-two of those variants were identified as having potential functional effect by several algorithms (RegulomeDB, CADDv1.3, GWAVA, VistaEnhancer, transcription factor binding sites (TFBS)) though only two of the variants showed evidence in more than one (both CADD and TFBS). Alternatively, 50 variants have an allele frequency of <10% in the SA individuals, 18 of which are also rare in the 1000 genomes dataset (<5%). All but one of the 50 show an allele frequency difference of at least 8-10% between affected and SA individuals. One noncoding variant (rs338950) has a CADD>15, overlaps a potential transcription factor binding site and has a 10% allele frequency difference in affected (~18%) versus SA individuals (~7%). This variant is intergenic, located ~23kb from the closest gene *FGF1*, which has been shown to prevent the death of dopamine neurons in animal models of PD. Contrary to our whole exome analysis, our initial genome data analysis has identified many potential noncoding risk variants, lying in the chromosome 5 region in our large Amish PD family. Current experiments are underway to assess the potential impact of these variants in PD.

1352T

Potential involvement of *DYSF* in Amyotrophic Lateral Sclerosis. *A. Orr-Urtreger[1,2], O. Goldstein[1], O. Nayshool[1], B. Nefussy[3], Z. Gan-Or[4], G.A. Rouleau[4], M. Gana-Weisz[1], V.E. Drory[3,2].* 1) Genetic Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Tel Aviv, Israel; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Neuromuscular Service, Department of Neurology, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel; 4) Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, H3A 0G4 Canada; Department of Human Genetics, McGill University, Montreal, Quebec, H3A 0G4 Canada; Montreal Neurological Institute, McGill University, Montreal, Quebec, H3A 0G4.

   Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease affecting 2-3 per 100,000 people per year worldwide. The disease features a wide phenotypic and genetic heterogeneity. The age at disease onset (AAO), the site of disease onset (bulbar, limb), the rate of progression, the disease manifestations and other characteristics, vary significantly. A large number of recently identified causative mutations in more than 20 genes accounts for the genetic heterogeneity of ALS. The various protein functions of these genes suggest multiple pathological pathways. Among the major pathways are proteosomal and autophagy pathways, RNA processing, and axonal dysfunction. We aimed to detect rare genetic variants underlying familial and sporadic ALS that are involved in both recessive and dominant modes of inheritance. Whole-exome-sequencing of Ashkenazi Jewish patients affected with ALS was carried out. A trio-analysis of a patient, which was diagnosed at the age of 21 years old, and her parents was performed. A homozygous *DYSF* variant was detected and subsequently further genotyped on 325 Ashkenazi Jewish patients with ALS and 970 ethnically matched controls .A mutation in *DYSF* was identified in a homozygous state in the ALS patient that changes a conserved amino acid at position 128 from glycine to glutamic acid (c.383G>A; p.G128E). Two additional ALS patients homozygous to the mutation were identified, accounting for about 1% of all Ashkenazi Jewish ALS patients. No homozygotes were observed in controls. Under recessive mode of inheritance these results show a significant association to ALS (p=0.009, χ2 test). The frequency of the mutated allele in the Ashkenazi controls is 0.056. The observed *DYSF* variant is associated with ALS in a homozygous state. For our knowledge, this is the first time that this myopathic gene is demonstrated to have a potential role in ALS. Its relatively high frequency in controls may suggest a model of age-dependent incomplete penetrance for homozygotes, or a risk factor. Further studies in large groups of patients and controls from other ethnicities are necessary to evaluate the role of *DYSF* in ALS pathogenesis.

1353F

Association of *SLC24A4* coding variants with the risk of posterior cortical atrophy. *M.M. Carrasquillo[1], I. Barber[2], S.M. Reeder[1], J. Siuda[1], M.L. Kachadoorian[1], J.D. Burgess[1], A. N'songo[1], S.J. Lincoln[1], T. Nguyen[1], K.G. Malphrus[1], L. Ma[1], G.D. Bisceglio[1], M. Allen[1], M.E. Murray[1], Q. ul Ain Khan[3], S.G. Younkin[1], D.W. Dickson[1], B.F. Boeve[4], N.R. Graff-Radford[3], R.C. Petersen[4], K. Morgan[2], N. Ertekin-Taner[1,3].* 1) Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA; 2) Human Genetics Group, University of Nottingham, Nottingham, UK NG7 2UH; 3) Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA; 4) Department of Neurology, Mayo Clinic Minnesota, Rochester, MN 55905, USA.

   Posterior cortical atrophy (PCA) is usually associated with Alzheimer's disease (AD) pathology. It is characterized by progressive visual impairment, relatively preserved memory and language in the early stages of the disease, and AD pathology predominantly in posterior brain regions. A recent PCA genome-wide association study (GWAS) identified study-wide significant associations at three loci not previously associated with AD, and only nominal significance was found at three of the twenty AD GWAS loci replicated in the International Genomics of Alzheimer's Project (IGAP) meta-analysis. In the present study we set out to evaluate if coding variants at the 20 AD-risk loci present in the NeuroX array, and not evaluated in the PCA GWAS study, may modify PCA risk. Gene-based analysis in our NeuroX cohort consisting of 124 PCA cases and 171 non-demented elderly controls identified significant association at the *SLC24A4* locus (p=2.5x10[-4]). Testing single variants for association with PCA risk (124 PCA cases vs. 3,557 non-demented elderly controls) demonstrated significant association with a rare variant not previously implicated in AD, *SCLC24A4*-rs62623445 (p.S362N), and conditional analyses suggest that the effect of this coding variant is independent from that detected with the AD GWAS intronic SNP *SCLC24A4*-rs10498633. In this first systematic evaluation of coding variants at AD-risk loci in a PCA cohort, we report a rare coding variant in *SCLC24A4* that associates with PCA risk, as well as significant association with *SCLC24A4* at the gene level. Thus, the results of this study provide support for further studies of the influence of *SLC24A4* in PCA-risk.

1354W

Apolipoprotein E and Alzheimer disease risk in a Puerto Rican Alzheimer disease data set. *B. Feliciano[1], K.L. Hamilton-Nelson[2], L.D. Adams[2], P.L. Whitehead[2], N.K. Hofmann[2], H.N. Cukier[2,3], A. Chinea[4], H. Acosta[5], N.I. Feliciano[6], J.L. Haines[7], J.L. McCauley[2,8], R.M. Carney[2], J.M. Vance[2,3,8], M. Cuccaro[2,8], G.W. Beecham[2,8], M.A. Pericak-Vance[2,3,8].* 1) PRADI, SJB School of Medicine, Caguas, PR; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Neurology, University of Miami Miller School of Medicine, Miami, FL; 4) San Juan MS Center Guaynabo, PR; 5) Centro Caribeño para el Estudio de la Memoria y Cognición, San Juan, PR; 6) Psychiatry, PRADI and Veterans Administration Hospital de San Juan, PR; 7) Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH; 8) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

   BACKGROUND: There are over 5 million individuals of Puerto Rican (PR) origin, making up over 1.5% of the US population and representing the 2nd largest Hispanic/Latino population in the continental US. The number of Alzheimer disease (AD) cases on the island of PR is estimated at 65,000. Genetic studies have been limited to a few rare families with early onset AD (PSEN mutations). The PR population is highly admixed, with average ancestry values of 64% European, 21% African, and 15% Native American, but individual ancestries are highly variable. The strong association between Apolipoprotein E 4 (*APOE4)* and AD has been confirmed world-wide but there is variability in the effect size in Hispanics and data suggesting that both African Americans and Hispanics have an increase incidence of AD regardless of *APOE* genotype. The purpose of the present study was to directly assess the effect of *APOE4* in a PR AD data set.   **ANALYSIS AND RESULTS:** To date we have ascertained 74 unrelated individuals into the PR AD Initiative (PRADI) study including 44 AD cases and 30 cognitively intact controls. Seventy five percent of the cases are female, and 63% of controls are female. The mean age at exam for the cases is 76.8 ± 9.0, while the mean age at exam for control individuals is 69.9 ± 7.7. The APOE e4 allele frequency in the cases is 22.7%, while the *APOE4* frequency in the controls is 21.7%. Logistic regression was performed to determine whether there was an association between affection status and *APOE* genotype, adjusting for age at exam and gender. The association between *APOE4* dose and affection status was not significant (OR=1.10, beta=0.091, se=0.45, p-value=0.84). However, given the current sample size an Odds Ratio of 2.55 would be needed to achieve 80% power. We do note that the three *APOE4* homozygotes all had AD.   **CONCLUSIONS:** Our data support the hypothesis that *APOE4* does not strongly contribute to risk in PR AD. These findings are interesting given a previous report suggesting that among PR non-demented nonagenarians the *APOE4* allele is associated with better cognition (Carrion-Baralt et al., 2009). GWAS genotyping is underway to examine local ancestry at the *APOE* locus as well as global ancestry in the PRADI data set as well as additional sampling and expansion of the PRADI data set.

1355T

A genome-wide study of structural variants in patients with epilepsy. *S.L. Girard[1,2,3], J. Monlong[2], C. Meloche[3], M. Cadieux-Dion[3], M. Gravel[3], S. Dobrzeniecka[3], D. Spiegelman[4], A. Dionne-Laporte[4], C. Boelman[5], G.A. Rouleau[4], J. Michaud[6], R.G. Lafreniere[3], B. Minassian[5], G. Bourque[2], P. Cossette[4].* 1) Université du Québec à Chicoutimi, Chicoutimi, Canada; 2) Genome Center, McGill University, Montréal, Canada; 3) Centre de Recherche du CHUM, Montréal, Canada; 4) Montreal Neurological Institute, Montréal, Canada; 5) Sick Kids Hospital, Toronto, Canada; 6) CHU Ste-Justine, Montréal, Canada.

   Background: Epilepsy is a complex neurological disorder that is characterized by abnormal excessive or synchronous neuronal activity in the brain. Although a small portion of the disorder can be explained by fully penetrant genetic mutations, a large fraction of the genetic heritability remains unknown to this day. The Canadian Epilepsy Network (CENet) is a Genome Canada funded initiative to bring epilepsy into the personalized medicine era. Methodology: A specific aim of this project was to evaluate the implication of structural variants (SVs) in the genetic mechanism of the disease. To do so, we performed whole genome sequencing on 200 patients suffering from various types of epilepsy, including generalized, partial and photosensitive epilepsy. We used PopSV, a detection algorithm developed by some members of our group to identify small and large structural variants in the genome of epilepsy patients. In order to evaluate the frequency of the identified SVs, we had access to more than 300 whole genome sequences from healthy individuals. Results: Out of the most promising SVs identified, a significant portion encompassed genes already linked to epilepsy thus showing the great importance of SVs in the genetic mechanism of the disease. We also found that on average, epilepsy patients carry more exon-disrupting SVs than healthy individuals. More specifically, we were able to identify structural variants in many genes previously known to harbour single nucleotide variants linked to epilepsy (*LGI1*, *CHD2*, *DEPDC5*) as well as larger chromosomal events (15q11.2 and 15q11.3). We also identified several new genes harbouring recurrent deleterious structural variants. These genes represent excellent new candidates to better explain the genetic etiology of epilepsy.

1356F

Benign familial neonatal seizures to early onset epileptic encephalopathy: What's mosaicism got to do with it? *A. Guerin, C. Kiss, A. Rusnak, J. MacKenzie.* Divison of Medical Genetics, Department of Pediatrics, Kingston General Hospital, Kingston, Ontario, Canada.

   Benign familial neonatal seizures is an autosomal dominant epilepsy syndrome that self resolves with favorable development. Up to 70% of families have a mutation in either *KCNQ2* or *KCNQ3*. However, it is recognized that the same mutations in *KCNQ2* also cause early onset epileptic encephalopathy, a severe, refractory form of epilepsy associated with significant developmental disability.   We present a patient who was born at term with refectory epilepsy developing at four days of life. Pregnancy was complicated by intrauterine growth restriction and neonatal 'hiccups'. EEG confirmed epileptiform activity and the patient had several medication trials. Family history was significant for benign neonatal seizures in the patient's mother, which self-resolved by the age of one year with no sequelae. On examination, the child was non-dysmorphic, with axial and appendicular hypotonia. Investigations including infection screening, microarray and biochemical analysis were negative. MRI brain revealed white matter abnormalities, consistent with micro-hemorrhages. An epilepsy gene panel revealed an undescribed heterozygous mutation in *KCNQ2* (c.790 T>A). The mutation is predicted to be pathogenic as it affects a conserved region essential for protein structure and function. This mutation with the clinical picture, fits a diagnosis of early onset epileptic encephalopathy. Parental testing revealed the mutation was maternally inherited, but present at much lower levels, suggesting somatic mosaicism. This may explain the history of benign neonatal seizures, in the context of a variable phenotype. The patient is currently two years old with well controlled epilepsy and developmental disability.   Our severe presentation of the patient with a more attenuated presentation in her mother, in the context of parental mosaicism, poses additional questions about the genotype–phenotype relationship of *KCNQ2, particularly with this mutation.* It could be proposed that mutations in *KCNQ2* have either a threshold effect, where a yet undetermined proportion of cells with the mutation must be present in order to have a phenotype, or that another gene is modifying the phenotype in our patient. Additionally, timing of expression may play a role, as benign epilepsy was present in our patient's mother and perhaps neonatal neurons are more susceptible and require a lower level of mutation to manifest as epilepsy. .

1357W

Genome-wide association scans of the brain's white matter microstructure reveals a pleiotropic relation with cardiovascular disease – Results from the ENIGMA- diffusion tensor imaging consortium. *N. Jahanshad, The Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) Consortium and DTI Working Group.* Neurology, University of Southern California, Marina del Rey, CA.

   Diffusion-weighted MRI (dMRI) reveals the brain's microstructure and connectivity using advanced metrics beyond those obtainable with standard anatomical MRI; dMRI offers insight into the properties of neural pathways, including axonal packing, myelination, and fiber connection patterns in the brain. Large-scale population studies of diffusion tensor imaging (DTI) are now underway, analyzing brain scans from many clinical and research dMRI protocols, to understand factors that affect white matter microstructure. One of the most common measures from DTI - fractional anisotropy (FA) - is a biomarker for numerous neuropsychiatric disorders, and is highly heritable in populations around the world. We present results from GWAS meta-analyses of FA derived from brain scans of over 10,000 individuals. Diffusion brain scans were analyzed using harmonized processing and quality-control protocols (http://enigma.ini.usc.edu/protocols/dti-protocols/). Several brain regions of interest are extracted from the protocols, of which 21 were found to be reliable and heritable across sites and protocols. These ROIs (including whole-brain average FA) were used for the GWAS; all ROI measures are bilaterally averaged across the left and right hemispheres. The genotypic data from each cohort were quality-controlled and imputed to the 1000 Genomes reference panel. All cohorts computed ancestry multi-dimensional ancestry scores with harmonized protocols. GWASs were conducted on all 21 ROIs. Corrections for sex, age (linear and quadratic effects) and their interactions were included. An additional covariate was included in the analysis for sites with patients and healthy controls. Patients included individuals with ADHD, Alzheimer's disease, bipolar disorder, Parkinson's disease and schizophrenia. The discovery sample included CEU individuals scanned at over 30 sites around the world. Results of the GWAS meta-analyses reveal an association between the FA in the anterior corona radiata and a locus in the *BCHE* gene, previously found to have strong influences over cardiovascular and metabolic risk factors. We report these results and present findings for many historically "candidate" SNPs for psychiatric disorders previously reported to affect FA in single-site studies. We relate findings to known biological and developmental processes affecting brain connectivity, and report the genetic covariation with other brain traits, as well as neuropsychiatric and cognitive phenotypes.

1358T

Mapping the effects of common variants within the human cortex: Results from the ENIGMA consortium GWAS meta-analyses of cortical thickness and surface area. *S.E. Medland, The Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium.* Psychiatric Genetics, Queensland Institute of Medical Research, Brisbane, Australia.

At the individual level, there is substantial variation in cortical thickness and surface area which has been implicated in a wide range of psychiatric and neurological traits. Thinning of the cortical starts in childhood and continues across the lifespan, whereas cortical surface area expands until about 12 years, remains relatively stable through middle age and then decreases. While cortical thickness and surface area are both strongly heritable, the two processes are genetically independent and there is little known about the loci influencing these morphological characteristics. Here we present results from GWAS meta-analyses of the thickness and surface area of cortical regions of interest derived from magnetic resonance imaging (MRI) scans from over 15,000 individuals.   Across cohorts, structural T1-weighted MRI brain scans were analysed locally using harmonized analysis and quality-control protocols (http://enigma.ini.usc.edu/protocols). All cortical parcellations were performed with the freely available and validated segmentation software. Cortical thickness and surface area were calculated for 68 (34 left and 34 right) cortical gray matter regions were visually inspected and statistically evaluated for outliers. The genotypic data from each cohort were quality-control and imputed to the 1000 genome reference panel and all cohorts computed ancestry multi-dimensional scores following harmonized protocols.   To improve measurement accuracy, the cortical measurements were averaged across the hemispheres resulting in the average thickness and surface area of 34 regions. We also analyzed two summary measures, average cortical thickness across regions and total surface area, resulting in a total of 70 cortical phenotypes. Corrections for sex, age (linear and quadratic effects) and their interactions were included in all GWAS analyses. For the regional measures corrections for the corresponding summary measures were also included in the GWAS to account for the omnibus effects of brain size.   The GWAS meta analyses presented provide an overview of the loci influencing univariate region level thickness and surface area. We also present the results of the multivariate analyses which focus on identifying patterns of genetic variation across regions. We relate these to known biological and developmental processes, examine the conservation of these loci and characterize the genetic covariation with psychiatric, neurological and cognitive phenotypes.

1359F

Genetics of trigeminal neuralgia: Insight from 15 families and whole exome sequencing. *S.J. Mosca[1], Z.H. Kiss[2], A.M. Innes[1,3].* 1) Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada; 2) Department of Clinical Neuroscience, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada; 3) Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada.

Trigeminal neuralgia (TN) (OMIM 190400) is a rare neuropathic pain condition typically characterized by intense unilateral paroxysmal attacks in the trigeminal nerve distribution, with an incidence of about 5 per 100,000 person-years. TN, which has also been called the "suicide disease", has been described as the worst pain known to man. The pathophysiology is thought to be related to neurovascular compression and demyelination of the nerve root entry zone. Treatment of TN includes anticonvulsant medications and surgeries that either relieve compression of the trigeminal nerve or deafferent the nerve. An enhanced understanding of the underlying disease mechanisms could inform future therapeutic considerations.   Evidence for an underlying genetic component includes a small number of reported pedigrees, with a familial form of TN accounting for about 1-2% of cases. The previous reports of familial TN have demonstrated that segregation of the disease is consistent with autosomal dominant inheritance with reduced penetrance. In addition, there is a co-occurrence of TN in certain Mendelian disorders, specifically Charcot-Marie-Tooth disease (CMT), a disorder of impaired myelination. Further evidence of dysmyelination in TN comes from the known association with multiple sclerosis (MS), as about 1-2% of MS patients develop TN.   Identifying genes predisposing to TN using conventional genetic techniques has been limited due to age-related reduced penetrance and typically small numbers of living affected individuals, but cohort analysis using whole-exome sequencing (WES) provides a new avenue of inquiry. Here we report our experience with a cohort of patients with familial TN, ascertained from a single academic neurosurgical practice. We have reviewed 15 affected individuals with a positive family history of TN. In three of these families at least some affected individuals had MS, but in most families MS was not present. Families had two to four affected individuals and were generally all consistent with autosomal dominant inheritance with reduced (and age-dependent) penetrance. 10 families presented for further genetic assessment, and interrogation of 3 common CMT-causing genes (PMP22, MPZ, GJB1) was negative in all families. We proceeded to WES on a proportion of cases (8 individuals from 5 families). Preliminary approaches and results from the data analysis will be discussed.

1360W

Searching for the missing genetic component of amyotrophic lateral sclerosis using a candidate gene approach. *E.P. McCann[1], K.L. Williams[1], J.A. Fifita[1], I.S. Tarr[1], Q. Zhao[2], D.B. Rowe[1], N. Wray[2], G.A. Nicoloson[1,3], I.P. Blair[1].* 1) Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia; 2) Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, Unviersity of Queensland; 3) ANZAC Research Institute, University of Sydney, Concord Hospital, Sydney, New South Wales, Australia.

   Amyotrophic lateral sclerosis (ALS, also known as Lou Gehrig's disease) is a debilitating and ultimately fatal neurodegenerative disease. Roughly 10% of cases are hereditary (familial ALS), while the remainder occur sporadically. The only known causes of ALS are gene mutations, with over twenty genes implicated to date. Gene mutations have driven the discovery of much functional knowledge underlying ALS. However, known mutations account for just two thirds of familial ALS and 11% of sporadic ALS. Thus, there is a pressing need to uncover the remaining proportion of the genetic causes underlying ALS. Classic gene discovery efforts in large ALS families have identified most known currently known ALS genes. The remaining ALS families are generally small and often exhibit incomplete penetrance of disease, complicating ALS gene discovery. Thus, we propose a candidate gene approach for identifying the remaining ALS mutations. Custom bioinformatics pipelines using the R interface have been developed to parse patient exome sequence data for shared variants in families or candidate genes, followed by Fisher's Exact testing for associations with ALS. Identified variants are then prioritised using a combination of the following: absence from control databases, gene function, predicted functional consequence, amino acid conservation, natural variation and variation tolerance of the gene, and presence in replication cohorts. We first investigated candidates proposed based on genome wide association data (n=4 genes), transgeneic mouse models (n=1 gene) or functional evidence (n=1 gene). This provided strong evidence implicating one of these genes in ALS. Further, shared variant analysis using exome data from eight ALS families comprised of multiple affected individuals identified between 26 and 112 candidate ALS mutations in each family. By applying this prioritisation approach to the first family, we have reduced the number of potential causal variants from 26 to 8 high priority and 2 medium priority variants. Identification of causal ALS mutations using this approach will enhance our understanding of ALS pathogenesis and inspire follow up studies investigating mutant effects on protein structure and function. Such discoveries are also imperative for providing patients and families the opportunity for diagnostic testing, which may be vital for future personalised therapeutics, as well as predictive testing or pre-implantation embryonic screening.

1361T

Phenotypic variability of *de novo* **mutations in** *ASXL3. S. Walker[1], R.K.C. Yuen[1], B. Thiruvahindrapuram[1], T. Woo[1], M. Woodbury-Smith[2], R. Dad[3], P. Szatmari[4], B.A. Minassian[3], S.W. Scherer[1].* 1) The Centre for Applied Genomics, Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Psychiatry & Behavioural Neurosciences, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; 3) Division of Neurology, Genetics and Genome Biology, The Hospital for Sick Children, University of Toronto, Ontario, Canada; 4) Centre for Addiction and Mental Health, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

   Bainbridge Ropers syndrome is an autosomal dominant disorder characterized by growth retardation, psychomotor delay, feeding difficulties and dysmorphology. The syndrome presents with an overlapping but distinct phenotype to Bohring-Opitz syndrome, which is caused by *de novo* mutations in *ASXL1*. Brainbridge-Ropers syndrome has been attributed to mutations in *ASXL3*, with all mutations resulting in premature truncation of the protein and arising *de novo* in affected cases. *ASXL3* (Additional sex combs-like 3) is a human homologue of the Drosophila gene asx and is involved in chromatin modification and transcription regulation. To date, 10 cases with *de novo* loss of function mutations in *ASXL3* have been published with a further 10 cases being described in DECIPHER. In 2014, *de novo* truncating mutations were also identified from exome sequencing in Autism Spectrum Disorder (ASD) cohorts. All but one of the mutations found are in exon 11 or 12, but are located throughout the protein since these two exons encode 83% of the coding sequence. Nearly all cases have developmental delay and language difficulties but there is variable presentation of additional symptoms, including motor-delay and hypotonia, ASD, intellectual disability and microcephaly. Here we present two new cases with *de novo* loss of function variants in *ASXL3* with differing phenotypes. Case one is a 15 year-old boy diagnosed with intellectual disability (ID). His phenotype is also characterized by delayed language development, microcephaly, short stature and craniofacial dysmorphology. He also presents with myoclonus and psychomotor slowing in the absence of any history of hypotonia. Through exome sequencing, a *de novo* frameshifting substitution was detected in exon 11 which results in premature termination at amino acid 445 (p.Ser439Argfs*7). Case two is a 9 year-old boy diagnosed with ASD, ID and comorbid ADHD. He also has a history of language delay and gross motor vulnerabilities. Through whole genome sequencing of the child and his parents, a *de novo* single base substitution was detected in exon 12 that results in the introduction of a premature stop codon at amino acid 1636 (p.Gln1636*). These two, along with other published cases, indicate that there is significant phenotypic variability arising from *de novo* truncating mutations in *ASXL3* and further assessment, including searching for other genetic contributors in these individuals, is necessary to correlate genotypes with phenotypes observed.

1362F

Identification of genes predisposing to syringomyelia associated with Chiari malformation I. *F. Ancot[1,2], P. Lemay[1,2], S.P. Knowler[3], G. Rouleau[4], C. Rusbridge[3], Z. Kibar[1,2].* 1) Sainte Justine Hospital Research Center, Montréal, Canada; 2) University of Montréal, Canada; 3) School of Veterinary Medicine and University of Surrey, Surrey, UK; 4) Montréal Neurological Institute and McGill University.

Chiari I malformation (CMI) is a common congenital abnormality of the craniovertebral junction (CVJ) affecting 1 in 1280. It is characterized by a descent of the cerebellar tonsils into the foramen magnum, often in association with a fluid filled cavity within the spinal cord called syringomyelia (SM). The exact pathogenesis of SM associated with CMI remains largely unknown. SM has emerged as one of the most common spinal cord diseases affecting toy breed dogs and as in humans it is mostly associated with Chiari-like malformation (CM). CM is ubiquitous in the Cavalier King Charles spaniels (CKCS) however not all dogs develop SM. To better understand the pathogenesis of SM/CM and attempt to predict its risk, there is a necessity to identify its genetic cause(s) in order to reduce or eliminate this condition in the numerous affected breeds. SM/CM has a complex inheritance. To investigate the genetics of SM/CM, we adopted a quantitative trait locus (QTL) approach where a total of 27 cranial measurements were taken in 65 CKCS dogs (including 33 SM unaffected) and were subjected to a whole-genome association study using the Illumina CanineHD beadchip array. Significant associations were identified in 2 candidate genomic regions of 0.8Mb on CFA26 and of 0.9Mb on CFA22.We have re-sequenced these 2 candidate regions using the SeqCap developer EZ and relaxed baits designed by Roche Nimblegen custom group and refined a significantly associated haplotype of 0,4 Mb on CFA26. Further bioinformatics and functional validation of associated variants residing in this haplotype is currently underway. Our study will help better understand the molecular pathogenic mechanisms underlying SM/CM and provides an entry point for identification of genes predisposing to SM/CMI in humans.

1363W

Polygenic overlap between amyotrophic lateral sclerosis and schizophrenia. *R.L. McLaughlin[1,2], D. Schijven[3,4], W. van Rheenen[3], K.R. van Eijk[3], M. O'Brien[1], R. Kahn[4], R.A. Ophoff[4,5,6], A. Goris[7,8], D.G. Bradley[1], A. Al-Chalabi[9], L.H. van den Berg[3], J.J. Luykx[3,4,10], O. Hardiman[2], J.H. Veldink[3], Project MinE GWAS Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium.* 1) Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland; 2) Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Dublin, Ireland; 3) Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 6) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA; 7) KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium; 8) VIB, Vesalius Research Center, Laboratory of Neurobiology, B-3000 Leuven, Belgium; 9) Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; 10) Department of Psychiatry, Hospital Network Antwerp (ZNA) Stuivenberg and Sint Erasmus, Antwerp, Belgium.

Amyotrophic lateral sclerosis (ALS) is a complex neurodegenerative disease characterized by rapid-onset loss of upper and lower motor neurones, resulting in progressive paralysis and death from respiratory failure, typically within 3-5 years of symptom onset. Schizophrenia is a neuropsychiatric disease with positive symptoms, negative symptoms and impairment over a broad range of cognitive abilities. We have recently shown that schizophrenia occurs with a higher frequency than expected in the pedigrees of ALS patients, suggesting an aetiological relationship between both diseases. Substantial SNP-based heritabilities and variance explained by polygenic risk scores (PRS) indicate polygenic components to both diseases. Using linkage disequilibrium score regression with summary statistics for GWAS of ALS and schizophrenia comprising over 100,000 unique individuals, we estimated the genetic correlation between ALS and schizophrenia derived from these polygenic components to be 14.3% (95% confidence interval 7.05-21.6; p = 1×10[-4]). We supported this estimate with analysis of schizophrenia PRS in a target of ALS cases without shared and related individuals in the schizophrenia dataset. Up to 0.12% of the variance in ALS was explained by PRS calculated on schizophrenia-associated alleles (p = 8.4×10[-7]). We leveraged the apparent pleiotropic relationship between ALS and schizophrenia to identify five potential novel ALS-associated genomic loci at conditional false discovery rate < 0.01. Diagnostic misclassification in the schizophrenia cohort did not contribute significantly to our observations (BUHMBOX p = 0.94) and we estimated that 4.86% (2.47-7.13%) of ALS cases would need to be misdiagnosed as schizophrenia in order to observe our genetic correlation estimate under a true genetic correlation of 0%. Our results indicate that the lifetime risk for comorbid ALS and schizophrenia increases from 1 in 40,000 to 1 in 34,336, which would require an incident cohort of 16,488 ALS patients to observe epidemiologically. Our findings suggest shared underlying biology between ALS and schizophrenia which will direct novel approaches in research and therapeutic development.

1364T

Use of NINDS Repository DNA samples in genomic studies available through dbGaP: A source of enhanced-annotation of biospecimens publicly available for neurological disorders research. *C.A. Pérez[1], S. Heil[1], J. Santana[1], A. Amberson[1], A. Green[1], R. Zhang[2].* 1) NINDS Repository, Coriell Institute for Medical Research, Camden, NJ; 2) NIH-NINDS, Bethesda, MD.

The burden of neurological disorders is a serious health concern and presents a massive challenge to healthcare systems globally. Studies aimed at identifying such pathogenic mechanisms will accelerate discovery of disease-specific genetic markers, the development of tools for monitoring disease onset and progression, and the implementation of therapies. The National Institute of Neurological Disorders and Stroke (NINDS) Repository is a public resource established to provide a centralized and open resource of DNA and lymphoblastoid cell lines with their corresponding de-identified clinical and phenotypic data. NINDS Repository samples cover a diverse population of neurological patients and neurologically normal controls. Since its establishment, biomaterials from more than 44,000 individuals with cerebrovascular diseases, Parkinsonism, motor neuron diseases, epilepsy, Tourette syndrome, Dystonia, and neurologically-normal controls have been submitted to the repository and more than 34,500 unique samples have been successfully banked in the NINDS Repository and are available via its online catalog (http://catalog.coriell.org/1/NINDS). To aid high-throughput gene discovery, the NINDS Repository offers disease- and neurologically-normal control panels of genomic DNA samples in 96-well plate formats as well as custom plate design services. As a result, since its inception back in 2002, more than 6,500 samples from the NINDS Repository have been used in GWAS, SNP, CNV, or sequencing studies and the results of such studies has been deposited in dbGaP, a NIH/NLM-sponsored repository for restricted-access data from studies investigating the interaction of genotype and phenotype. dbGaP, the Database of Genotypes and Phenotypes is available online (http://www.ncbi.nlm.nih.gov/gap). The NINDS Repository regularly monitors new dbGaP studies to identify those that include genomic information about NINDS Repository samples. In addition, a faceted-search tool in the Coriell Biorepositories online catalog allows users to filter samples of interest to identify those with additional genomic characterization data in dbGaP. This cross-referencing of the NINDS repository and dbGaP databases allows NINDS Repository users to access additional characterization data for thousands of biospecimens and complements the available samples and their associated clinical data.

1365F

Identification and characterization of multiple sclerosis associated loci in the continental Italian population. *M. Sorosina[1], N. Barizzone[2,3], F. Clarelli[1], S. Anand[4], S. Lupoli[5], F. Esposito[1,6], E. Mangano[4], R. Bordoni[4], V. Martinelli[6], G. Comi[6], M. Leone[2], D. Cusi[5], N.A. Patsopoulos[7,8,9,10], P. De Jager[7,8,9,10], G. De Bellis[4], S. D'Alfonso[2,3], F. Martinelli Boneschi[1,6], PROGEMUS, PROGRESSO.* 1) San Raffaele Scientific Institute, Milan, Italy; 2) Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy; 3) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 4) National Research Council of Italy, Institute for Biomedical Technologies, segrate, Milano, Italy,; 5) Department of Health Sciences, University of Milan, Italy; 6) Department of Neurology and Neurorehabilitation, San Raffaele Scientific Institute, Milan, Italy; 7) Ann Romney Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital, Boston, MA; 8) Harvard Medical School, Boston, MA; 9) Program for Medical and Population Genetics, Broad Institute, Cambridge, MA; 10) Program in Translational Neuropsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, MA.

Multiple sclerosis (MS) is a complex disorder characterized by myelin damage and neurodegeneration. Until now 103 susceptibility genetic loci have been identified though genome-wide association studies (GWAS), however they involved samples from several different countries that could have a different genetic and environmental predisposition. The aim of the present project was to identify and characterise the most associated regions in the Italian continental population following a multi-step approach. We started from two cohorts of Italian continental MS patients and healthy controls (HC): ITA$_{GWAS}$ (750 MS, 1291 HC) with available imputed genome-wide genotypes and ITA$_{iChip}$ (977 MS, 1000 HC) with genotyping data in 184 immune regions. To boost the statistical power of the study, the top associated signals of association (p<0.005) in the two Italian cohorts were meta-analysed with an imputation-based meta-analysed cohort of 20,512 cases and 19,145 HC of European ancestry from US. A total of 16 loci of association were identified (p<5x10^{-7}), and then tested for genetic association in a third Italian cohort (903 MS, 884 HC). This replication phase confirmed the association of two loci on chromosome 17 and chromosome 3, already known to be associated with the disease. To better characterize them, we did a target sequencing study in an Italian cohort of 600 MS and 408 HC using a pooling approach and comparing high-risk MS patients and low-risk HC according to their genetic burden. In the locus on chromosome 17, sequencing data revealed the presence of 203 SNPs associated with disease (best signal: p= 1.11x10^{-23}). We tested, based on their availability, these 203 SNPs for the existence of cis-eQTL association using Braineac, Gtex Portal and SNPExpress databases. We found an association of several SNPs with the expression of *EFCAB13* in brain tissues, while in immunological tissues we observed that a different gene, *TBKBP1*, is also regulated, with an opposite direction compared to brain. A total of 12 signals were identified by sequencing analysis in the intergenic locus on chromosome 3 with no eQTL associations for any of those SNPs. Concluding, we confirmed that two loci on chromosome 17 and 3 are highly associated also in the Italian population, and we better characterised the origin of these signals using next-generation sequencing approach. Moreover, we observed tissue-dependent regulatory functions, opening questions on the role of *EFCAB13* and *TBKBP1* in MS.

1366W

Genetic contributions to motor learning and spindle density through sleep: Recall by genotype investigation of ZNF804A. *M. Taylor[1], U. Bartsch[1], C. Hellmich[1], C. Durant[2], L. Corbin[3], M. Jones[1], N. Timpson[3].* 1) School of Physiology and Pharmacology, Bristol University, Bristol, Bristol, United Kingdom; 2) Clinical Research and Imaging Centre, Bristol University, Bristol, Bristol, United Kingdom; 3) MRC Integrative Epidemiology Unit at Bristol University, Bristol, Bristol, United Kingdom.

rs1344706 in *ZNF804A* was among the first genetic variants associated with schizophrenia and since then a number of studies have examined the influence of rs1344706 on cognition, brain structure and physiology. We used a Recall by Genotype (RbG) design implemented in the Avon Longitudinal Study of Parents and Children to investigate impacts of rs1344706 on sleep architecture, neurophysiology and sleep-dependent memory consolidation in healthy young adult homozygotes CC (n=12) and AA (n=14). We used a combination of Actigraphy and sleep diaries (collected over two weeks) to assert sleep wake cycles. Participants were monitored using controlled sleep laboratory polysomnography (PSG) on two nights. A motor sequence task (MST) was used to probe sleep-dependent memory consolidation. We ran a detection algorithm to quantify non- rapid eye movement neurophysiology characteristics including sleep spindles, slow waves and delta rhythms. Measures were inverse rank transformed and linear regression was used to test for between group differences. Actigraphy revealed similar rest-activity cycles between groups. As expected, we found no evidence for differences in sleep architecture (such as sleep latency, total sleep time, or time spent in sleep stages) between groups. There was evidence that the AA group performed worse (mean rank change in number of correct sequences = -0.86 95% CI -1.64, 0.09, p=0.03) and had a slower reaction time (mean rank change of button press latency = 0.79 95% CI 0.002, 1.58, p=0.05) on the MST. We found weak evidence for reduced spindle density (mean rank change of number of spindles per minute = -0.62 95% CI -1.39, 0.14, p=0.10) and delta frequency (mean rank change in time between delta events = -0.70, 95% CI -1.45, 0.05, p=0.07) in AA carriers. The overall pattern of association signal results suggested a reduction in coordinated neural activity which was complemented by impaired performance in the MST testing. Although current evidence is weak, findings here (along with the overall consistency across multiple PSG phenotypes) suggest that rs1344706 is likely to be correlated with finely measured neurological phenotypes during sleep. This work not only highlights the utility of RbG as an efficient design for following up genetic association signals, but has allowed direct comparison of animal model results to human experiments and development of functional interpretation of established gene effects.

1367T

Copy number variation in mesial temporal lobe epilepsy. *T.K Araujo[1], F.R. Torres[1], R. Secolin[1], A. Donatti[1], M.K.M Alvim[2], C.S. Rocha[1], M.E Morita[2], C. Yasuda[2], B.S. Carvalho[3], F. Cendes[2], I.L. Cendes[1].* 1) Department of Medical Genetics , School of Medical Sciences, University of Campinas – UNICAMP , Campinas, SP, São Paulo, Brazil; 2) Department of Neurology; School of Medical SciencesUniversity of Campinas – UNICAMP , Campinas, SP, São Paulo, Brazil; 3) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing; University of Campinas – UNICAMP and Brazilian Institute of Neuroscience and Neurotechnology, Campinas, SP, BRAZIL.

Epilepsy is a common brain disorder characterized by recurring seizures. Mesial temporal lobe epilepsy (MTLE) is the most common form epilepsy in adults and it is frequently associated with medically refractory seizures. Most patients with MTLE do not present positive family history of the disease, which is believed to have a complex mode of inheritance. Copy number variations (CNV) have been identified in a wide-range of neurodevelopmental disorders, many of which are also associates with epilepsy. The objective of the present study was to investigate the distribution of CNVs in patients with MTLE. To date, we have studied a total 396 individuals, 198 patients with MTLE and 198 control subjects. CNVs were identified by array genomic hybridization (aGH) using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). We considered only CNVs that were ≥ 100kb and that spanns 25 probes for deletion and 50 probes for duplications. We identified CNVs in several genomic hotspots, which were previously associated with different epilepsy syndromes including 1q21.1, 15q11.2, 15q13.3, 15q11–q13, 16p11.2, and 16p13.11. Furthermore, we identify many overlapping CNVs that are present in different patients (until 90 patients). The size of CNVs ranges between 245kb to 2840kb.We identified microdeletions and microduplications at the following chromosomal regions 1q21.2, 1p21.1, 2p11.2, 2q35, 3p12.2, 3q12.2, 3q26.1, 4q13.2, 5p15.33, 5p15.2, 5q13.2, 6p25.3, 7q35, 8p11.22, 8p23.1, 9p11.2, 9p12 p11.2, 9q13-q21.11, 10q11.21, 10q26.3, 11p11.12, 11q14.1, 12q14.2, 13q21.1, 14q11.2, 15q11.2, 15q11.1-q11.2, 15q13.2, 16p11.2, 17p13.3, 17q12, 17q2131, 19q13.31, 22q11.21 and 22q11.22. Although we are reporting here only preliminary findings and additional analysis are under way, we found evidence that CNVs may be an important genetic susceptibility factor involved in the pathogenesis of epilepsy syndromes with complex inheritance such as MTLE. Supported by: FAPESP, SP, BRAZIL.

1368F

Genetic overlap between epilepsy and schizophrenia, using GWAS summary statistics and raw genotypes. *S.S. Cherny[1,2,3], H. Gui[2], M. Li[1,2], P.C. Sham[1,2,3], L. Baum[1,2], P. Kwan[4,5].* 1) Psychiatry, Univ Hong Kong, Pokfulam, Hong Kong; 2) Centre for Genomic Sciences, University of Hong Kong; 3) State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong; 4) Department of Medicine & Therapeutics, Chinese University of Hong Kong; 5) Department of Neurology, Royal Melbourne Hospital, Australia.

   Epilepsy and schizophrenia both common disorders, and neurological and psychiatric illnesses, respectively. A patient suffering from epilepsy may have depression or other mental illnesses, and some forms of seizures may mimic the symptoms of psychosis. Better understanding of epilepsy may therefore provide insights on schizophrenia, and vice versa. However, genome-wide association studies (GWAS) of schizophrenia and epilepsy have identified no common genetic loci. In this study, summary statistics were obtained from both the Psychiatric Genomics Consortia (PGC) and International League Against Epilepsy (ILAE) Consortium on Complex Epilepsies meta-analyses of schizophrenia and epilepy. In addition, raw genotype data were employed from our own GWAS of schizophrenia and epilepsy on Chinese populations. To investigate the genetic relationship between the two disorders, we calculated the overall genetic correlation by LD-score regression analyses and polygenic risk scores (PRS) projected on each other. In addition, we searched for shared susceptibility genetic loci by cross-phenotype analyses on three different levels (SNPs, genes, and gene-sets). A non-significant very low genetic correlation (r=-0.01) was found between these two disorders. PRS on different *p*-values for one phenotype cannot predict the other one. Genetic sharing was only found for one SNP located at 2p16.1 locus. These negative findings may partially be due to low power of the epilepsy meta-analysis GWAS, which only explained 7% of phenotype variance. Nevertheless, this also suggests the comorbidity between these two brain diseases is less likely to be caused by genetic sharing than other confounding environmental risk factors.

1369W

Effect of CYP2C19 polymorphisms on stiripentol administration in cases of Dravet syndrome. *T. Kouga[1], H. Osaka[1], H. Shimbo[2], K. Takano[2], M. Iai[2], S. Yamashita[2], T. Yamagata[1], A. Ishii[3,4], S. Hirose[3,4], K. Yamakawa[5].* 1) Department of Pediatrics, Jichi Medical University, Tochigi, Japan; 2) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Pediatrics, School of Medicine, Fukuoka University, Fukuoka, Japan; 4) Central Research Institute for the Pathomechanisms of Epilepsy, Fukuoka University, Fukuoka, Japan; 5) Laboratory for Neurogenetics, RIKEN Brain Science Institute, Saitama, Japan.

   [;**Objective**];Dravet syndrome (DS) is an intractable epilepsy characterized by the recurring febrile and afebrile seizures in early childhood. Stiripentol (STP), an inhibitor of cytochrome P450, increases plasma concentrations of concomitant antiepileptic drugs and showed efficacy in DS. An increase in blood concentration of clobazam (CLB) and its major metabolite, norclobazam (N-CLB), had been attributed to the antiepileptic efficacy of STP. STP increases the N-CLB through attenuation of its hydroxylation by CYP2C19. The objective is to investigate STP administration in cases of DS by comparing CYP2C19 allelic polymorphisms with clinical effects of STP. [;**Method**];CYP2C19 polymorphism were examined in eleven cases of DS.CYP2C19 has two main allelic polymorphisms, CYP2C19*2 and CYP2C19*3, in addition to the wild genotype, CYP2C19*1. The *2 variant derives from a single-nucleotide polymorphism (G to A), and creates an exon aberrant splice site resulting in aberrant splicing and out of frame transcription. The *3 variant derives from a single-nucleotide polymorphism (G to A), resulting in a premature stop codon. Each polymorphism were examined with clinical characteristics of the cases; age at the study period, body weight, mean dose and plasma concentration of valproate acid (VPA)/CLB off and on STP, mean plasma concentration of N-CLB off and on STP, degree of seizure reduction, and adverse effects of STP. [;**Result**];There were 3 cases of DS with wild type (*1/*1), 6 with intermediate type (*1/*2, *1/*3), and 2 with poor type of CYP2C19 polymorphisms (*2/*2, *2/*3). The N-CLB concentration/CLB dose ratio and N-CLB/CLB concentration ratio off STP were significantly higher in poor metabolizers. Three of 8 cases showed no effectiveness of STP regardless of the N-CLB concentration increase, and 1 of 3 cases showed effectiveness of STP regardless of N-CLB concentration decrease. In total, six of 11 cases with DS had >50% reduction in seizure frequency without significant differences over the CYP2C19 polymorphisms. [;**Conclusion**];This study demonstrated an effect of CYP2C19 polymorphisms on STP administration in cases of DS. There were cases of seizure reduction regardless of N-CLB concentration decrease on STP, which indicates a significant anti-convulsant action of STP in itself. N-CLB concentration decrease on STP was observed in 2 cases with *3 allelic polymorphisms. STP may partially restore the enzyme activity of CYP2C19*3.

1370T

NTRK2 is associated with multidrug-resistant epilepsy in European and African American populations. *L. Vazquez[1], B. Almoguera[1], E. Robbins[2], D. Dlugos[2], J. Connolly[1], F. Mentch[1], R. Pellegrino[1], F. Mafra[1], J. Garifallou[1], L. Tian[1], S. Manzi[3], P. Sleiman[1], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pharmacy, Boston Children's Hospital, Boston, MA.

   Introduction: Up to 40% of patients with epilepsy do not respond to currently available antiepileptic treatment options. Although genetic factors are likely to play a role in drug-resistant epilepsy (DRE), no strong gene candidate has emerged yet. Our aim was to investigate the genetic factors involved in multidrug-resistant epilepsy.   **Methods:** Identification of subjects with DRE and responders was done using an EMR based algorithm, which mines EMR data for pertinent diagnostic and medication information. The algorithm was implemented at the Children's Hospital of Philadelphia, revealing positive predictive values of 84.6% for responders and 83.3% for DRE. It was validated in a research cohort at Boston Children's Hospital with a positive predictive value of 100%. We used the PGRN sequencing platform of 82 genes to sequence a cohort of 97 DRE, 337 drug-responders and 866 non-epilepsy controls. The effect of genetic variants on DRE was assessed in European Americans (EA) and African Americans (AA) separately comparing DRE versus drug-responders and DRE versus non-epilepsy controls. Single variant analysis, gene-based analysis of rare functional variants and meta-analysis were performed with raremetalworker and raremetal, correcting for cryptic relatedness and population structure.   **Results:** No significant results were found in the DRE versus drug-responders analysis. However, in the analysis of DRE versus non-epilepsy controls, *NTRK2* was significantly associated with DRE in both the single variant as well as the gene-based analyses. A rare missense variant, rs76950094 (minor allele frequency –MAF- of 0.0016) was associated with DRE with a p-value=4.94×10^{-8} (p-value in EA=1.19×10^{-4}; p-value in AA=1.10×10^{-4}, with a cutoff for significance of 1.31×10^{-5} after correcting for multiple testing). In the gene-based analysis, seven functional variants with MAF<0.05 were associated with DRE (p-value=5.65×10^{-5}; with a cutoff for significance set as 1.87×10^{-4} after correcting for 276 genes).   **Conclusion:** *NTRK2* is involved in the development and maturation of the central and the peripheral nervous systems and has a role in neuronal excitability. The identification of variants in the *NTRK2* gene in association with DRE may contribute to a better understanding and improved therapeutic management of these patients in the future.

1371F

Whole exome sequencing implicates VLDLR as a risk factor for cerebral palsy. *J. Connolly[1], J. Li[1], F. Mentch[1], Y. Guo[1], D. Abrams[1], M. Oskui[2], M. Shevell[2], S. Lafleur[2], L. Tian[1,3], H. Hakonarson[1,3,4].* 1) Ctr Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Department of Neurology, McGill University, Montreal, Canada; 3) Division of Human Genetics Dept. of Pediatrics, Perelman School of Medicine, Philadelphia, PA; 4) Dept. of Pulmonary Medicine, CHOP, Philadelphia, PA.

   Several recent studies have suggested a prominent role for genomic factors in susceptibility to cerebral palsy (CP). To address the genetic underpinnings of CP, we conducted a genome-wide association study (GWAS) on our cohort of 697 cases and >2,000 controls. While no single SNP reached genome-wide significance, several leading signals were notable as functional candidates, and were used to inform subsequent analyses. Next, we selected 65 patients for whole exome sequencing (WES). Exclusion criteria for cases were prematurity, complications of pregnancy, complications of labor, and co-morbidity with known Mendelian disorders. Results were compared against 688 internal controls, selected on the basis of stringent exclusion criteria including pre-/post-/peri-natal complications, and a wide range of neurological deficits. After quality control filtering, we analyzed subjects of European ancestry and African American subgroups separately. We conducted gene-wise tests on the CP phenotype, using 7 different algorithms implemented in software package EP-ACTS (Efficient and Parallelizable Association Container Toolbox) to compare cases *versus* controls. Then we performed meta-analysis of the two groups. The gene *VLDLR*, encoding for Very Low Density Lipoprotein Receptor, came out with P-value < 5×10^{-5} in 6 out of 7 tests in subgroup of European ancestry and in 2 tests in African Americans. In all instances, meta-analyses yielded a more significant P-value for each type of test. Mutations in this gene have been associated with cerebellar hypoplasia and disequilibrium syndrome, suggesting functional plausibility and valuable insight into possible etiology of CP.

1372W

Early-onset epilepsy: Diagnostic yield and treatment impact of whole-exome sequencing. *I. Guella[1], M.B. McKenzie[1], S.E. Buerki[2], E. Toyota[2], S. Adam[2], M. Van Allen[3], G. Sinclair[4], G. Horvath[4], C.D. Van Karnebeek[4], P. Eydoux[5], D. Evans[1], A. Datta[2], C. Boelman[2], L. Huh[2], A. Micholaus[2], B. Bjornson[2], T. Nelson[6], M. Connolly[2], M.K. Demos[2], M.J. Farrer[1].* 1) Centre for Applied Neurogenetics, Djavad Mowafaghian Centre, Dept of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Division of Neurology, Department of Pediatrics, University of British Columbia and BC Children's Hospital, Vancouver, BC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Division of Biochemical Diseases, Department of Pediatrics, University of British Columbia and BC Children's Hospital, Vancouver, BC, Canada; 5) Department of Pathology and Laboratory Medicine, University of British Columbia and British Columbia Children's Hospital, Vancouver, BC, Canada; 6) Department of Pathology and Laboratory Medicine, Children's and Women's Health Center of British Columbia, Vancouver, BC, Canada.

Epilepsy is one of the most common neurological disorders and affected children are at increased risk of having developmental delay, autism and a psychiatric illness. Modern genomic technologies, such as targeted next generation sequencing and whole exome sequencing (WES), enable the identification of pathogenic variants in 10 – 78 % of selected patients with unexplained epilepsy. The clinical impact is significant and includes an earlier diagnosis of disorders with specific treatment implications. Exome trio intersections, of parents and children, may identify novel causes of disease. Here, we report results on 148 patients with early onset epilepsy (≤ 5 years), evaluating diagnostic yield, time to diagnosis and treatment impact. WES was performed in 109 retrospective (epilepsy > 6 months) and 39 prospective (epilepsy < 6 months) patients. Detailed clinical data was abstracted in a REDCap database. WES was performed using Ion Proton™ System within 2 weeks of receiving samples. Average coverage was >80x across all samples. Initial reporting was restricted to ~550 genes previously implicated in epilepsy. Putative causative mutations were validated by Sanger sequencing and by parental testing. In patients with no immediate genetic diagnosis we also performed proband-parent trio-based WES analyses. We established a genetic diagnosis in 23% of patients which included the identification of variants in 25 known epilepsy genes. A possible diagnosis was identified in an additional 11% of patients for whom supporting evidence is pending. Moreover, genetic diagnosis had potential treatment implications in 18 probands (12%) which resulted in anti-seizure medication changes and/or treatment with specific therapies. Time to genetic diagnosis from enrollment with genetic counselling averaged 46 days for the first 50 patients studied. The analysis of 22 family trios is currently ongoing and may further increase diagnostic yield. Clinical utility of WES in this epilepsy patient cohort was supported by a timely potential diagnoses in 34% of patients, with a higher yield in the prospective cohort compared to the retrospective cohort (44% vs 32%). The protocol accelerates triage in pediatric epilepsy, for patients and their families, and will likely improve patient care and potentially outcomes.

1373T

Whole exome sequencing approach for the identification of pain-related mutations in an Italian cohort of familial and early-onset patients affected by idiopathic peripheral neuropathy. *F. Martinelli Boneschi[1,7], A. Zauli[1], S. Santoro[1], M. Marchi[2], M. Sorosina[1], L. Grevendonk[1], D. Cazzato[2], R. Lombardi[2], C.G. Faber[3], M.M. Gerrits[4], R. Almomani[4], J.G.J. Hoeijmakers[3], I.S.J. Merkies[3], H. Fadavi[5], R. Malik[5], D. Ziegler[6], G. Boenhof[6], G. Comi[1,7], S.D. Dib-Hajj[8], S.G. Waxman[8], G. Lauria[2].* 1) Laboratory of Human Genetics of Neurological Disorders, San Raffaele Scientific Institute, Milan, IT; 2) Neuroalgology Unit, Neurological Institute Carlo Besta, Milan, IT; 3) Department of Neurology, Maastricht University Medical Center, Maastricht, NL; 4) Clinical Genetics, Maastricht University Medical Center, Maastricht, NL; 5) Centre for Endocrinology and Diabetes, University of Manchester and Central Manchester NHS Foundation Trust, Manchester, UK; 6) German Diabetes Center, Medical Faculty Heinrich Heine University, Düsseldorf, DE; 7) Department of Neurology, San Raffaele Scientific Institute, Milan, IT; 8) Department of Neurology, Yale University School of Medicine, New Haven, CT, USA.

Neuropathic pain is a frequent feature of peripheral neuropathy causing a significant impact on patients' quality of life and health care costs. The Consortium of 9 partners from 6 countries coordinated by IRCCS Neurological Institute Carlo Besta is engaged in the project for Probing the Role of Sodium Channels in Painful Neuropathies (the PROPANE STUDY). We propose to use a whole exome sequencing approach in a cohort of familial and early-onset painful neuropathic patients to identify rare variants causative of the disease, to achieve a stratification of patients at high-risk for neuropathic pain and to enhance our understanding of underlying mechanisms. Inclusion criteria were (i) families with ≥2 affected members or (ii) patients with an early-onset of disease (<40 years). Ten Italian families with a total of 29 patients and 11 healthy controls, as well as 36 early-onset Italian patients were recruited. DNA extracted from peripheral blood was enriched using the Agilent SureSelect Human All Exon QXT V5. Paired-end sequencing (2x101bp) was performed on Illumina Hiseq 2500 platform. Reads were trimmed and aligned with BWA using hg19 as reference genome. Variants were identified using GATK pipeline and annotated using SnpEff suite. We focused on variants according to their depth (<6), evolutionary conservation (GERP++RS>0), mutation frequency (<0.01 for MODERATE and <0.05 for HIGH impact variants), in silico predicted effect (Polyphen2 >0.15, SIFT <0.05, and MetaSVM >0) and segregating under a transmission model according to the pedigree in familial cases or shared among patients in the early-onset sporadic cohort. Until now we sequenced 23 subjects belonging to 8 families and 36 early-onset patients reaching a mean coverage of 90X. In two large families, characterized by an autosomal dominant transmission, a mean of 17.000 variants of HIGH and MODERATE impact were identified. Following a dominant transmission model, a total of 49 and 40 rare/novel variants with a putative functional role were identified respectively in the first and second family. Analyses on additional families and early-onset cases are ongoing at present time. The application of exome-sequencing in familial or atypical cases of painful neuropathy is effective in identifying causative genes, and can help to better understand the pathophysiology of the disease.

1374F

Exome sequencing identifies targets in ophthalmoplegic myasthenia gravis. *M. Nel[1], M. Jalali Sefid Dashti[2], S. Prince[3], J. Gamieldien[2], J.M. Heckmann[1].* 1) Department of Medicine, University of Cape Town, South Africa; 2) South African National Bioinformatics Institute, University of the Western Cape, South Africa; 3) Department of Human Biology, University of Cape Town, South Africa.

Background: We described a distinct ophthalmoplegic complication (OP-MG) in myasthenia gravis (MG) patients of African-genetic ancestry that remains treatment-resistant. A previous candidate approach identified 2 African-specific gene expression traits linked to OP-MG, decay accelerating factor (*DAF*) and transforming growth factor beta-1 (*TGFB1*). We speculate that OP-MG may result from a genetic network activated in extraocular muscles (EOMs) by autoimmune MG, in these and as yet unknown targets. Objective: To use the unbiased approach of whole exome sequencing (WES) in a selected cohort of juvenile-onset OP-MG vs control MG individuals to discover OP-MG susceptibility variants. Method: Next-generation sequencing (Agilent v5, Otogenetics, USA) of the coding and untranslated regions (UTRs) was performed on genomic DNA from 15 African ancestry individuals at 50x coverage (11 OP-MG vs 4 control MG). Read alignment and variant calling were carried out. Filtering strategies were employed to prioritize variants for Sanger sequencing in a replication sample (14 OP-MG, 46 control MG). Gene expression of top variants was assessed by qPCR in a phenotype-specific cell culture model. Results: We identified 356 potentially pathogenic variants (>60% in 3' UTRs) which were more common in OP-MG (MAF ≥0.27) compared to control MG (MAF ≤0.12). Two filtering strategies were then applied: unbiased prioritization of variants in genes transcribed in EOM according to strength of association with OP-MG (p ≤0.05), and functional filtering of variants related to biological relevance for known MG disease pathways. This approach identified 13 variants associating with OP-MG in the discovery set; the top candidate, a 3'UTR variant in *DDX17*, is implicated in myogenesis and was also shown to be associated with OP-MG in the replication set (OR 2.4, p=0.014). Consistent with the hypothesis that OP-MG is a polygenic problem (multiple genetic risk loci with small effect sizes), we found a weighted genetic risk score for candidate and unbiased variants positively correlated with OP-MG. Preliminary results show that our "muscle model" was informative in showing altered transcript levels in at least 2 of our gene candidates. Discussion: Exome sequencing in a well-characterized discovery set has identified several potentially pathogenic variants that associate with OP-MG. Preliminary results suggest novel gene variants involved in myogenesis may cross-talk with known MG-associated pathways.

1375W

Rapsyn N88K mutation is not a prevalent cause of Congenital Myasthenic Syndrome in Iranian patients. *S. Parvizi Omran[1], P. Karimzadeh[2,3], R. Parvizi Omran[4], M. Houshmand[5].* 1) Department of Biology, Damghan Branch, Islamic Azad University, Damghan, Iran; 2) Department of Pediatric Neurology, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 3) Pediatric Neurology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 4) Department of Biology, Concordia University, West, Montreal, Canada; 5) Department of Human Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran.

Congenital myasthenic syndromes (CMSs) are considered as a rare heterogeneous group of disorders that cause dysfunction of neuromuscular transmission. In CMSs impaired neural transmission have identified in pre-synaptic, synaptic and most frequently in post-synaptic stages. Mainly, post-synaptic CMSs are caused by deficiency or kinetic abnormalities of acetylcholine receptor (AChR). The major clinical signs of CMSs are hypotonia, episodic apnea, ptosis and fatigability that appear in early infancy or childhood. Mutations in *RAPSN* [MIM 601592] are the second most common cause of autosomal recessive post-synaptic CMSs. In the present study, were aimed at identifying mutations in *RAPSN* gene in 18 patients with confirmed post-synaptic CMSs. For the mutation screening, we used polymerase chain reaction (PCR) to amplify exon 2 of *RAPSN* gene, then direct DNA sequencing was used for mutation identification. All the patients were negative for mutations c.264C/A (p.Asn88Lys), c.271C/T (p.Arg91Cys), c.280G/A (p.Glu94Lys), c.490C/T (p.Arg164Cys), and c.493G/A (p.Val165Met) of *RAPSN* gene. We found a previously reported polymorphism c.193-15C/T (IVS1-15C/T) in a 4-year-old boy , while there is no relation between the known polymorphism to CMSs. Regarding to these findings, N88K is not a recurrent mutation in CMSs patients from Iran. It could be concluded that mutations spectrum of *RAPSN* gene in CMSs is population specific. Further studies that screen RAPSN gene full sequence for possible variations are warranted to more elucidate mutational basis of CMSs in Iran. **Keywords:** Congenital mysthenic syndrome; Rapsyn; Mutation N88K; Neuromuscular disease.

1376T

Whole exome sequencing of multiplex families with late onset Alzheimer's disease identifies rare coding variants in known candidates and novel genes. *S. Rolati[1], H.N. Cukier[1,2], B.W. Kunkle[1], K.L. Hamilton-Nelson[1], N.K. Hofmann[1], M.A. Kohli[1], P.L. Whitehead[1], J.M. Vance[1,3], M.L. Cuccaro[1,3], R.M. Carney[1,4], J.R. Gilbert[1,3], E.R. Martin[1,3], G.W. Beecham[1,3], J.L. Haines[5], M.A. Pericak-Vance[1,2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Miami, Miami, FL; 3) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 4) Mental Health and Behavioral Sciences Service, Miami Veterans Affairs, Miami, FL; 5) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH.

Alzheimer's disease is a neurodegenerative disorder that can present with either an early (<65 years) or late (65+ years) age of onset. While greater than 25 loci have been associated with late onset Alzheimer's disease (LOAD), there is still a large number of genetic factors that have yet to be revealed. We performed whole exome sequencing (WES) on 28 multi-generational LOAD families to identify rare LOAD risk variants. The pedigrees chosen display a dominant inheritance pattern and have an average of eight Alzheimer affected subjects. Between 1-9 individuals were sequenced for each family. Variants were filtered for segregating, nonsynonymous, and loss-of-function rare variants (MAF <0.01). Variants were given a high priority if they were predicted to have a functional consequence as represented by a high CADD score (>15), fell within a family-specific or previously reported linkage peak, and by comparing results across families. Following WES, 10 families were identified to have variants of interest located in peak LOD regions (LOD=1.02-2.95). Examples include a family with a variant co-segregating in 11 AD individuals in *TTC3*, a regulator of Akt signaling (LOD=2.66, Cukier, et al, 2016), and a different family with 10 affected individuals carrying an alteration in *CD163L1*, a gene involved in immunity and endocytosis (LOD=2.95). Another segregating variant was identified in the tumor suppressor gene *CTNNA1*, which encodes a catenin protein involved in the cerebral cortex (LOD=2.36). We also discovered segregating variants in the reported LOAD gene *SORL1*. When comparing variants across families, genes with alterations identified in more than one family include *PLEKHG5*, a gene reported to be involved in Charcot-Marie-Tooth Disease as well as distal spinal muscular atrophy, and *THBS2*, which is involved in synaptogenesis in the central nervous system. In conclusion, WES of large LOAD pedigrees identified genes with rare variants co-segregating with disease, some of which were identified in more than one family. Given the known functions of some of these genes, they have the potential to be involved in AD risk.

1377F

Familial essential tremor candidate genes have been identified by whole exome sequencing study. *M. Sen[1,2], I.O. Tuncay[1,2], S. Yaylacı[2], O.E. Onat[3], C. Ulusoy[4], C. Akbostancı[4], O. Dogu[5], H. Topaloglu[6], T. Ozcelik[2,3], A.B. Tekinay[1,2].* 1) Neuroscience, Bilkent University, Ankara 06800, Turkey; 2) Institute of Materials Science and Nanotechnology, National Nanotechnology Research Center, Ankara 06800, Turkey; 3) Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800, Turkey; 4) Department of Neurology, Faculty of Medicine, Ankara University, Ankara 06100, Turkey; 5) Department of Neurology, Faculty of Medicine, Mersin University, Mersin 33343, Turkey; 6) Pediatrics, Neurology Unit, Faculty of Medicine, Hacettepe University, Ankara 06100, Turkey.

Essential tremor (ET) is the most common inherited movement disorder and it is known to be generally transmitted as an autosomal dominant trait. Although the main cause of ET is not known, there is a consensus about the presence of etiologic heterogeneity and a strong genetic predisposition. Genetic mapping and candidate gene sequencing revealed that there are few susceptibility genes related with ET. Our approach is to search for rare, high-penetrance, and gene-disrupting mutations through whole exome sequencing of the ET subjects in the setting of familial cases. For identifying genes that cause ET, clinical evaluation and correct diagnosis for probands and their relatives were performed based on WHIGET and MDS criteria. For whole exome sequencing, at least two affected and one control individuals were selected in the same family. Read pairs were aligned to the reference human genome hg19 (GRCh37) using the Burrows-Wheeler Aligner (BWA). Removal of PCR duplicates, sorting and indexing were done using SAMtools v0.1.18. Local realignments around insertions or deletions (indels) were performed using the Genome Analysis Toolkit (GATK v.3.1-1). Finally, whole exome sequencing data was analyzed with annotation tools. For segregation analysis, Sanger sequencing was carried out with selected candidate genes primers. We are searching for candidate genes which segregated with disease for different families. Variants that included any variation that was predicted to alter gene expression or protein function have been prioritized. Candidate alleles have also been selected predicted as damaging with in silico analysis tools. Identification of the causal mutations that are correlated with the ET phenotype will provide invaluable insights into underlying mechanism of the disease and will potentially open new avenues for treatment.

1378W

Whole exome sequencing approach in a cohort of multiplex Italian families with multiple sclerosis. *A. Zauli[1], C. Guaschino[1,2], E. Mascia[1], F. Esposito[1,2], M. Sorosina[1], A.M. Osiceanu[1], S. Peroni[1], S. Santoro[1], D. Biancolini[3], D. Lazarevic[3], S. Bonfiglio[3], D. Cittaro[3], V. Martinelli[2], G. Meola[4], S. D'Alfonso[5,6], G. Comi[1,2], F. Martinelli Boneschi[1,2].* 1) Laboratory of Human Genetics of Neurological Disorders, San Raffaele Scientific Institute, Milan, IT; 2) Department of Neurology, San Raffaele Scientific Institute, Milan, IT; 3) Center for Translational Genomics and BioInformatics, San Raffaele Hospital, Milan, IT; 4) Department of Neurology, San Donato Hospital, San Donato Milanese, IT; 5) Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, IT; 6) Department of Health Sciences, University of Eastern Piedmont, Novara, IT.

While the role of common genetic variants is established from whole-genome association studies (GWAS) in Multiple Sclerosis (MS), pointing to the existence of more than 100 non-HLA genetic loci, the heritability of the disease is still poorly captured, suggesting the existence of rare variants implicated in the susceptibility to the disease. The aim of the study is to identify rare variants involved in the predisposition of MS by analyzing multiplex families using whole exome sequencing approach. 12 Italian families with at least two affected relatives (1 with 5, 4 with 4, 6 with 3, and 1 with 2 affected) have been recruited as part of a large multicentric Italian study. DNA extracted from peripheral blood was enriched using the Agilent SureSelect Human All Exon QXT V5. Paired-end sequencing (2x101bp) was performed on Illumina Hiseq2500 platform. Reads were trimmed and aligned with BWA using hg19 as reference genome. Variants were identified using GATK pipeline and annotated using SnpEff suite. We focused on variants according to their depth (<6), evolutionary conservation (GERP++RS>2), mutation frequency in the general population (MAF<0.05), in silico predicted effect (Polyphen2 >0.15, SIFT <0.05, and MetaSVM >0), and segregating under a transmission model according to the pedigree. WES was performed in 37 relatives with a general coverage >70X in all individuals. Around 40.000 functional variants were found in affected relatives, of which 16.000 with frequency <5%. We focused on variants shared among families, as well as variants which segregate across families. We also looked at variants which fall in MS-associated genetic loci, as well as genes known to be implicated in MS-like diseases (mithocondrial lysosomal and peroxisomal disorders, leukoencephalopathies and neurometabolic diseases). These preliminary data represent a potential pool of rare functional variants implicated in MS pathophysiology that should be further analyzed in larger datasets of familial and sporadic cases.

1379T

Genome-wide linkage analysis in Dutch families identifies novel candidate genes for Alzheimer's disease. *S. Ahmad[1], N. Amin[1], E. Blue[2], S.J. van der Lee[1], A.Q. Nato[2], H. Sohi[2], B. Wang[5], E. Boerwinkle[3], A.L. DeStefano[4], E. Wijsman[2,6], C.M.V. Duijn[1], Alzheimer's disease sequencing project (ADSP).* 1) Erasmus University Medical Center, Department of Epidemiology, Genetic Epidemiology Unit. PO Box 2040 3000CA Rotterdam , Netherlands; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; 3) Human Genetics Center, University of Texas Health Sciences Center at Houston, Houston, TX, 77030, USA; 4) The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham MA, USA; 5) Department of Statistics, University of Washington, USA; 6) Department of Biostatistics and Genome Sciences, University of Washington, USA.

Purpose: Alzheimer's disease is one of the most heritable disorders in old age, and both rare as well as common variants are implicated. Whole genomes of 30 Alzheimer's disease (AD) patients from two large and complex multigenerational Dutch families were sequenced as a part of the ongoing Alzheimer's disease sequencing project (ADSP) to identify rare sequence variation that can have large effects on AD. To identify regions that might harbour rare variants, multipoint parametric and non-parametric linkage analyses were performed, with implicated regions evaluated for potential causal variants. **Methods**: We selected maximally informative heterozygous single nucleotide polymorphisms (SNPs) from an Illumina SNP microarray (370K) for linkage analysis. Affected-only linkage analysis was performed in MERLIN with 10,431 SNPs by splitting the 30 cases into six smaller sub-families. Findings were validated using Markov chain Monte Carlo (MCMC) in the two intact large families with analysis from 2734 uniformly-spaced, maximally informative SNPs in the genome scan. To identify the potential explanatory variants, we searched for sequence variation shared by the AD patients of the contributing families in the LOD-2 support interval of the linked regions. **Results**: Significant linkage evidence was observed for chromosome 5p14.3 (logarithm of odds (LOD) score MERLIN=3.29; MCMC=3.12). The LOD score was mainly determined by one family with 7 AD patients. The 7 cases shared at least 294 rare sequence variants (including 2 deleterious variants) within the LOD-2 interval. The region encompassing these variants is a highly conserved region and contains a cluster of Cadherin genes most notably *CDH12* and *CDH18*. *CDH* genes are highly expressed in the nervous system and are known to be involved in presenilin-mediated signaling. Furthermore, *CDH12* is predicted to be involved in the development of hippocampus and cerebellum and *CDH18* is predicted to be involved in memory. We validated the region using MCMC and extend the pedigree with 6 new patients. There is a 50 bp haplotype shared among subjects who drive the linkage sgnal in the extended pedigree. **Conclusions**: AD is linked to chromosome 5p14 in a large pedigree from the Netherlands, with 13-18 patients out of 30 contributing substantially to the evidence for linkage. Based on known and predicted function, the cadherin genes are the most likely candidate genes.

1380F

Whole genome sequencing analysis of a human with total insensitivity to capsaicin. *H. Kim[1], J. Park[2], J. Gill[1], S. Yun[3], A. Cashion[1], H. Na[2].* 1) SML, NINR/NIH, Bethesda, MD; 2) Dept. Physiology, Korea University, Seoul, Korea; 3) Yotta Biomed, Bethesda, MD.

Because the transient receptor potential V1 (TRPV1) is known to mediate chemical, thermal and pain stimuli, the capsaicin, the source of spicy taste in hot peppers and the agonist of TRPV1, has been suggested to affect multiple neuroscience behaviors. Various genetic variations in TRPV1 and other TRP channel encoding genes have been discovered, but it is not yet clear what genetic factors influence capsaicin sensitivity.We tested and sequenced a 51 year old female Korean subject who has a complete lack of sensitivity to capsaicin but hypersensitivity to allicin, along with her 55 year old sister control subject with normal to both capsaicin and allicin. Both subjects had normal responses to other sensory tests. Illumina Hiseq 2500 was used to sequence whole genome for both subjects. For analysis, base call quality was determined using FASTQC software. Alignment was performed on hg18 with BWA, version 0.7.10. Variation calling was performed using the haplotypecaller in GATK, version 3.4-46. ANNOVAR was used to annotate variants.We obtained 324 million reads for the patient, and 318 million reads for her normal sister in paired-end sequencing, which corresponds to ~ 32x coverage. We obtained 4.9 million and 4.8 million variants for the patient and her normal sister, respectively. We selected homozygous insertions/deletions, stop-gain/loss, missense, or splice-site mutation. Then, we filtered out her normal sister's variants and Korean common SNPs. We filtered out the variants further by requiring minor allele frequency less than 0.01 in 1000 Genome database. Finally, we obtained 57 genes, and performed pathway analysis with Ingenuity IPA. IPA revealed 3 associated network functions (cancer, molecular transport, drug metabolism) and 5 upstream regulators. These findings suggest an analytical tool to investigate whole genome sequencing data with small sample sizes. It will also lead to an increased understanding of how dynamic regulation of the genome is associated with capsaicin sensitivity and other neuroscience behaviors.

1381W

Deciphering the genetic basis of Hirschsprung disease by whole genome sequencing. *C. Tang[1,2], H.S. Gui[3], M.T. So[1], S.S. Cherny[3,4], P.C. Sham[3,4], M.M. Garcia-Barceló[1], P.K.H. Tam[1].* 1) Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong; 2) Dr. Li Dak-Sum Research Centre, The University of Hong Kong - Karolinska Institutet Collaboration in Regenerative Medicine, The University of Hong Kong, Pokfulam, Hong Kong; 3) Genome Research Centre, The University of Hong Kong, Pokfulam, Hong Kong; 4) Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.

Hirschsprung disease (HSCR) is a congenital disorder characterized by the absence of enteric ganglia along the lower intestinal tract. The disease is caused by the failure in migration, differentiation and/or proliferation of the enteric neural crest cells to the hindgut, leading to colonic aganglionosis. The differential contributions of rare versus common variants as well as coding versus noncoding variants tend to vary in accordance with the length of aganglionosis. Short-segment HSCR (S-HSCR), accounting for ~80% of the patients, is believed to be genetically complex whereas the rarer subtype, long-segment HSCR (L-HSCR), could be attributed to damaging *de novo* or recessive mutation. To understand comprehensively the genetic basis across the disease spectrum, we carried out an integrative analysis by high coverage whole-genome sequencing (30-45X) of 443 Chinese S-HSCR, 11 L-HSCR of 10 families (9 trios and 1 quartet) and 493 controls of East Asian ethnicity, aiming to identify novel HSCR genes. Gene-based association analysis on rare variants of S-HSCR not only vindicated the causal roles of the two known genes—*RET* (SKAT-O P=6.8x10^{-6}) and *EDNRB* (CMC P=0.0055)—but also revealed higher burden of rare protein-altering variants in *PTGS2* and *BACE2* (P<0.001) in S-HSCR. Our data also show that *de novo* as well as inherited variants contribute to the development of enteric nervous system and thereby to HSCR. Recent progress of CNV analysis will also be presented.

1382T

Project MinE: Study design and pilot analyses of a large-scale whole genome sequencing study in amyotrophic lateral sclerosis. *W. van Rheenen, S.L Pulit, R.A.A. van der Spek, A.M. Dekker, B.M. Middelkoop, L.H. van den Berg, J.H. Veldink, Project MinE Sequencing Consortium.* Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease affecting 1 in 400 people. We recently showed a disproportionate contribution from low-frequency variants in genetic susceptibility to ALS. We therefore have begun Project MinE, an international collaboration that seeks to whole-genome sequence 15,000 ALS patients and 7,500 controls. Here, we report on the design of Project MinE and pilot analyses including 1,264 ALS patients and 611 controls. We find an abundance of rare genetic variation (allele frequency < 0.1%), of which the vast majority is absent in public data sets. Principal component analysis reveals that strong geographical clustering of these variants even within The Netherlands. We further test the implications of poor geographical matching of cases and controls and investigate how externally sequenced controls, selected from the identical population as cases, can induce false positive associations and reduce power to discover true genetic signal. Finally, we present the project MinE variant server: a publicly available server to interrogate the genome for (rare) genetic variation observed in project MinE.

1383F

Investigating the genetic contribution to Amplified Musculoskeletal Pain Syndromes (AMPS) in pediatric populations. *M.V. Gonzalez[1,2], S.A. Gmuca[2], J. Tress[2], F.D. Mentch[1], R. Chiavacci[1], P. Sleiman[1], G.A. Barr[3], D.D. Sherry[2,4], H. Hakonarson[1,4].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA. 19104; 2) Amplified Pain Program, Division of Rheumatology, The Children's Hospital of Philadelphia, Philadelphia, PA. 19104; 3) Department of Anesthesiology and Critical Care Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA. 19104; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA. 19104.

Chronic pain is one of the most costly conditions in modern western society. Pediatric chronic pain conditions have been estimated to account for more than 5.2 billion dollars annually. Many diagnostic terms are used, however, a common thread in these pain conditions is central and/or peripheral sensory amplification. Amplified musculoskeletal pain syndromes (AMPS) comprise a range of disorders that can present with a variety of signs and symptoms related to increasing pain over time, allodynia, and disproportional dysfunction. AMPS are most prevalent in adolescent females of European Caucasian (EC) descent (approx. 75%) and often significantly interfere with daily function and quality of life. Additionally, pediatric amplified musculoskeletal disorders differ significantly from those occurring in adults. The etiology and pathophysiological mechanisms of AMPS are currently unknown. Familial aggregation and candidate gene studies suggest a strong genetic component in the development of AMPS, but many of the genetic susceptibility factors remain unidentified. Genome-wide association studies (GWAS) facilitate impartial queries of common genetic variation, allowing identification of novel genetic contributions to the development of complex disorders. To our knowledge, this is the first report of a genome-wide investigation of AMPS in a large pediatric cohort. To this end, 952 male/female AMPS cases and 8,930 controls of African American and European Caucasian descent were selected for genotyping utilizing single nucleotide polymorphism (SNP) arrays. Logistic regression models were used to determine association between AMPS incidence and SNP genotype. Because European Caucasian females are predominantly affected with AMPS, these samples were separated from the total population and analyzed separately (359 cases, 4151 controls). The total AMPS and female EC analyses identified 6 and 15 genomic regions respectively. Many of these implicated genomic loci included genes influencing neurogenesis, neurotransmission, and regulation of the innate and adaptive immune response. Replication and validation of these results in an independent cohort of 629 AMPS cases is currently underway. The genes identified in these initial genome-wide investigations of AMPS deliver a number of possible functional candidates and may provide novel therapeutic targets for children suffering from neurologic pain disorders.

1384W

A common genetic basis between neurodevelopmental disorders of childhood. *F. Tsetsos[1], D. Koukousi[1], M. Mpampali[1], S.S. Padmanabhuni[1], J. Alexander[1], I. Karagiannidis[1], D. Mantzaris[1], M. Georgitsi[1,2], P. Drineas[3], P. Paschou[1].* 1) Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Evros, Greece; 2) Laboratory of General Biology, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece; 3) Computer Science Department, Rensselaer Polytechnic Institute, Troy, Ny, USA.

Gilles de la Tourette Sydrome (TS) is a childhood onset neurodevelopmental disorder, characterized phenotypically by the presence of multiple motor and vocal tics. It is often accompanied by multiple neuropsychiatric comorbidities, with Attention Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorders (ASD) among the most common. The extensive co-occurrence of the three disorders suggests a shared genetic background. A major step towards the elucidation of the genetic architecture of TS was performed by the first TS Genome-wide Association Study (GWAS) reporting 552 SNPs that were moderately associated with TS (p<1E-3) [Scharf et al, 2013]. The latest genome-wide attempts towards the identification of the genetic background of ADHD and ASD were described in the cross-disorder meta-analysis of the Psychiatric Genomics Consortium [Smoller et al. 2013]. In this study we examine the common genetic background of the three neuropsychiatric phenotypes, by meta-analyzing the 552 top hits in the TS GWAS with the results of the GWAS for ADHD and ASD that preceded the cross-disorder meta-analysis undertaken by the PGC. We identified 19 significant SNPs in the TS/ADHD meta-analysis, with the top four implicated genes being TBC1D7, GUCY1A3, RAP1GDS1 and CHST11. TBCD17 harbors the top scoring SNP, rs1866863 (p:3.23E-07), located in a regulatory region downstream of the gene, and the third best-scoring SNP, rs2458304 (p:2.54E-06), located within an intron of the gene. Both variants were in linkage disequilibrium with eQTL rs499818, indicating a role in the expression levels of the gene. TBC1D7 is the third subunit of the TSC1/TSC2 complex, with a Rheb-GAP activity. The top genes indicate a complex and intricate interplay between them, warranting further investigation into a possibly shared etiological mechanism for TS and ADHD. In the TS/ASD meta-analysis, we identified 10 significant SNPs, with the top two residing in the same region (rs8004624, p: 5.57E-06 and rs4898764, p:1.11E-05), in an LD-block that harbors the genes STYX, PSMC6 and GNPNAT1. Our study will also include the results of the EMTICS and TS-EUROTRAIN initiatives that will generate genome-wide genetic data from another 2600 European patients with TS. Our results provide some interesting insight into the common genetic mechanisms that underlie neurodevelopmental disorders, with the main theme of the TS/ADHD overlap being the structural and circulatory function of specific areas in the brain.

1385T

Integration of genetic, epigenetic, and transcriptional variation in the context of multiple disease phenotypes reveals diverse molecular signatures of Alzheimer's disease. *M. Kellis[1,2], Y. Park[1,2], Z. Zhang[1,2], M. Eaton[1,2], D. Bennett[3], P. De Jager[4].* 1) Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Rush University, Chicago, IL; 4) Harvard Medical School, Brigham and Women's Hospital, Boston, MA.

Perhaps the greatest surprise of human genome-wide association studies (GWAS) is that 90% of disease-associated regions do not affect proteins directly, but instead lie in non-coding regions with putative gene-regulatory roles. This has increased the urgency of understanding the non-coding genome, as a key component of understanding human disease. To address this challenge, we generated maps of genomic control elements across 127 primary human tissues and cell types, and tissue-specific regulatory networks linking these elements to their target genes and their regulators. We have used these maps and circuits to understand how human genetic variation contributes to disease, providing an unbiased view of disease genetics and sometimes re-shaping our understanding of common disorders. For Alzheimer's Disease, we find evidence that genetic variants contributing to disease predisposition act primarily through immune processes, rather than neuronal processes. To gain insights on the intermediate phenotypes that lead to disease onset, we study epigenetic, transcriptional, and phenotypic information from risk and non-risk individuals using postmortem brain samples and longitudinal phenotypes in a cohort of aging. We find that methylation differences are highly correlated with Alzheimer's disease diagnosis, and lead to significantly higher classification power than genetic information alone. The genomic regions that show the highest correlation with Alzheimer's disease status are significantly enriched for genetic variants associated with Alzheimer's disease, including large numbers of loci that do not reach genome-wide significance, indicating the potential to identify new disease loci. We find that the genomic regions showing methylation differences associated with Alzheimer's disease can be partitioned according to the subset of phenotypes that they are associated with, including neuritic plaques, neurofibrilary tangles, and cognitive decline. These subsets are enriched in diverse functional categories based on the gene ontology classifications of nearby genes, indicating diverse functions in adherens, signaling, and immune functions. They are also enriched in enhancer and promoter regions active in distinct groups of cell types and differentiation stages, including embryonic stem cells, immune cells, and fetal and adult brain cells. These results indicate that epigenetic profiling can help partition complex trait associations into diverse intermediate phenotypes.

1386F

Identification and functional characterisation of a novel regulatory variant in monoamine oxidase-A (*MAO-A)* **gene in Parkinson's disease.** *S. Kumar[1], B.K. Thelma[1], K.A. Michealraj[1], U. Muthane[2].* 1) Dept of Genetics, Delhi University, New Delhi, India; 2) Parkinsons & Ageing Research foundation, Bangalore, India.

Parkinson's disease (PD) is a common progressive complex disorder of the elderly caused by the interaction of genes and environmental factors. Loss of dopaminergic neurons in *substantia nigra* results in dopamine (DA) depletion which results in motor symptoms associated with this disease. Oxidative stress is believed to be a key player in this neuronal loss. *Post mortem* reports of PD patients show higher levels of lipid peroxidation. Degradation of DA leads to formation of reactive oxygen species and thus DA metabolism may itself be a major contributor to oxidative stress. Consequently, DA metabolising genes such as *COMT, DBH, DDC* and *MAO* etc. are strong candidates in PD pathology as variants in these genes will affect the DA determined phenotypes. Inhibitors of such gene products are routinely used in PD treatment. Genetic variants in *COMT*, *MAO*, *DAT, NET,* and *5- HTT* genes from DA metabolism pathways have been documented to alter amines and their metabolic product levels. With this background, we attempted to establish genotype-phenotype correlations in two dopamine metabolising genes namely *MAO-A* and *MAO-B.* Enzymatic activity in the blood sera of same cohort was measured using Amplex Red Monoamine oxidase kit (Thermo Fisher Scientific, USA) and protein quantitation was done by by Elisa (MyBioSource, USA). Deep sequencing of exonic and regulatory regions of these two genes in 50 healthy individuals was carried by Dr. K. A. Michealraj using Ion torrent platform; A novel 5' UTR variant was identified in an individual with higher MAO-A quantity and activity. Interestingly, the same variant was also identified in a multi-generation family with late onset of PD. The proband in the family had increased sleep disturbance and hypotension in later stages which is in accordance with the reported effects of higher MAO-A enzyme levels. To establish the genotype-phenotype correlation of this novel regulatory variant, wild type and variant alleles with native MAO-A promotor were cloned in the luciferase expression vector. Results from ongoing expression assays in human cell line SH-SY5Y expressing MAO-A together with test of association of this variant in an independent PD case-control cohort will be presented.

1387W

Whole genome sequencing and rare variant discovery in the ASPIRE autism spectrum disorder cohort. *B. Callaghan[1,2], P. Tan[1,2], S. Rogic[1,2], K. Calli[3], Y. Qiao[4], M. Jacobson[1,2], M. Belmadani[1,2], N. Holmes[1,2], C. Yu[5], Y. Li[5], Y. Li[5], F. Kurtzke[3], A. Yu[3], M. Hudson[6,7], A. Dionne-Laporte[8,9], S. Girard[10], P. Liang[11], E. Rajcan Separovic[4], X. Liu[6,7], G. Rouleau[8,9], P. Pavlidis[1,2], S. Lewis[3].* 1) Michael Smith Laboratories, UBC, Vancouver, BC; 2) Department of Psychiatry, UBC, Vancouver, BC; 3) Department of Medical Genetics, UBC, Vancouver, BC; 4) Department of Pathology and Laboratory Medicine, UBC, Vancouver, BC; 5) BGI Tech Solutions; 6) Queen's Genomics Lab at Ongwanada, Ongwanada Resource Center, Kingston, ON; 7) Department of Psychiatry, Queen's University, Kingston, ON; 8) Montreal Neurological Institute, Montréal, QC; 9) Department of Neurology and Neurosurgery, McGill University, Montréal, QC; 10) Department of Human Genetics, McGill University, Montréal, QC; 11) Department of Biological Sciences, Brock University, St. Catharines, ON.

We characterized rare genetic variation in a cohort of 119 individuals with autism spectrum disorder (ASD) using whole genome sequencing. The cases were drawn from the ASPIRE (Autism SPectrum Interdisciplinary REsearch) British Columbia cohort. Individuals in this cohort have been diagnosed with ASD (DSM-IV; ADOS-G/ADI-R) and underwent a detailed standardized phenotyping protocol including morphometrics. We obtained whole genome sequences (Illumina paired-end 100 base pair reads, average depth 30x) for each subject. Variants were called using the Genome Analysis Toolkit (GATK) against the human reference genome, and filtered for quality and rarity in population data. To prioritize variants further, we relied on a combination of existing and in-house bioinformatics tools incorporating both gene-level metrics (genic intolerance to mutation, functional effect prediction) and variant level metrics (population frequency, predicted damage). Our efforts included recurating variants reported in the ASD literature, totalling 4768 variants from 28 publications. We have made the resulting database available as MARVdb (www.chibi.ubc.ca/marvdb). To collaboratively analyze variants we also applied our general interactive assessment tool, ASPIREdb (aspiredb.chibi.ubc.ca). In total 97 high-priority candidate variants were identified, affecting 66 subjects. All of these variants were heterozygous and all but one were autosomal. Of these, 31 were predicted loss-of-function mutations, 3 affecting genes previously associated with ASD and 3 affecting genes associated with other neurodevelopmental disorders. A total of 66 candidate MS mutations were prioritised as potentially pathogenic, including 12 affecting literature-associated ASD genes, and 8 affecting genes associated with other neurodevelopmental disorders. High priority variants are being subjected to trio resequencing to assess inheritance. Using this approach we have thus far identified a novel de novo splice site variant in *SCN2A*, predicted to result in a loss of function, in an individual with severe autism and intellectual disability. This finding adds to the evidence that mutations in this gene can be associated with autism without comorbid seizure disorders.

1388T

Novel gene discovery in familial amyotrophic lateral sclerosis. *J.A. Fifita, K.L. Williams, S. Yang, K. Zhang, G.A. Nicholson, D.B. Rowe, I.P. Blair.* Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South WaFaculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW Australia, Australia.

   Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a fatal neurodegenerative disease that is caused by the progressive death of motor neurons. Ten percent of cases are familial (FALS), with the remaining 90% considered sporadic (SALS). To date, the only known causes of ALS are gene mutations. Mutations in ALS genes, including *SOD1, TARDBP, C9ORF72,* and *CCNF* account for approximately 60% FALS and 10% of SALS. We aim to identify remaining ALS genes and mutations. Gene discovery strategies utliize exome sequence data from Australian familial ALS cases. We have completed whole exome capture and sequencing of probands from 74 ALS families, as well as ten individuals from three multigenerational ALS families negative for all known ALS genes. Custom bioinformatics analysis of these families has identified candidate gene mutations in genes not previously linked to ALS. A custom *in silico* prioritisation pipeline was designed to predict pathogenicity of all candidates and prioritize for functional studies. To predict pathogenicity the pipeline includes standard use of multiple protein prediction software and analysis of conservation data. In addition to genic tolerance information, large control datasets are used to determine the natural variation of each gene. To predict a link between potentially pathogenic variants and ALS, protein expression, function and interacting partners are compared to what is known of ALS pathobiology. This pipeline has prioritized candidate variants. All candidates from Family One have undergone functional studies to examine potential pathogenicity *in vitro* using neuronal cell lines. These analyses begin with a general toxicity assay to determine pathogenicity. To determine causality, specific aspects of ALS-linked pathology were investigated, with results reflecting the *in silico* predictions. The pathogenic basis of ALS remains poorly understood. The identification of novel ALS genes increases our knowledge of disease biology, adds to diagnostic regimes, and provides tools for the development of cell and animal models and long-term therapeutic discovery.

1389F

Short sequence variations (SSVs) may explain multiplicity of *SOD1*-**SNP mutations in Amyotrophic Lateral Sclerosis.** *A.D. Roses[1,2], P.A. Akkari[2], M.W. Lutz[1], D.K. Burns[2], R. Saul[3], T. Siddique[4].* 1) Neurology, Duke University School of Medicine, Chapel Hill, NC; 2) Shiraz Pharmaceuticals, Inc., Chapel Hill, NC; 3) Polymorphic DNA Technologies, Inc, Alameda, CA; 4) Neurology, Northwestern University School of Medicine, Chicago, Il.

   There are multiple SNPs assigned as specific "mutations" for each of several amyotrophic lateral sclerosis [ALS] genes. More than 150 separate SNP variants within SOD1 have been proposed as familial ALS-*SOD1* "mutations" to date. We initially screened short structural variants [SSVs] located near SOD1 in familial and sporadic ALS patients [fALS and sALS], and observed an association between rs71714698 18T allele and ALS patients. This initial survey was performed with 159 fALS and sALS cases and 33 age matched controls. Additional SOD1, TARBDP and C9ORF72 cases and controls were added to this cohort: 205 ALS cases consisting of 147 SOD1 cases with 32 different SOD1 mutations, 11 C9ORF72, 15 TARDBP familial ALS patients, 32 sporadic ALS cases, and 459 controls and genotyped in a follow-up study. The addition of cases and controls strengthened the initial association between rs71714698 18T allele and ALS. Patients who carried at least one rs71714698 18T allele included; 29 SOD1 cases (28 A4V, 1 G100K); 3 ALS cases formerly scored as sporadic, 1 TARDBP N325B case and 2 C9ORF72 fALS cases compared to controls (p=1*10^{-12}, Likelihood Ratio test). 27 out of 31 ALS cases who carry the rs71714698 18T allele were positive for the common A4V mutation. Examination of the A4V and G100K ALS missense mutations reveals that they are 11327bp and 3790kb upstream from rs71714698 respectively. We then examined duration and age of onset for rs71714698 in ALS cases and observed that 18T was associated with the shortest duration until death of a patient when compared to 17T or 16T [p< 0.0054]. We hypothesize that the varying length of rs71714698 polyT may affect transcription similarly in those SOD1 SNPs clustered on 18T-carrying phylogenetic clade branches. SOD1 is flanked by POLR2A binding sites (https://genome.ucsc.edu) which may extend the length of the SSV affects to include many uncommon SNPs sharing the phylogenetic clustered clades carrying 18T, appearing as separate SNP "mutations." The common causative variant may be 18T for each uncommon SNP recognized as a "SOD1 mutation" that is cis-linked to 18T. This is analogous to the rs10524523 TOMM40-Long polyT in Alzheimer's disease clades. In-vitro confirmation of the functional effect of rs71714698 18T could therefore provide a rational for a pathogenetic mechanism and a strategy to select targets for drug discovery and pharmacogenetic-assisted drug development.

1390W

Next-generation sequencing of Alzheimer's and Parkinson's Disease genes in neuropathologically confirmed DLB, PDD and AD cases. *D. Tsuang[1], I. Mata[2], J. Leverenz[3], D. Bennett[4], J. Schneider[5], A. Buchman[5], J. Kaye[6], R. Woltjer[7], J. Quinn[6], J. Trojanowski[8], D. Weintraub[9], D. Irwin[10], V. Van Deerlin[8], E. Larson[11], P. Crane[12], W. Kukull[13], O. Lopez[14], R. Hamilton[15], J. Kofler[15], P. Nelson[16], G. Jicha[17], J. Neltner[16], D. Galasko[18], E. Masaliah[19], T. Bird[2], M. Dorschner[20], D. Hanna[20], M. Weaver[20], D. Yearout[2], C. Zabetian[2].* 1) Psychiatry and Behavioral Sciences, University of Washington/VA Puget Sound, Seattle, WA; 2) Neurology, University of Washington/VA Puget Sound, Seattle, WA; 3) Cleveland Clinic, Cleveland, OH; 4) Neurology, Rush University Medical Center, Chicago, IL; 5) Pathology, Rush University Medical Center, Chicago, IL; 6) Neurology, Oregon Health Sciences, Portland, OR; 7) Pathology, Oregon Health Sciences, Portland, OR; 8) Pathology, University of Pennsylvania, Philadelphia, PA; 9) Psychiatry, University of Pennsylvania, Philadelphia, PA; 10) Neurology, University of Pennsylvania, Philadelphia, PA; 11) Group Health Cooperative Research Insititute, Seattle, WA; 12) Medicine, University of Washington, Seattle, WA; 13) Epidemiology, University of Washington, Seattle, WA; 14) Neurology, University of Pittsburgh, Pittsburgh, PA; 15) Pathology, University of Pittsburgh, Pittsburgh, WA; 16) Pathology, University of Kentucky, Lexington, KY; 17) Neurology, University of Kentucky, Lexington, KY; 18) Neurology, University of California at San Diego, San Diego, CA; 19) Pathology, University of California at San Diego, San Diego, CA; 20) Pathology, University of Washington, Seattle, WA.

Background: Next-generation sequencing (NGS) has been useful in gene discovery for complex genetic disorders. We sought to use a comprehensive NGS-based panel to identify pathogenic variants in a large sample of autopsied subjects with dementia. Lewy body dementias (LBD) encompasses a spectrum of clinicopathologic entities that include Parkinson disease dementia (PDD), and dementia with Lewy bodies (DLB), which are differentiated from one another based on occurrence of parkinsonism relative to dementia. Neuropathologic changes include classic histologic inclusions of Lewy bodies, but whether these disorders share genetic risk factors remain unclear. Methods: All 314 subjects were diagnosed with dementia and classified as having neuropathologically-confirmed Alzheimer's disease (AD; n=71), Lewy body dementia with concomitant AD (LBD-AD; n=109), pure dementia with Lewy bodies (pDLB; n=78), and Parkinson's disease dementia (PDD; n=56). For each individual, we screened the entire coding region of 27 causative genes for AD, PD, and other dementing neurodegenerative disorders as follows: sample libraries were constructed and enriched for the genes of interest using xGEN custom probes. Using Illumina technology, >100X mean sequence coverage was generated for each sample; the resulting sequences were aligned to the human genome reference (hg19) using the Burrows-Wheeler Aligner and variants identified with the Genome Analysis Took Kit. We excluded those variants that did not meet standard quality values, those with >1% minor allele frequency in Exome Aggregation Consortium (ExAC) database and those classified as low impact by SnpEff prediction tools. Results: Twenty-five potentially pathogenic variant changes were identified, including 5 high-impact changes (e.g., stop-gained, frameshift mutations) and 20 moderate-impact changes (e.g., missense mutations, inframe deletions). All five of the high-impact variants were either previously reported but with very low frequencies in the ExAC database (< 0.0001) and/or were not found in ClinVar. These five high-impact novel variants were found in PDD (n=2) and LBD-AD (n=2) and pDLB (n=1) cases. Conclusions: Using a custom genetic sequencing panel, we identified potentially pathogenic variant changes in ~9 percent (23 out of 243) of our subjects with LBD. Our findings suggest that NGS methods are likely to uncover meaningful genetic changes and should be incorporated in future studies of LBD.

1391T

Association of genetic factors increasing serum glutamine metabolite levels with risk for migraine and depression. *D.R. Nyholt, IHGC and ENGAGE consortia.* Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Brisbane, QLD, Australia.

Migraine and major depressive disorder (MDD) are two frequent complex disorders, each affecting around 10–20% of the global population. Twin and family studies indicate moderate heritabilities ranging from 30 to 50% and point to the presence of shared genetic components underlying the two conditions. In 2013, the International Headache Genetics Consortium (IHGC) published results of a genome-wide association (GWA) meta-analysis of 23285 migraine cases and 95425 controls. This study of 1.8 million single nucleotide polymorphisms (SNPs) identified 12 loci robustly associated with migraine susceptibility ($P < 5×10^{-8}$).[1] In contrast, the 2013 Psychiatric Genomics Consortium (PGC) MDD GWA mega-analysis of 1.2 million SNPs in 9240 cases and 9519 controls failed to robustly identify any MDD risk loci.[2] While more recently, the ENGAGE (European Network for Genetic and Genomic Epidemiology) Consortium performed a large GWA meta-analysis of 131 serum metabolites in human blood (measured via mass spectrometry at Biocrates Life Sciences AG). This study of 2.6 million SNPs in 7478 individuals of European descent, found 4068 genome- and metabolome-wide significant associations, involving 59 independent SNPs and 85 metabolites ($P < 1.09×10^{-9}$).[3] To identify biological processes and pathways involved with migraine and depression aetiology, we tested for an association between genetic factors influencing the 131 ENGAGE serum metabolites and genetic risk factors for migraine and depression. Analysis of ~24000 independent SNPs with the most significant association *P*-values in the IHGC migraine GWA and PGC MDD GWA, found genetic factors for increased Glutamine metabolite levels were significantly correlated with genetic risk for both migraine and MDD. Our results strongly support previous suggestions that alterations in the glutaminergic system play a key role in migraine and MDD pathogenesis. References: 1. Anttila et al. Nat Genet 2013; 45(8):912-917. 2. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium et al. Mol Psychiatry. 2013;18:497-511. 3. Draisma et al. Nat Commun 2015; 6:7208.

1392F

LCM coupled with single-cell gene expression assay: Isolating neurons from archive frozen brains for functional genomic studies of neuro-degenerative diseases. *L. Tagliafierro[1,2], K. Bonawitz[1,2], O.C. Glenn[1,2], O. Chiba-Falek[1,2].* 1) Neurology, Duke University Medical Center, Durham, NC; 2) Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC.

Changes in gene expression in brain tissues from neurodegenerative disease patients compared to healthy controls have been reported. However, the majority of these studies used brain tissue homogenates that represent multiple cell-types. Gene expression profiles in a specific population of cells is more informative than studying the whole tissue, particularly in the brain, where different pathologies affect a specific subset of cells. Due to the difficulty to obtain homogenous cell populations, gene expression in specific cell-types (neurons, glial-types) has been understudied. Thus the specific cell-type responsible for the observed alteration in expression levels remained largely elusive. To leverage the archive resources of frozen human brains and address these key gaps in knowledge, we developed and calibrated a method to quantify cell-type specific – neuronal, astrocytes, microglial – expression profiles of neurodegenerative disease-risk genes. Archive human frozen brain tissues were used to prepare slides for rapid immunostaining. The immunoreactive-cells were isolated by Laser Capture Microdissection (LCM). We optimized the technique to preserve the RNA integrity, so that the RNA was suitable for downstream analyses. Following RNA extraction, the gene expression was determined digitally using nCounter Single Cell Gene Expression Assay (NanoString). The enrichment for a particular cell-type was determined by the expression of cell-specific markers. Comparisons of mRNA levels between brains from patients affected by neurodegenerative diseases, including Alzheimer's (AD) and Parkinson's (PD), and normal controls, revealed trends of differential expression profiles for several genes implicated in these pathologies. Noteworthy, *APOE-* and *TOMM40*-mRNAs were higher in neurons from Mild Cognitive Impairment (MCI) due to AD compared to neurons from normal brains, while the neuronal *SORL1*-mRNA levels were lower in MCI. *SNCA*-mRNA levels showed elevated levels in PD neurons. These results were consistent with our previous findings using brain tissues homogenates from AD and PD vs. control, and suggested that the alterations in expression of these genes occur in neurons. Experiments to identify *cis*-variants that contribute to the neuronal-regulation of these genes are underway. In conclusion, we suggest that neuronal-regulation of expression of disease critical genes may have a causative role in the etiology of neurodegenerative diseases of the aging brain.

1393W

HIV-related neurocognitive impairment is related to polymorphisms in *CCR2* and *CD163* in the absence of cocaine and opiate dependency. *M.M. Jacobs[1], D. Byrd[1,2], S. Morgello[1,3,4].* 1) Neurology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Pathology, Icahn School of Medicine at Mount Sinai, New York, NY.

Despite the widespread use of efficacious antiretroviral therapies, HIV-associated neurocognitive disorder (HAND) remains highly prevalent, and its dissociation from HIV replication makes it imperative to understand non-viral factors related to its neuropathogenesis. Chronic immune activation has been implicated, but the role of genetics has not been explored. Using an advanced-staged multi-ethnic HIV-positive population (n=276), we examined 230 polymorphisms within 55 genes proposed to play a role in both immune dysregulation and cognition. We examined these polymorphisms for associations with neuropsychological performance in global and cognitive domain T-scores (Motor, Processing Speed, Verbal Fluency, Learning, Memory, Executive Functioning, Working Memory) while controlling for opiate and cocaine dependency using linear regression analysis. While significant associations were observed in nearly every domain across both populations for multiple polymorphisms, one of the most significant effects in Caucasian subjects was observed in the Motor domain with the *CCR2* V64I polymorphism (rs1799864), such that nonsubstance users carrying the mutation had poorer motor performance while substance users with the mutation had better motor performance (p=0.004). When we examined levels of *CCR2* mRNA expression in peripheral blood mononuclear cells, we observed that nonsubstance users with the mutation had decreased expression of *CCR2* as compared to those without the mutation; such differences were not seen in drug users. For African- American subjects, the most significant effects were observed for several *CD163* polymorphisms in multiple domains (Global, Processing Speed, Verbal Fluency, and Working Memory). In African-Americans, nonsubstance users carrying *CD163* mutations had poorer performance, as compared to substance using individuals, who had better performance across multiple domains. Gene expression studies for *CD163* are ongoing, as are replication of these results with additional samples (n=500) from the National NeuroAIDS Tissue Consortium. These results suggest that substance use changes that neurobiological relationship of cognitive impairment to inflammatory processes, and should be factored into genetic studies of HAND.

1394T

Pathogenic *SORL1* mutations in Alzheimer's disease. *M. Cuccaro[1], R. Carney[2], Y. Zhang[4], C. Bohm[4], B. Kunkle[1], B. Vardarajan[3], P. Whitehead[1], H. Cukier[1], R. Mayeux[3], P. St. George-Hyslop[5], M. Pericak-Vance[1].* 1) University of Miami Miller School of Medicine, Miami, FL; 2) Bruce W. Carter VA Medical Center, Miami, Florida; 3) College of Physicians and Surgeons, Columbia University, New York, New York; 4) University of Toronto, Toronto, Ontario, M5T 2S8, Canada; 5) University of Cambridge, Cambridge, CB2 0XY, United Kingdom.

The *sortilin-related receptor LDLR class A repeats containing* (*SORL1*) gene has been implicated in both early and late onset Alzheimer's disease (AD). Located on chromosome 11q23.2-q24.2, *SORL1* plays a key role in differential sorting of the amyloid precursor protein (APP) and regulation of amyloid-β (Aβ) production. We describe clinical and molecular findings among individuals with AD and *SORL1* mutations. *SORL1* changes were identified by whole exome sequencing in 50 early onset AD families which contained at least one case with AD onset less than 60 years of age. The clinical consequences associated with SORL1 mutations were characterized based on clinical reviews of medical records. Functional studies were completed to evaluate (Aβ) production and APP trafficking associated with SORL1 mutations. Constructs were generated using human SORL1-MYC pcDNA3.1. Site directed mutagenesis was used to insert the T588I and T2134IM mutations verified by sequencing, and either the wild type or mutant constructs transfected into HEK293 cells expressing the Swedish APP mutant (APPsw). Cell culture and transfection followed previously described standard protocols. Aβ, Western blot, and co-immunoprecipitation assays were performed. A novel *SORL1* T588I change was identified in four individuals with AD from one EOAD family, two of whom also presented with Parkinsonian features. A second EOAD family was found to carry the T2134 alteration in three of four AD cases; one case with the T2134M variant was identified postmortem to have Lewy bodies. In a third family, two individuals, one with AD and one with MCI, were found to carry a p.Cys1431fs mutation in *SORL1*. Functional studies demonstrate that the variants weaken the interaction of *SORL1* with full-length APP, resulting in altered levels of Aβ and interfering with APP trafficking. Finally, a review of unrelated cases from a previously published study of *SORL1* mutations in late onset AD (Vardarajan 2015) identified four individuals with either *SORL1* A528T or T947M alterations who also presented with Parkinsonian features. The findings from this study support an important role for *SORL1* mutations in AD pathogenesis by way of altering Aβ levels and interfering with APP trafficking. In addition, the results point to a potential association of *SORL1* in the manifestation of Parkinsonian features among individuals with AD, expanding the phenotypic spectrum of *SORL1* mutations.

1395F

The evolving genetic risk landscape of sporadic amyotrophic lateral sclerosis. *J.M. Downie[1], S.B. Gibson[2], K.P. Figueroa[2], M.B. Bromberg[2], S.M. Pulst[2], L.B. Jorde[1].* 1) Department of Human Genetics, University of Utah School of Medicine, 15 North 2030 East Room 5100, Salt Lake City, UT 84112; 2) Department of Neurology, University of Utah School of Medicine, 175 North Medical Drive East, Salt Lake City, UT 84132.

Amyotrophic lateral sclerosis (ALS) is a disease in which the upper and lower motor neurons progressively degenerate, leading to paralysis and death. Understanding the genetic landscape of ALS is important in determining the pathogenic mechanisms of ALS and in the development of better clinical care. 90% of ALS cases occur sporadically (SALS) with no family history of the disease. It is known that genetic factors play a role in the pathogenesis of SALS; more than 33 genes have been associated with ALS. However, there is disagreement regarding the percentage of SALS cases that are caused by mutations in these genes, with estimates ranging between 11% and 28%. These calculations have been derived by determining the proportion of SALS cases that have a rare, protein-altering variant in a set of ALS genes. However, a large proportion of nonsynonymous variants are not deleterious. Thus, a shortcoming of these studies is that variant pathogenicity is not considered. To improve the understanding of ALS risk factors, we performed whole-exome sequencing of 96 SALS patients. BWA and GATK were used to perform read mapping and variant calling. Pathogenic repeat expansions of *ATXN2* and *C9orf72* were detected using PCR-based assays. Principal components analysis and *ADMIXTURE* identified 87 high-quality European SALS samples, which were used for analysis. Variants from these individuals were annotated by SnpEff to determine which ones produced nonsynonymous or splice-site gene changes. These variants were also annotated by dbNSFP v2.9, which contains allele frequency information and predictions of variant pathogenicity from 11 different variant prediction methods. Exome-sequencing data from 324 healthy adults from the Simons Simplex Collection (SSC) underwent the same process. An odds ratio comparing the proportion of SALS cases with a rare (<0.001 ExAC European MAF) and deleterious variant or pathogenic repeat expansion in at least 1 of 33 ALS-associated genes compared to SSC controls was calculated for each prediction method. 7 of these methods showed a significant excess of genetic burden in ALS genes compared to SSC controls indicating they are making accurate predictions. From these 7 methods, the average proportion of SALS cases with an identifiable genetic variant likely responsible for disease was 16.5%. This estimate, which incorporates both variant deleteriousness and frequency, is likely more accurate than previous estimates based on only variant frequency.

1396W

NOX genes - Genetic markers for severity of multiple sclerosis? *A. Gyllenberg[1], L.M. Olsson[2], B. Acharjee[1], M. Khademi[1], H. Lund[1], R. Parsa[1], F. Piehl[1], L. Alfredsson[3,4], J. Hillert[5], R. Holmdahl[2], T. Olsson[1], I. Kockum[1].* 1) Neuroimmunology Unit, Dep. of Clinical Neuroscience, Karolinska Institutet, Centre for Molecular Medicine, Karolinska University Hospital (Solna), S-171 76 Stockholm, Sweden; 2) Division of Medical Inflammation Research, Dep. of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden; 3) Cardiovascular Epidemiology, Inst. of Environmental Medicine, Karolinska Institutet, S-171 77 Stockholm, Sweden; 4) Stockholm Center for Public Health, Stockholm County Council, Stockholm, Sweden; 5) The Multiple Sclerosis Research group, Dep. of Clinical Neuroscience, Karolinska Institutet, Centre for Molecular Medicine, Karolinska University Hospital (Solna), S-171 76 Stockholm, Sweden.

Statement of purpose: The focus in this project is to analyze how NADPH oxidases (NOX enzymes), essential for the formation of reactive oxygen species (ROS), govern neuroinflammation in Multiple Sclerosis (MS [MIM 126200]). We aim to identify severity variables and investigate genetic association between NOX related genes and MS progress in relation to NOX enzyme activity. Methods used; Using genotypes from a custom made Illumina chip from 7701 MS patients and 6637 healthy controls, we analyzed 1733 Single Nucleotide Polymorphisms (SNPs) in 34 NOX-related genes and for 4 MS-associated biomarkers measured in cerebrospinal fluid (CSF). We also performed qPCR measuring the mRNA levels of gene-expression in CSF samples. Oxidative burst in peripheral blood mononuclear cells (PBMC) was measured in 156 patients with MS or other neurological disease (OND). In the analysis, we compared both MS patients and healthy controls, but also used variables developed to estimate severity or progress of disease. Summary of results; 29 SNP markers in high linkage disequilibrium in the cytochrome b-245 beta chain (*CYBB*) gene on chromosome Xp21.1, coding for a subunit of the enzyme complex NADPH oxidase, where analyzed. We found association to patients with both high MS severity scores (MSSS) and second line treatment compared to individuals with low scores (quartiles) and first line treatment (p=0.003 OR (CI):1.4 (1.1-1.7)) and with strength of burst (Mean Fluorescent Intensity, MFI) in stimulated monocytes (Beta: -0.12, p=0.01) for twenty of these markers. The associated allele correlates with lower burst and also, *CYBB* mRNA is expressed at a lower level in CSF samples from MS patients compared to OND patients. We also identified the same alleles in the *CYBB* gene to be associated with increased levels of neurofilament light (NFL) in CSF, a biomarker for neuronal damage. This association was only among MS patients and not in controls. Taken together, decreased activity of *CYBB* seem to associate with more severe MS. Our results provide preliminary evidence for the involvement of the NOX related gene *CYBB* in severity and progress of MS disease, possibly through the mechanisms of oxidative burst. Importantly, these results do not reach formal criteria for genome wide significance and need confirmation in independent cohorts. We acknowledge the International Multiple Sclerosis Genetics Consortium (IMSGC) for the genotyping. This specific project have support from EU fp7, Neurinox.

1397T

A method for the identification of variants in Alzheimer's disease candidate genes and transcripts using hybridization capture combined with long-read sequencing. *S. Kujawa[1], J. Ekholm[1], M. Jarosz[2], K. Giorda[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) Integrated DNA Technologies, Coralville, IA.

Alzheimer's disease (AD) is a devastating neurodegenerative disease that is genetically complex. Although great progress has been made in identifying fully penetrant mutations in genes such as *APP*, *PSEN1* and *PSEN2* that cause early-onset AD, these still represent a very small percentage of AD cases. Large-scale, genome-wide association studies (GWAS) have identified at least 20 additional genetic risk loci for the more common form of late-onset AD. However, the identified SNPs are typically not the actual risk variants, but are in linkage disequilibrium with the presumed causative variant (Van Cauwenberghe C, et al., The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* 2015;18:421-430).Long-read sequencing together with hybrid-capture targeting technologies provides a powerful combination to target candidate genes/transcripts of interest. Shearing the genomic DNA to ~5 kb fragments and then capturing with probes that span the whole gene(s) of interest can provide uniform coverage across the entire region, identifying variants and allowing for phasing into two haplotypes. Furthermore, capturing full-length cDNA from the same sample using the same capture probes can also provide an understanding of isoforms that are generated and allow them to be assigned to their corresponding haplotype.Here we present a method for capturing genomic DNA and cDNA from an AD sample using a panel of probes targeting approximately 20 late-onset AD candidate genes which includes *CLU, ABCA7, CD33, TREM2, TOMM40, PSEN2, APH1* and *BIN1*. By combining xGen® Lockdown® probes with SMRT® Sequencing, we provide completely sequenced candidate genes as well as their corresponding transcripts. In addition, we are also able to evaluate structural variants that due to their size, repetitive nature, or low sequence complexity have been un-sequenceable using short-read technologies.

1398F

Association of CpG-related SNPs (CGSs) with Alzheimer disease. *Y. Ma[1], K.L. Lunetta[1], G. Jun[1,2], J. Chung[1], R. Koester[3], C.C. White[4,5,6], P.L. De Jager[4,5,6], A.C. Naj[7], B.W. Kunkle[8], R. Mayeux[9], J.L. Haines[10], M.A. Pericak-Vance[8], G.D. Schellenberg[7], L.A. Farrer[1], the Alzheimer's Disease Genetics Consortium.* 1) Boston University, Boston, Massachusetts, USA; 2) Andover Production Creation Innovation System (A-PCIS), Eisai Inc., Andover, MA, USA; 3) Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA; 4) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts, USA; 5) Harvard Medical School, Boston, Massachusetts, USA; 6) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 7) University of Pennsylvania, Philadelphia, Pennsylvania, USA; 8) University of Miami, Miami, Florida, USA; 9) Columbia University, New York, New York, USA; 10) Case Western Reserve University, Cleveland, Ohio, USA.

There is growing evidence of epigenetic influences on Alzheimer disease (AD) risk, but little is known about associations with CpG-related single nucleotide polymorphisms (CGSs) that govern site-specific methylation. We conducted an AD GWAS focused specifically on CGS using 24 cohorts containing 11,956 AD cases and 12,252 controls assembled by the Alzheimer's Disease Genetics Consortium. We used a sliding-window approach with a window size of 1 kb and a sliding interval of 500 bp to capture the potential regional effects of methylation. After filtering out those CGSs with imputation quality ≤ 0.4 and with genotype information in ≤ 12 cohorts, we obtained 2,288,371 windows containing at least 2 CGSs per window. We identified genome-wide significant (GWS, $P<5\times10^{-8}$) associations with windows spanning 932 kb in the *APOE* region; in *BIN1* (best window: chr2:127,847,001-127,848,000, $P=1.3\times10^{-13}$); the intergenic region between *LRFN2* and *UNC5CL* (best window, chr6:40,825,501-40,826,501, $P=1.2\times10^{-8}$); *TREM2* (best window, chr6:41,128,501-41,129,500, $P=1.7\times10^{-8}$); *MS4A6A* (best window, chr11:59,923,001-59,924,000, $P=2.7\times10^{-10}$); *MS4A4A* (best window: chr11:60,030,001-60,031,000, $P=4.8\times10^{-11}$); and *PICALM* (best window: chr11:85,759,501-85,760,500, $P=6.3\times10^{-9}$). Associations in 13 of the 15 GWS windows were no longer significant in analyses that controlled for non-CGSs in these windows, suggesting a unique role of CGSs in AD. We investigated the functional relevance of these findings by evaluating the association of CGSs in the best window of each significant locus with methylation levels of CpG sites reported to be related with neurofibrillary tangle counts in brains of 708 random subjects from the Religious Order Study and the Memory and Aging Project. The combined effect of CGSs in the best window of *BIN1* is associated with methylation level of cg22883290 ($P=1.4\times10^{-3}$), and that in the best windows of *MS4A6A* ($P=3.9\times10^{-3}$) and *PICALM* ($P=0.02$) are associated with methylation level of cg04585722. Methylation data were unavailable to evaluate these associations for top windows in chromosome 6. Our study suggests that DNA methylation has an important role in pathways leading to AD.

1399W

Genome-wide significant associations with AD and variants in *UGT2B28*, *AXIN2*, *SYCE1*, *BCAT1*, *FAM13A*, **and** *DQ594761*: **The Alzheimer's Disease Sequencing Project Dominican families.** *A. Parrado[1], L. Henehan[2], Q. Li[3], P. Peeters[4], G. Rajagapol[1], K. Bachman[1].* 1) Discovery Sciences, Janssen Pharmceutical R&D, Spring House, PA, USA; 2) Trinity Prep High School, Winter Park, FL, USA; 3) Janssen Pharmaceutical, Titusville, NJ, USA; 4) Janssen Pharmaceutical, Beerse, Belgium.

Introduction: To identify novel genetic variants contributing to increased risk or protection against developing Alzheimer's disease (AD) among Caribbean Hispanic and Caucasian families. **Methods:** We preformed family-based association tests (FBAT) to identify an over or under-transmission of alleles among affected offspring that deviate from the expected distribution, derived using Mendel's law of segregation, using the Alzheimer's Disease Sequencing Project (ADSP – discovery phase) whole-genome sequence (WGS) cohort consisting of 84 highly informative multiplex families. We studied the WGS data of 362 AD diagnosed patients and 68 unaffected relatives, the largest proportion (n=67) of families self-reported from Dominican Hispanic ancestry (the first of the six consent groups from the ADSP Study). **Results:** We identified six variants located in the intronic or promoter region of six genes (*UGT2B28, AXIN2, SYCE1, BCAT1, FAM13A* and *DQ594761*) that achieved a genome-wide level of statistical significant association (FBAT $P < 5\times10^{-8}$) with Alzheimer's disease diagnosis. Additionally, we identified six variants in or near six genes (*POU6F1, NRG3, SLITRK1, SMIM15, AC092684.1,* and *RP11-986E7.2*) with a suggestive level of statistical significant association (FBAT $P < 1\times10^{-6}$) with AD diagnosis. **Conclusions:** Interestingly, *UGT2B28, SLITRK1, AXIN2, BCAT1,* and *NRG3* are expressed in the brain (i.e. hippocampus, glial cells, and ganglion), furthermore, *NRG3* and *SLITRK1* have been associated with neurological diseases (Schizophrenia and Tourette's, respectively). *Axin2* plays an important role in the Wnt/Hedgehog/Notch signaling pathway, in concert with *PSEN1* and *PSEN2.* Neuregulin 3 *(NRG3)* is involved in the nuclear signaling pathway through *ErbB4* with *PSEN2.* Noteworthy, similar to *APP* and *PICALM, SLITRK1* (SLIT And NTRK-Like Family, Member 1) plays a role in Axonogenesis.

1400T

Identification of amyotrophic lateral sclerosis loci using distantly related relatives in large pedigrees. *K.L. Russell[1], J.M. Downie[1], S.B. Gibson[2], K.P. Figueroa[2], M.B. Bromberg[2], S.M. Pulst[2], L.B. Jorde[1].* 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112; 2) Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT 84132.

Amyotrophic lateral sclerosis (ALS) is a fatal disease that causes both upper and lower motor neurons to progressively degenerate, culminating in paralysis and respiratory failure. Approximately 68% of familial ALS cases (i.e., those in which a first- or second-degree relative is affected) have an identifiable mutation in one of more than 30 ALS-causing genes. Among non-familial ALS cases, about 16% have a mutation in one of these genes. The purpose of this study is to identify ALS genetic risk loci by performing whole-genome sequencing and shared genomic segment analysis between pairs of ALS patients who are considered to be non-familial but are distantly related and may share disease-causing variants. The Utah Population Database (UPDB) was used to identify 35 ALS patients who were distantly related to at least one other ALS patient. UPDB identified 30 pairwise relationships among these 35 patients, which range from 6-14 degrees of relationship (median = 10). Whole-genome sequencing was performed on these 35 ALS patients to an average coverage of 60X. BWA and GATK (v3.3) were used to perform read alignment and variant calling. Genomic segments that were on average 589 kb in size and were shared between the ALS patients were identified in a pairwise manner using IBDseq. To validate our approach, we determined whether these shared genomic segments contained any of 33 genes known to be associated with ALS pathogenesis (*ANG, ATXN2, CHCHD10, CHMP2B, C9orf72, DAO, DCTN1, ELP3, ERBB4, EWSR1, FIG4, FUS, GLE1, GRN, HNRNPA1, HNRNPA2B1, MATR3, NEFH, NEK1, OPTN, PFN1, SETX, SOD1, SPAST, SQSTM1, SS18L1, TAF15, TARDBP, TBK1, TUBA4A, UBQLN2, VAPB, VCP*). Four pairs of samples shared a genomic segment that contained *SOD1, CHCHD10, CHMP2B,* or *DCTN1*. Next, we sought to determine if any nonsynonymous variants could be identified in these genes. Ensembl's Variant Effect Predictor tool was used to identify variants that result in protein coding sequence changes. A single shared nonsynonymous variant in *CHCHD10* (rs9153) was identified. The encoded protein is located in the intermembrane space of mitochondria and is hypothesized to be involved in cristae structure maintenance. These results suggest that shared genomic segment analysis may be helpful in identifying both known and novel ALS risk loci. Future efforts will be made to identify specific variants in these shared genomic segments that may be responsible for ALS pathogenesis.

1401F

Rescue of locomotion defect in Drosophila model of spastin hereditary spastic paraplegia. *F.V. Bolduc[1,2,3], C. Rosenfelt[1], J.A. Parker[4], P. Drapeau[4], G.A. Rouleau[5].* 1) Dept Pediatrics, Univ Alberta, Edmonton, Alberta, Canada; 2) Neuroscience and Mental Health Institute, University of Alberta; 3) Dept Medical Genetics, Univ Alberta, Edmonton, Alberta, Canada; 4) CRCHUM and Department of Neuroscience, Université de Montréal, Montréal, Québec, Canada; 5) Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada.

Background. Hereditary spastic paraplegia (HSP) is associated with progressive locomotion defects in human. Several causative genes have been identified but there remain no clinical treatment available. *SPASTIN* is the most common gene mutated in autosomal dominant HSP. We recently identified a role for oxidative stress in *SPASTIN* homologues mutants using various model organisms. Furthermore, we were able to show that pharmacological treatments aimed at reducing oxidative stress (methylene blue, n-acetylcysteine and phenazine). In all three models, improvement in locomotion was observed. Furthermore, cellular markers of oxidative stress were reduced closer to control level. Here we tested another potential treatment for HSP not tested in the Drosophila model: guanabenz. Methods. Transgenic Drosophila expressing spastin RNAi with the pan neuronal driver ELAV were used. Groups of 20 flies are used for each trial. Flies are incubated for 21 hours at 22C before being inserted into a glass cylinder for the assay. Flies are then tapped at the bottom of the cylinder and allowed to climb. The percentage of flies climbing above the 190mL line is recorded every 10 seconds for 2 minutes. For drug treatment, flies were transferred to drug/vehicle 24 hours prior to testing. Statistical analysis is performed using Tukey test to compare each group (untreated and treated for each genotype). Results. We observed a significant improvement (P=0.0091) in the percentage of ELAV-spas RNAi flies climbing above the target line after treatment with 300uM guanabenz. Conclusions. HSP is a devastating disorder associated with progressive loss of mobility and spasticity that remains without treatment. These results combined with our previously published data provide strong pre-clinical support for the use of antioxidants in HSP.

1402W

Genetic variants in the *SNCA* and alpha-synuclein immunostaining in the gastric and colonic mucosa in Parkinson's disease. *S. Chung[1], K. Lohmann[2], J. Kim[3], H. Ryu[1], F. Hinrichs[2], H. Lee[4], K. Kim[1], J. Lee[4], K. Jung[4], M. Kim[5], Y. Kim[1], S. Yun[6], J. Lee[7], S. Hong[8], S. Myung[4], C. Klein[2].* 1) Department of Neurology, Asan Medical Center, Seoul; 2) Institute of Neurogenetics, University of Luebeck, Luebeck; 3) Department of Neurology, Metro hospital, Anyang; 4) Department of Gastroenterology, Asan Medical Center, Seoul; 5) Health Screening and Promotion Center, Asan Medical Center, Seoul; 6) Division of Biostatistics, Center for Medical Research and Information, Asan Medical Center, Seoul; 7) Asan Institute for Life Sciences, Asan Medical Center, Seoul; 8) Department of Pathology, Asan Medical Center, Seoul.

Background: Alpha-synuclein (a-Syn) deposition in the enteric nervous system (ENS) has been reported in patients with Parkinson's disease (PD). However, enteric α-Syn has also been reported in healthy individuals. The determining factors for the a-Syn deposition in the gastric and colonic mucosa of human are unclear. **Objectives:** We aimed to investigate the association between genetic variants in the *SNCA* and a-Syn immunostaining in the ENS in PD patients and controls. **Methods:** Study subjects consisted of 38 patients with PD and 53 controls. The a-Syn immunohistochemistry was performed for gastric and colonic mucosal tissues that was obtained by endoscopic gastroduodenoscopy and colonoscopy. The *SNCA* SNP rs11931074 and REP1 variants were genotyped in all subjects. **Results:** The positivity of a-Syn immunostaining was not different between PD patients (31.6% for stomach and 10.4% for colon) and controls (33.3% for stomach and 18.5% for colon). In comparisons between subjects with positive a-Syn immunostaining and those without, *SNCA* SNP rs11931074 (G allele) were significantly associated with positive a-Syn immunostaining for the stomach and colon mucosa in analyses for PD patients only (OR = 5.96, 95% CI = 1.70 – 20.97, P = 0.005, additive model) and in analyses for all subjects (PD patients + controls) (OR = 1.98, 95% CI = 1.02 – 3.82, P = 0.043, additive model). The longer *SNCA* REP1 alleles are associated with positive a-Syn immunostaining for the stomach and colon in all subjects (PD patients + controls) (OR = 4.16; 95% CI = 1.27 – 13.67; P = 0.019). **Conclusion:** Genetic variants in the *SNCA* may have an effect on a-Syn deposition in the gastric and colonic mucosa in PD patients as well as in healthy individuals.

1403T

***NUBPL* mutations link Parkinson's disease and other movement disorders to recessive Complex I deficiency.** *P.S. Eis[1], B. Schüle[2], S. Kim[2], J.W. Langston[2], V.E. Kimonis[3], E. Hatchwell[1].* 1) Population Diagnostics, Inc., Melville, NY; 2) The Parkinson's Institute and Clinical Center, Sunnyvale, CA; 3) Division of Genetics and Metabolism, Department of Pediatrics, University of California, Irvine, Children's Hospital of Orange County, Orange, CA.

Complex I Deficiency (CID) is an autosomal recessive disorder caused by homozygous or compound heterozygous mutations in one of at least 23 nuclear-encoded genes, which include Complex I (CI) subunits and assembly factors (MIM 252010). Parkinson's Disease (PD) pathology in many patients includes deficiency in CI activity (first reported in 1989). In a genome-wide PD gene discovery study, we performed array comparative genomic hybridization (aCGH) on a PD cohort of 468 patients to identify copy number variants (CNVs) impacting genes potentially causative of PD. This genome-wide screen led to the identification of one patient with a large, complex chromosomal rearrangement that was an apparent loss of function mutation impacting CI assembly factor *NUBPL* (nucleotide binding protein-like), an Fe-S protein. Mutations in *NUBPL* (MIM 613621) were identified as a cause of CID in 2010, and intriguingly, the first patient was reported to have a CNV identical to that found in our PD patient on one allele, and a splicing mutation (c.815-27T>C) on the other allele. Given the strong biology linking CI activity and PD, we sequenced *NUBPL* in the PD cohort (n=477) and found three novel SNVs that may be causing or contributing to PD. We also observed three patients that are heterozygous for the splicing mutation (c.815-27T>C). Subsequent identification of two new patients with *NUBPL* CID revealed a family history, in older relatives, of PD, essential tremor (ET), or restless legs syndrome (RLS), which further supports the role of this CI assembly factor in late-onset movement disorders. Analogous to mutations in *GBA* (MIM 606463) linking autosomal recessive Gaucher disease to an increased risk of PD in *GBA* mutation carriers, we hypothesize that heterozygous *NUBPL* mutation carriers have an increased risk for late-onset movement disorders that include PD, ET, and RLS.

1404F

Assessing the contribution of common polygenic risk and rare variation in 1,975 migraine families from Finland. *P. Gormley[1,2,7], M.I. Kurki[1-3,7], A. Mitchell[4], M. Hiekkala[5], P. Häppölä[3], D. Lal[1,2], I. Surakka[3], P. Palta[3], G. Srivastava[4], M. Kaunisto[3], J. Rämö[1-3], E. Hämäläinen[3], M.J. Daly[1,2], S. Ripatti[3], M. Kallela[6], H. Runz[4], B.M. Neale[1,2], M. Wessman[3,5], A. Palotie[1-3].* 1) Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) Department of Genetics & Pharmacogenomics, Merck Research Laboratories, Merck and Co., Inc., Boston, MA; 5) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 6) Helsinki University Central Hospital, Helsinki, Finland; 7) These authors contributed equally.

Migraine is a complex neurological disease affecting almost one billion people worldwide (PMID:26063472). It is known to aggregate in families and twin studies have established a strong genetic component (PMID:25985137). Disease presentation ranges from the rare (and severe) hemiplegic migraine (HM) to the more common forms; migraine with aura (MA) and migraine without aura (MO). Likewise, the spectrum of genetic variation conferring risk ranges from three Mendelian gene variants known for HM (*CACNA1A, ATP1A2,* and *SCN1A*) to the 38 common variant loci identified by genome-wide association studies (Gormley et al., *Nat. Genet.*, In press). We investigate the contribution of common polygenic and rare variation to migraine in 1,975 families from Finland, including 563 HM, 2,649 MA, 2,622 MO, and 2,813 unaffected members. It is the largest collection of migraine families to our knowledge. We genotyped 8,992 family members on either the Illumina Core-Exome or PsychArray. We compared the within-family imputation accuracy by three approaches; 1) a Finnish reference panel of 1,941 individuals whole-genome sequenced (WGS) at low coverage (4.6x); 2) a joint Finnish + 1000 Genomes panel (2,504 individuals at 7.4x WGS + 65.7x exome); and 3) a family-based imputation using high coverage (30x) WGS from 631 members of the same families. Both population panels accurately imputed allele frequencies down to 0.1% and while sensitivity was 4% higher for the joint panel, the Finnish-specific panel introduced only a fraction (27.5%) of the false positives imputed by the joint panel. In the family-based approach, imputation was accurate for individuals with two sequenced parents and one sibling (95–99% of true variants were imputed with 0.2% discordance). However, much lower accuracy was observed for those more distantly related to sequenced individuals (e.g. with two sequenced grandparents only 30% of variants were imputed with 10% discordance). To assess the relative contribution of common polygenic and rare variation, we calculated polygenic risk scores (PRS) for each individual using effect sizes from a recent migraine GWAS (*N*=375,000). We identified 442 families (22.4%) with a median PRS in the highest quartile of the PRS distribution (i.e. migraine risk was mostly polygenic) and 383 families (19.4%) in the lowest quartile where rare variation likely contributes more. Future work will assess rare variants aggregating in these low PRS families by association and linkage analyses.

1405W

Whole-exome sequencing identifies *FIG4* as a risk gene for ALS in a central European cohort. *A. Osmanovic[1], I. Rangnau[1], A. Kosfeld[1], J.H. Martens[1], S. Abdulla[2], C. Janssen[2], B. Auber[1], P. Raab[3], M. Preller[4], S. Petri[2], R.G. Weber[1].* 1) Hannover Medical School, Department of Human Genetics, Hannover, Germany; 2) Hannover Medical School, Department of Neurology, Hannover, Germany; 3) Hannover Medical School, Department of Neuroradiology, Hannover, Germany; 4) Hannover Medical School, Institute for Biophysical Chemistry and Center for Structural Systems Biology, German Electron Synchrotron (DESY), Hamburg, Germany.

We aimed to identify the genetic cause of the devastating neurodegenerative disease amyotrophic lateral sclerosis (ALS) in a German family with two affected individuals, and to assess the prevalence of mutations in the identified risk gene, *FIG4*, in a central European ALS cohort. Therefore, whole-exome sequencing (WES) and an overlapping data analysis strategy were performed in an ALS family with autosomal dominant inheritance and incomplete penetrance. Additionally, 200 central European ALS patients were analyzed using whole-exome or targeted sequencing. All patients were subjected to clinical, electrophysiological, and neuroradiological characterization to explore genotype-phenotype relationships. WES analysis of the ALS family identified the rare heterozygous frameshift mutation *FIG4*,c.759delG;p.(F254Sfs*8) predicted to delete the catalytic domain and active center from the encoded phosphoinositide 5-phosphatase with a key role in endosomal vesicle trafficking. Additionally, novel or rare heterozygous *FIG4* missense mutations predicted to be deleterious were detected in five sporadic ALS patients revealing an overall *FIG4* mutation frequency of 3% in our cohort. Four of six variants identified were previously associated with ALS or the motor and sensory neuropathy Charcot-Marie-Tooth disease type 4J (CMT4J), while two variants were novel. In *FIG4* mutation carriers, disease duration was longer and upper motor neuron predominance was significantly more frequent compared to ALS patients without *FIG4* mutation. In summary, our study confirms *FIG4* as an ALS risk gene in a central European cohort, adds new variants to the mutational spectrum, links ALS to CMT4J on a genetic level, and describes a distinctive ALS phenotype for *FIG4* mutation carriers.

1406T

Targeted high-throughput screening of olivocerebellar motor circuitry genes in essential tremor. *J-F. Schmouth[1], C. Leblond[2], G. Houle[2], S. Beatrice-Laurent[1], C. Bourrassa[1], C. Vilarino-Guell[3], A. Rajput[4], P. Dion[1], G. Rouleau[1].* 1) Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada; 2) Human Genetics, McGill University, Montréal, Quebec, Canada; 3) Medical Genetics, University of British Columbia, Vancouver, British-Columbia, Canada; 4) University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

   Background: Essential Tremor (ET) is a prevalent neurological disorder of unknown etiology that is characterized by the presence of action tremors occurring during voluntary motion. The worldwide prevalence of ET is 0.9% in the general population and 4.6% in individuals ≥65 years old. Twin studies are suggesting that genetic risk factors likely contribute to the pathology. To date, no gene was reproducibly reported to cause ET across unrelated cohorts which is likely due to the genetic heterogeneity of the disorder. Mounting evidence currently suggests dysfunctions affecting the olivocerebellar motor circuitry as the primary source of tremors; with the inferior olive, Purkinje cells and deep cerebellar nuclei being the main affected brain regions. **Hypothesis:** Genes that are phylogenetically conserved for their expressions in the olivocerebellar circuitry are susceptible to present genetic variations that would be ET risk factors. **Methods:** A preliminary list of 11 candidate genes that were selected based on their high level of expression in the olivocerebellar circuitry in both human and mouse has been generated. The 11 genes from this list, which noticeably appear to encode several proteins involved in calcium and glutamate signaling pathways, will be screened for the presence of genetic variations in a cohort of cases and controls. Calcium related genes are plausible candidates since previous evidence supports calcium dependent pathways to play a role in the generation of neuronal oscillations, resonance and pacemaker activities. Glutamate signaling related genes are also good candidates since NMDAR agonists are known to alleviate harmaline induced tremors. The mutation screen was done using a Molecular Inversion Probes (MIPs) capture methodology. For this particular project, a total of 991 MIPs were designed to screen 269 cases and 288 matching control individuals. The primary focus of our genetic analysis will be on the identification of nonsense and missense mutations. Further examination will use a statistical burden method that will collapses all the rare variants located in the different candidate genes into a single value, which is then tested for association with the phenotype. **Outcome:** The list of 11 candidates contains genes functionally related to three defined pathways. Hence, if proven true, this approach has not only the potential in identifying novel ET risk factors but could also link this pathology to a particular signaling pathway.

1407F

Linkage analysis on the 20 consanguineous singleton families with multiple system atrophy. *T. Yasuda, T. Matsukawa, H. Ishiura, J. Mitsui, S. Tsuji.* Neurology, The University of Tokyo, Bunkyo-ku, Tokyo, Japan.

   Background & Purpose: Although MSA is mostly a sporadic disease, mutations in *COQ2* with an autosomal recessive inheritance was associated with MSA in rare familial cases. Assuming there may be other causative genes with autosomal recessive inheritance, we focused on patients with MSA born to consanguineous parents. We conducted a parametric linkage analysis with an autosomal recessive inheritance model. **Methods:** We analyzed 20 Japanese MSA patients born to consanguineous parents. The consanguineous parents were all first cousins. We calculated the multipoint parametric LOD scores based on SNP typing. Unaffected siblings of the MSA patients, if available, were also included in the analysis. We used Genome-Wide Human SNP Array 6.0 (Affymetrix) for SNP typing. LOD scores were calculated using SNP HiTLink and Allegro 2.0 under an autosomal recessive model (disease allele frequency of 10^{-3} and a penetrance of 1.0). **Result:** In these 20 families, the onset ages ranged from 35 to 69 years old, including 11 female and 9 male cases. These cases included 14 MSA-C, 4 MSA-P and 2 unclassifiable cases. In these cases, one MSA-C patient carried a homozygous V393A mutation in *COQ2*. Cumulative multipoint parametric LOD scores did not reveal loci with positive LOD scores, suggesting genetic heterogeneity. There were multiple loci giving positive HLOD scores exceeding 1.0. **Conclusion:** Presence of homozygous V393A in one patient further supports our previous findings that *COQ2* variants are associated with MSA. Further detailed linkage analyses are required to dissect genetic loci involved in MSA. In particular, detailed linkage analyses of the MSA patients born to consanguineous parents should facilitate the search for genes involved in MSA.

1408W

Genome wide association study of cisplatin-induced peripheral neuropathy (CIPN) in testicular cancer survivors. *O. El-Charif[1], H.E. Wheeler[2], B. Mapes[1], T. Mushiroda[3], M. Kubo[3], E. Gamazon[4], D.R. Feldman[5], R.J. Hamilton[6], D.J. Vaughn[7], C.J. Beard[8], C. Fung[9], S.D. Fossa[10], C. Albany[11], N.J. Cox[4], L.H. Einhorn[11], L.B. Travis[11], M.E. Dolan[1].* 1) Medicine, Univ Chicago, Chicago, IL; 2) Departments of Biology and Computer Science, Loyola University, Chicago, IL; 3) RIKEN Center for Integrative Medical Science, Yokohama, Japan; 4) Division of Genetic Medicine, Vanderbilt University, Nashville, TN; 5) Department of Medical Oncology, Memorial Sloan-Kettering Cancer Center, Memorial Sloan-Kettering, New York, NY; 6) Department of Surgical Oncology, Princess Margaret Cancer Centre, Princess Margaret Cancer Centre, Toronto, ON; 7) Department of Medicine University of Pennsylvania, Philadelphia, PA; 8) Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA; 9) J.P. Wilmot Cancer Institute, University of Rochester Medical Center University of Rochester Medical Center, Rochester, NY; 10) Department of Oncology, Oslo University Hospital, Radiumhospital, Oslo Norway; 11) Department of Medical Oncology Indiana University, Indianapolis, IN.

Chemotherapy-induced peripheral neuropathy, one of the most common and potentially permanent side effects of modern chemotherapy, can adversely affect survivors' quality of life due to the lack of treatment that can effectively prevent or mitigate it. Although cisplatin-based therapy is curative for testicular cancer, this drug is associated with multiple long-term serious toxicities, including peripheral neuropathy. We performed a genome-wide analysis of neuropathy in a cohort of 847 testicular cancer survivors using EORTC-CIPN20, a validated patient-reported questionnaire. Median age at diagnosis was 30 ± 9 (SD) years and median cumulative dose was 400 mg/m². The most clinically relevant symptoms (tingling, numbness and pain) were common (55% having any). The subscale from these items indicated excellent internal consistency (Cronbach $\alpha = 0.88$) and was subsequently used to model CIPN. Patients were placed into 3 severity groups: 45% none, 39% a little, and 16% quite a bit/very much. 5.1 million SNPs passed quality control in 680 genetically European individuals; SNP dosages and covariates including age, dose and 10 genetic principal components were regressed on this phenotype using an additive ordinal model. SNPs within several genes involved in cytochrome C-dependent apoptosis, neural excitability, intracellular trafficking, and vascularization were identified ($p < 5 \times 10^{-5}$). Chip heritability analysis revealed that CIPN likely has a polygenic architecture ($h^2 = 0.70 \pm 0.49$, $P = 0.049$). An eQTL for *GSTT1*, a gene known to play a role in cisplatin detoxification was also identified ($p = 1.8 \times 10^{-5}$). We performed permutation-resampling analysis to identify enrichment of candidate genes involved in cisplatin metabolism, i.e., Mendelian Charcot-Marie-Tooth (CMT) genes, and SNPs associated with preclinical cellular sensitivity to cisplatin. Enrichment of CMT genes was significant across a range of p-values (0.005-0.05, $\alpha = 0.10$) and implicated the gene *FGD4*, previously identified in a GWAS of paclitaxel-induced neuropathy. Further studies will focus on replication of findings in an independent cohort and functional validation of the implicated pathways.

1409T

The genetics of levodopa-induced dyskinesia in Parkinson's disease. *O.V. Sazonova, I. Hallgrímsdóttir, J. Shelton, P. Cannon. Research, 23andMe, Mountain View, CA.*

Parkinson's disease (PD) is a progressive movement disorder that afflicts approximately one percent of the population. PD is caused in part by the progressive destruction of dopamine-producing neurons of the substantia nigra, precipitating the onset of a wide range of symptoms. Levodopa, a compound converted into dopamine in the brain, is the most effective therapy for treating the stiffness, tremors, spasms, and poor muscle control of PD. However, the majority of patients using levodopa-containing therapies go on to develop levodopa-induced dyskinesia (LID), a disorder of involuntary, jerky movements. Many discontinue the therapy or reduce drug dosage due to LID severity, all to the detriment of managing their PD symptoms. Recent candidate gene studies have demonstrated a genetic predisposition for LID involving a number of genes that play a role in neurotransmission. 23andMe has a large research database of genetic and phenotypic information collected from individuals diagnosed with PD, with over 12,200 genotyped individuals who have consented to research. We hypothesized that there are genetic markers of LID that can be discovered through genotyping and self-reported dyskinesia among individuals on levodopa therapies. To test this hypothesis, we developed a new survey to define binary and quantitative phenotypes for individuals who have altered or discontinued their levodopa regimens due to dyskinesia symptoms. By capturing timelines of PD onset, the start of levodopa treatment, the start of LID, and the disruption of treatment due to LID, the data enabled genome-wide association studies (GWAS) of LID onset and severity. We present these results along with their comparison to previous candidate gene studies, thereby deepening our understanding of the role of genetic variation in levodopa-induced dyskinesia. Leveraging the unique resources of 23andMe's PD cohort, the present study is the first genome-wide investigation of side effects associated with levodopa therapy. Our findings offer important insight into the complex mechanisms of this disorder, and lay the groundwork for developing new therapies to reduce LID and for personalizing PD treatment based on patient genotype.

1410F

Impaired mitochondrial function and dynamics in the pathogenesis of FXTAS. *M.I. Alvarez-Mora[1], L. Rodriguez-Revenga[1], I. Madrigal[1], M. Guitart-Mampel[2], G. Garrabou[2], M. Milà[1].* 1) Biochemistry and Molecular Genetics, Hospital Clinic and CIBERER, Barcelona, Spain; 2) Muscle Research and Mitochondrial Function Laboratory, CELLEX, Faculty of Medicine-University of Barcelona, Internal Medicine Department, Hospital Clinic and CIBERER, Barcelona, Spain.

Mitochondrial involvement plays an important role in neurodegenerative diseases. At least one-third of adult carriers of a *FMR1* premutation (55-200 CGG repeats) are at risk of presenting an adult-onset neurodegenerative disorder known as fragile X-associated tremor/ataxia syndrome (FXTAS). In an attempt to provide new insights into the mechanisms involved in the pathogenesis of FXTAS, we characterized mitochondrial function and dynamics by the assessment of oxidative respiratory chain function, mitochondrial content, oxidative stress levels and mitochondrial network complexity. Regarding mitochondrial function we found that mitochondrial respiratory capacity is compromised in skin fibroblasts whereas in blood no differences were observed between the FXTAS and control groups. Furthermore, fibroblasts from FXTAS patients presented altered mitochondrial architecture, with more circular and less interconnected mitochondria being observed. Mitochondrial function and dynamics deregulation, characteristic of neurological disorders, are present in FXTAS patients. These features might be limiting temporal and spatial bioenergetics cells supply and thus contributing to disease pathogenesis.

1411W

Generation of disease-specific autopsy-confirmed iPSCs lines from post-mortem isolated peripheral blood mononuclear cells. *A. Ali[1], K. Belle[1,2], F. Shabazz[2], K. Nuytemans[2], D. Davis[3], J. Young[1,2], W. Scott[1,2], D. Mash[3,4], J. Vance[1,2], D. Dykxhoorn[1,2].* 1) University of Miami, Miami, FL., Select a Country; 2) John P. Hussman Institute for Human Genomics,; 3) Department of Neurology; 4) Department of Molecular and Cellular Pharmacology,.

Understanding the molecular mechanisms that underlie neurodegenerative disorders has been hampered by a lack of readily available model systems that replicate the complexity of the human disease. Recent advances in stem cell technology have facilitated the derivation of patient-specific stem cells from a variety of differentiated cell types. Induced pluripotent stem cells (iPSCs) are an attractive disease model since they can be grown and differentiated to produce large numbers of disease-relevant cell types. However, most iPSC lines are derived in advance of, and without the benefit of, neuropathological confirmation of the donor – the gold standard for many disease classifications and measurement of disease severity. While others have reported the generation of autopsy-confirmed iPSC lines from patient explants, these methods require outgrowth of cadaver tissue, which require additional time and is often only successful ~50% of the time. Here we report the rapid generation of autopsy-confirmed iPSC lines from peripheral blood mononuclear cells (PBMCs) drawn postmortem. Since this approach doesn't require the propagation of previously frozen cadaver tissue, iPSC can be rapidly and efficiently produced from patients with autopsy-confirmed pathology. These matched iPSC-derived patient-specific neurons and postmortem brain tissue will support studies of specific mechanisms that drive the pathogenesis of neurodegenerative diseases.

1412T

BLM DNA helicase protects against (CAG)•(CTG) repeat instability.

M.H.M. Schmidt[1,2], G.B. Panigrahi[1], S. Lanni[1], M. Kekis[3], T. Ezzy[1,4], S. Sample[5], C.E. Pearson[1,2]. 1) Genetics & Genome Biology, The Hospital for Sick Children, Toronto, Canada; 2) Molecular Genetics, The University of Toronto, Toronto, Canada; 3) Cytogenetics & Molecular Genetics, Emory University, Atlanta, GA; 4) Molecular Biology & Biotechnology, University of Toronto at Scarborough, Toronto, Canada; 5) Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Canada.

Trinucleotide repeat expansions are associated with 43 neuromuscular and neurodegenerative disorders. Ongoing repeat expansions drive disease progression, and are caused by aberrant DNA repair. The Bloom (BLM) protein regulates sister chromatid exchange with it's DNA unwinding (helicase) and DNA annealing activities, critical during DNA replication and DNA repair. BLM's branch migration activity at three- and four-way DNA junctions suggests a possible role for BLM in unwinding unusual DNA structures such as slipped-DNAs, which we have previously shown to arise at expanded (CAG)•(CTG) repeat tracts of the myotonic dystrophy locus in patient tissues. Here we examined the replication products of (CAG)•(CTG) repeat tracts in the presence or absence of BLM. In the absence of BLM, we detect the formation of slipped-DNA structures that are indicative of repeat instability (expansions and contraction events). These structures formed only when (CTG) was used as the lagging strand template for DNA replication, where repeats tended to contract. Inclusion of functional BLM eliminated the accumulation of slipped-DNAs. Point mutations that inactivate the DNA unwinding helicase or DNA annealing domains of BLM revealed that the helicase domain was necessary to avoid slipped-DNA formation. Further, we show that BLM helicase activity, but not BLM annealing activity, is required for *in vitro* repair of DNA containing a single (CTG) repeat unit slip-out. The BLM helicase activity is not required for the repair of a longer slip-out of 20 repeats, suggesting that BLM protects against small changes of repeat lengths. Together, these results support a new model through which disease-associate trinucleotide repeat are protected from contractions by the BLM helicase.

1413F

Repetitive and restrictive behaviors associate with the prohibitin gene in autism.

R. Cantor-Chiu[1], L. Navarro[1], D. Geschwind[2], J. Lowe[2]. 1) Dept Human Gen, UCLA Sch Med, Los Angeles, CA; 2) Dept Neurology, UCLA Sch Med, Los Angeles, CA.

Autism Spectrum Disorders (ASDs) are behaviorally defined phenotypes that manifest prior to age 3 as deficits in communication skills and social interactions, including restricted and repetitive behaviors (RRBs). A salient conclusion of current genetic studies is that ASDs are polygenic and genetically complex, with as many as 1,000 predisposing genes. We hypothesized that focusing analyses on RRBs, a feature of ASDs, might reduce genetic heterogeneity and complexity. We mined Caucasian GWAS data from two of the largest ASD family cohorts. RRBs were measured in the Autism Genetics Resource Exchange (AGRE) and the Autism Genome Project (AGP) families using the Autism Diagnostic Inventory –Revised (ADIR), an interview administered to the parents of potential ASD cases. In the full AGRE sample, seven of twelve RRBs measured were significantly familial. The seven RRBs were tested for association, genome-wide, in the 1592 ASD affected children from 846 AGRE families and the 1108 ASD affected children from 1092 AGP families that were genotyped. Imputation resulted in 5,755,879 AGRE and 3,972,813 AGP SNPs for association analyses with the seven RRBs. GWAS were conducted using the EMMAX software to account for family structure. A stringent p-value, correcting for testing multiple traits, was employed (p<7.1E-9). Replication was set at p<0.01. In the AGP sample, significant association (p<6.8E-9) was seen with rs2898883 within an intron of the Prohibitin (PHB) gene at 17p21 for the RRB described as the repetitive use of objects or interest in parts of objects. Association was seen with multiple independent PHB SNPs for the same trait in the AGRE sample (p<0.01). Remarkably, 17p21 previously showed linkage and its replication in earlier analyses of ASD in the AGRE sample. PHB, a regulatory gene for cell growth that inhibits DNA synthesis, has not previously been associated with ASD. These analyses identify a novel risk gene for a behavior used to define ASD and support the hypothesis that reducing phenotypic complexity can enhance the power to identify novel risk genes for complex traits.

1414W

De novo loss of function mutations in *KMT2A* is a risk factor for autism spectrum disorder and other neurodevelopmental phenotypes. *A. Chan[1,2], M. Uddin[1], R. Yuen[1], T. Higginbotham[1], M. Zarrei[1], S. Walker[1], S.W. Scherer[1,2,3].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, On, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, On, Canada; 3) McLaughlin Centre, University of Toronto, Toronto, On, Canada.

De novo loss of function (LoF) mutations in *KMT2A* have been associated with Wiedemann-Steiner Syndrome (WSS), which is characterized by hypertrichosis cubiti, dysmorphic facial features, short stature, developmental delay, and intellectual disability. Recently, whole-exome sequencing studies have found de novo LoF mutations in some individuals with neurodevelopmental disorders (NDDs), including autism spectrum disorder (ASD). *KMT2A* encodes a histone methyltransferase that is required for neurogenesis, early development and hematopoiesis. Here, we present the clinical and genomic characterization of two unrelated individuals with ASD, who possess de novo frameshift mutations in *KMT2A* identified by whole-genome sequencing. By performing a meta-analysis on WSS cohorts (n=16), ASD (n=5615) and other NDD cohorts (n=1797), and control trios (n=2169), we found that 0.32% (24/7412 cases) of individuals with NDDs have de novo LoF mutations in *KMT2A*, while no de novo LoF mutations were found in controls. These mutations are significantly enriched in individuals with NDDs compared to unaffected control trios (p=0.0021), which collectively indicate that these mutations are highly penetrant for NDDs and NDD-associated conditions. Collectively, de novo LoF mutations were found in eight individuals with ASD, suggesting that *KMT2A* is a novel ASD risk gene. De novo LoF mutations in *KMT2A* were also present in individuals with other NDD phenotypes, such as developmental delay and intellectual disability. This finding supports the notion that *KMT2A* mutations possess a shared genetic risk affecting common genes and pathways for different NDDs. The genetic and clinical findings in this study establish a role of *KMT2A* in ASD and neurodevelopmental disorders, which will aid clinical interpretation of *KMT2A* mutations. These findings also contribute to the accumulating evidence of the role of chromatin-associated genes in ASD and other NDDs.

1415T

Properties of pathogenic *de novo* missense mutations guide discovery of new neurodevelopmental disease risk genes. *M. Geisheker[1], B. Coe[1], T. Wang[2], H. Stessman[1], T. Turner[1], E. Eichler[1,3].* 1) University of Washington School of Medicine, Seattle, WA; 2) The State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 3) Howard Hughes Medical Institute, Seattle, WA.

De novo missense mutations contribute to more cases of autism spectrum disorders than likely gene-disruptive mutations but they have been understudied because an estimated 87% of them are incidental based on studies of simplex autism. We created a database from 23 published studies of genetically-overlapping neurodevelopmental disorders (NDDs) and compared the properties of 6,451 de novo missense mutations in 9,267 cases with 1,514 in 2,243 controls. We find a significantly elevated rate of de novo missense mutations in cases. De novo missense mutations in cases are predicted to be more deleterious as their distribution shows a significant skew towards higher CADD scores (P = 3.364×10^{-6}, Wilcoxon rank-sum test). Further, the likelihood ratio for a mutation increases with CADD scores, reaching 1.5 at a CADD score of 29. Out of 4,620 genes with de novo missense mutations in cases, 1,126 (24.3%) bear mutations in two or more unrelated individuals. In 10,000 down-sampling permutations matching the number of de novo missense mutations in cases with the number in controls, the average number of genes hit recurrently in cases was greater than that seen in controls; significance was reached at three or more mutations per gene (simulated P = 0.002). Thirty-four genes had significantly more de novo missense mutations in cases than expected, and 12 of them have already been implicated as NDD risk genes. Strikingly, 28 amino acid positions in 25 genes were hit by two or more rare de novo missense mutations with a CADD score over 20 in unrelated cases; this was not observed in controls. Targeted sequencing of these recurrently mutated amino acids and other clusters of mutations in 9,126 additional cases and 2,599 controls identified 54 more mutations in cases and 5 in controls. Eleven of these mutations map to nine of the 23 targeted amino acids; four establish two new sites of recurrence. Some of these newly found mutations establish significance in genes associated with NDDs, such as *CLCN4* and *ALG13*, and genes critical for neuronal function, such as *GRIA1*. Preliminary phenotypic and functional data suggest syndromic effects of some of these recurrent mutations and strengthen their association with disease. Our analysis reveals that clustered, recurrent and severe de novo mutations provide a useful method to identify new causal genes, expanding our understanding of the pathology of NDDs.

1416F

Noncoding autism spectrum disorder risk variants identified by whole genome sequencing in extended multiplex families. *J.R. Gilbert[1,2], A.J. Griswold[1], H.N. Cukier[1], D. Van Booven[1], P.L. Whitehead[1], N.K. Hofmann[1], E.R. Martin[1,2], M.L. Cuccaro[1,2], J.P. Hussman[3], M.A Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami Miller School of Medicine, Miami, FL; 2) John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School Medicine, Miami, FL; 3) Hussman Institute for Autism, Baltimore, MD.

Autism spectrum disorder (ASD) is a neurodevelopmental condition with a strong, but heterogeneous, genetic etiology with more than 100 genes and genetic loci implicated. Recent massively parallel sequencing efforts in ASD have focused on whole exome (WES) or whole genome sequencing (WGS) in cohorts of simplex trios and the identification of *de novo* loss of function coding variants. While these are strong risk factors for ASD, we hypothesize that there are other factors contributing to the underlying genetic risk including poorly covered coding regions, functional and regulatory variants in the 98% of the genome that is noncoding, and unidentified genomic structural variation. To identify such factors, our study applied WGS to extended, multiplex families with at least one pair of cousins with ASD. We performed WGS on two or more ASD individuals across six families (15 total) where compelling candidate coding variants had not been identified by WES (Cukier HN, 2014). We sequenced each genome to 40X and analyzed with a GATK best practice pipeline and structural variants (SVs) called with the SWAN algorithm. Annotations were applied with ANNOVAR including functional predictions for noncoding variants (GWAVA, CADD, FATHMM-MKL, and Eigen). Because each genome contained more than ~4 million single nucleotide variants and indels and more than 100 SVs, we refined our regions of interest by calculating identity by descent (IBD) with existing genotyping data and filtering IBD alterations for each family. Variants were prioritized by 1) sharing in all ASD individuals per family, 2) low frequency in the population (<1%), and 3) functionality based on computational predictions. Among coding variants poorly covered by previous WES, a family carried a rare missense mutation in the neurogenesis growth factor *GDF11* and another family a frameshift in the first exon of the axonal development gene *SLAIN1*. Furthermore, WGS identified noncoding variants including two in the promoter of the ASD candidate gene *CNTN4* in one family and two upstream of potassium channel *KCTD1* in a different family. Rare SVs in single families were identified including one disrupting the promoter of the neurodevelopmental *WWOX* gene and another deleting an exon of the lincRNA *FIRRE* involved in chromosomal organization. By studying these unique pedigrees, we establish that WGS in extended families can be used to identify ASD risk alterations beyond *de novo* protein coding variants.

1417W

Increased burden of deleterious variants in essential genes in autism spectrum disorder. *X. Ji[1,2], R. Kember[2], C. Brown[2], M. Bucan[2,3].* 1) Genomics and Computational Biology Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Autism spectrum disorder (ASD) is a heterogeneous, highly heritable neurodevelopmental syndrome characterized by impaired social interaction, communication and repetitive behavior. The highly polygenic nature of ASD suggests that the analysis of the full spectrum of sequence variants, including those with small effects, will be necessary for deeper understanding of disrupted neuronal function. Human orthologs of genes with an essential role in pre- and postnatal development in the mouse ("essential genes") are enriched for human disease genes and under strong purifying selection relative to human orthologs of mouse genes with known non-lethal phenotype ("non-essential genes"). This intolerance to deleterious mutations, coupled with commonly observed haploinsufficiency of essential genes (EGs), suggests a possible cumulative effect of a large number of small-effect variants in essential genes on a disease phenotype such as ASD. We hypothesized that a cumulative effect of deleterious variants in EGs, within the same pathway or across pathways, may underlie impaired brain development and individual's ASD risk. To test this hypothesis, we identified the most comprehensive list of 3,915 EGs established to date by combining human orthologs of EGs in the mouse and genes that are essential for viability in human cell lines. We assessed the cumulative contribution of EGs to ASD risk by combined analysis of genomic and phenotypic data in ASD families and transcriptomic data in developing human brain. We found that EGs were enriched in the recently identified high-confidence ASD-associated genes. By examining the *de novo* and inherited variations from 1,781 ASD families from the Simons Simplex Collection, we demonstrated a higher burden of damaging mutations in EGs in ASD probands compared to their non-ASD siblings. In addition, we found that deleterious variants in EGs contribute to decreased social skills in males. Furthermore, the co-expression analysis of EGs in the developing brain identified three clusters of EGs implicated in ASD. Finally, we suggested 29 EGs as strong candidates for further investigation in their role in ASD. Overall, we provide multiple lines of evidence for a strong contribution of EGs to ASD risk. Our findings show that large-scale studies of gene function in model organisms provide a powerful approach for prioritization of genes and pathogenic variants identified by sequencing studies of human disease.

1418T

Deciphering the non-coding regulatory landscape in autism spectrum disorders. *J. Li[1], J. Zhou[1], Z. Ma[1], M. Shi[1], D. Phanstiel[1], R. Yuen[5], D. Merico[5], G. Li[1], H. Wang[3], M. Deurloo[3], Q. Li[3], B. Zhou[2], Y. Cheng[1], J. Hallmayer[2], A. Urban[2], Z. Feng[3], S. Scherer[5], M. Pletcher[4], M. Snyder[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94304, USA; 3) Department of Physiology, Collaborative Program in Neuroscience, University of Toronto, Ontario M5S 1A8, Canada; 4) Genomic Discovery, Autism Speaks Inc., 1 East 33rd Street, New York, NY 10016; 5) The Centre for Applied Genomics (TCAG), Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada M5G 0A4.

For decades, technical and cost hurdles have prevented the systematic investigation of non-coding sequences in complex human diseases, and thus our knowledge about autism spectrum disorders (ASD) has been primarily obtained from analysis of protein-coding sequences. We have combined the analysis of whole genome sequencing with global studies of regulatory sequences of human cortical neurons to reveal the regulatory architecture of ASD. Analysis of de novo mutations from whole genome sequencing of 261 autism families revealed the physical proximity of ASD de novo mutations specifically to the cortical expression quantitative loci (eQTLs) of synaptic genes. We performed ATAC-Seq, ChIP-Seq, RNA-Seq and Hi-C experiments on human cortical neurons, which for the first time provided a paranormal view of the regulatory landscape in these cells. We found that ASD de novo mutations preferentially affect regulatory elements, and the associated genes are shared targets of two ASD syndromic factors, CHD8 and PTEN. Analyzing 15 chromatin states across 127 human tissue/cell types revealed a significant enrichment of ASD de novo mutations in active transcription start sites and the perturbed genes implicated in neuron functions; this distribution enabled us to develop a machine-learning algorithm to assess potential ASD risk for a given individual. Taken together, our study for the first time revealed the regulatory landscape in human neurons, demonstrated the importance of the non-coding genome in ASD and provides a general framework for analyzing regulatory mutations for other complex human diseases.

1419F

CDH4 and TSNAX-DISC1 moderate a novel quantitative autism phenotype. *S. Luzi[1], L. Gomez[1], H.N. Cukier[1], S. Slifer[1], P.L. Whitehead[1], J.R. Gilbert[1,2], J.L. Haines[3], J.P. Hussman[4], E.R. Martin[1,2], M.L. Cuccaro[1,2], M.A. Pericak-Vance[1,2].* 1) Human Genomics, University of Miami - Miller School of Medicine (HIHG), Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA; 4) Hussman Institute for Autism, Baltimore, MD, USA.

Autism spectrum disorder (ASD) is highly prevalent and has a complex genetic architecture. The ASD phenotype is multi-dimensional and variable, consisting of core social-communicative and behavioral features which vary considerably among individuals. Changes in diagnostic criteria have contributed to an increased prevalence of ASD as well as phenotypic heterogeneity. We aim to identify a narrower ASD phenotype based on core ASD features and to find genes associated with this phenotype.We hypothesize that genetic variants modify the ASD phenotype and lead to an increased number of core autism features. We developed a Quantitative Autism Score (QAS) using items from the ADI-R which consistently discriminate ASD from non-ASD, occur early in development and remain stable throughout changing diagnostic criteria. Our initial discovery sample consisted of 1118 individuals with ASD, 268 ascertained at the Hussman Institute for Human Genomics (HIHG) and 850 from the Simons Simplex Collection (SSC).We targeted 689 genes (Griswold, 2015) and conducted a SKAT-O gene-based test using the QAS as a continuous trait. Results showed the *CDH4* and *LGR5* genes to be significant, adjusting for number of genes tested (p-values=9.20E-06 and 3.59E-05, respectively).To follow up these results, a meta-analysis was run on whole exome sequencing (WES) data available for the HIHG sample (N=253) and the SSC sample (N=1130). Again, results showed *CDH4* to be among the top nominally significant genes (p-value=0.0002). The *TSNAX-DISC1* gene was also nominally significant with a p-value of 0.0003.No genes were significant when we compared ASD cases to controls in a binary SKAT-O analysis.The QAS is a novel trait whose application in the current study narrowed the clinical phenotype to individuals with core features of ASD as a strategy for gene discovery. Refining the ASD phenotype using the QAS allowed for identification of two potential risk genes.*CDH4* is a neuronal cell adhesion molecule which plays a role in neuronal outgrowth and has been previously associated with ASD. *TSNAX-DISC1* is a transcription between the neighboring TSNAX and DISC1 whose alterations in gene processing have been associated with psychiatric disorders like Schizophrenia.These genes are associated with a greater presence of core ASD features, which can be an indicator of ASD severity.This novel approach is the first step towards dissecting a polygenic and multi-dimensional condition to clarify its underlying biology.

1420W

Identification of candidate genes for IQ discrepancy in extended families with autism using whole exome sequencing data: An update. *A.Q. Nato[1], N.H. Chapman[1], H.K. Sohi[1], B. Wang[2], H.D. Nguyen[1], J.M. Viskochil[3], H. Coon[3], E.M. Wijsman[1,4,5].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Statistics, University of Washington, Seattle, WA, USA; 3) Department of Psychiatry, University of Utah, Salt Lake City, UT, USA; 4) Department of Biostatistics, University of Washington, Seattle, WA, USA; 5) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Performance IQ (PIQ) is typically greater than verbal IQ (VIQ) in individuals with autism spectrum disorders (ASD). The strongest evidence of linkage on chromosomes (chr) 10p12 and 17 was previously implicated in a linkage analysis of IQ discrepancy (PIQ-VIQ) in University of Washington (UW) nuclear families [Hum Genet 2011:129,59-70]. We also reported a region of interest (ROI) on chr 15 from variance component (VC) linkage analysis of 4 large multi-generational families (457 individuals) from Utah (UT1), where 195 individuals were typed for ~750K SNPs, and 110 have IQ data. In the Utah (UT) data, we defined IQ discrepancy (PIQ-VIQ) using Wechsler block design or DAS pattern construction subtests (BD) to represent PIQ, and Wechsler vocabulary or DAS naming vocabulary and word definition subtests (VOC) to represent VIQ. Here, we analyzed 12 trimmed UT families (1,064 individuals), ranging from 23 to 153 individuals, where 520 individuals were typed for ~750K SNPs and 387 individuals have IQ data. UT1 is a subset of UT. Our current VC linkage analysis of IQ discrepancy (BD-VOC) in UT showed evidence of linkage on chr 1q, 9q, 10p, and 22q. The previous linkage signal on chr 15 decreased, however, we now have a linkage peak (lod=1.46) that partially overlaps the strongest evidence in UW PIQ-VIQ on chr 10p. The ROI on chr 10 spans ~22 cM. We obtained whole exome sequencing (WES) data of 86 individuals in 7 of the 12 UT families (782 individuals), and identified 16,560 variants on chr 10, where 1,383 are within this ROI. Within each family, we selected single nucleotide variants (SNVs) which have ≥2 copies, resulting in 160 SNVs (range: 19-114) after excluding indels. We further selected nonsynonymous SNVs (using SeattleSeq 138 annotation) that have minor allele frequency <0.05 (from 1000 Genomes European genotype data). This filtration step resulted in 23 nonsynonymous SNVs (range: 0-9 per family) in 13 genes (range: 1-7 per family), where 6 SNVs and 5 genes with nonsynonymous SNVs are observed in 2-4 families. 3 of these SNVs are predicted by PolyPhen to be possibly- or probably-damaging. We performed file manipulation through PBAP and genotype imputation for the remaining 696 individuals using GIGI. The dosages of both observed and imputed data are used as a measure of gene burden, which we use as a covariate in VC linkage analyses in SOLAR, to identify genes that *as covariates* decrease linkage signals.

1421T

Affected sib-pair analysis to identify risk variants for autism endophenotypes. *M. Pirooznia[1], T. Niranjan[2], F. Goes[1], D. Avramopoulos[1,2], J. Potash[3], P. Zandi[1], T. Wang[2].* 1) Psychiatry, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Psychiatry, University of Iowa.

Case-control and family-based studies have implicated many autism risk genes by identifying shared common variants of small effects, higher mutation load for rare variants of large effects, and recurrent and/or *de novo* variants in affected probands. Identification of genetic factors contributing to specific autism endophenotypes remains a big challenge. In certain multiplex families, affected sibs manifest significant difference in severity of one or more endophenotypes. Since sibs often grow up together during infancy and early childhood, familial contributions are reduced as a confounding factor for differences in severity of phenotype. Because sibs share a large fraction of their genome, we speculate that genetic contributions to differences in phenotypic severity likely reside in the fraction of their unshared genome. We developed a novel strategy using affected male sibs with extreme difference in severity to identify risk alleles for autism. We sequenced the exons of synaptome genes (n=1,886) in 274 pairs of affected male sibs from multiplex families. These sib-pairs were selected based on their extreme differences in severity of one or more endophenotypes. We first compared the cumulative burden of rare (MAF≤1%) and presumably functional variants from synaptome and pathway-specific genes between sibs with severe versus mild social deficits as determined by ADI-R social interaction scores and found no difference (*p*=0.96). We next identified subsets of affected sibs either with large (SOCT_CS≥10; n=92 pairs) or small (SOCT_CS≤4; n=108 pairs) differences in severity of social deficits as determined by cumulative ADI-R social interaction scores. We reason that sib pairs with large differences in severity harbor a higher load of variants contributing to social deficits as compared to sib pairs with small differences. Preliminary studies on the cumulative mutation loads in synaptome genes identified a significant difference in variant loads between the cohorts with large versus small differences in severity (large difference cohort: 705.1±16.2; small difference cohort, 668.3±9.0; *p*=0.025). Genetic and functional studies on selected variants potentially contributing to social deficits in autism are ongoing. Exploring variants in affected sibs with significant differences in severity of individual endophenotypes may be a valuable approach for identification of autism risk genes.

1422F

Dysregulation of translation processes: A persistent signal in the diagnostic prediction of autism in toddlers from community pediatric practices. *T. Pramparo[1], K. Pierce[1], C. Carter Barnes[1], L. Lopez[1], S. Nalabolu[1], Z. Guo[2], J. Li[2], N. Lewis[3], P. Yu[2], E. Courchesne[1].* 1) UC San Diego, Autism Center of Excellence, Department of Neuroscience, University of California San Diego School of Medicine, La Jolla; 2) Texas A&M University, Department of Electrical and Computer Engineering & TEES-AgriLife Center for Bioinformatics and Genomic Systems Engineering (CBGSE); 3) Pediatrics, University of California San Diego School of Medicine, La Jolla.

Recent efforts have provided evidence that accessible tissues, such as peripheral blood, may carry genetic signatures or biomarkers useful to predict infants and toddlers with Autism Spectrum Disorder (ASD) recruited from community practices (Pramparo et al., 2015). These findings hold the promise to be validated and further developed into a clinical test for ASD that can be used as a population screening in 1 year olds. However, a major obstacle in the study of genetic signatures of ASD is the identification of reproducible signals surviving the intrinsic large heterogeneity of the disorder as well as the presence of confounding variables. To continue our effort in the identification of early biomarkers of ASD, we sought to identify genetic signatures of ASD in blood at young ages using RNA sequencing data from a largely new pediatric sample. Our first goal was to validate at the functional level the previously reported gene expression signature which displayed gene enrichment for translation and immune functions and high specificity and sensitivity in diagnostic classification of ASD. Recruitment of participants was done in general pediatric clinics and community settings following criteria of the 1-Year Well-Baby Check-Up Approach. Diagnostic judgment followed DSM-V and Autism Diagnostic Observation Schedule criteria for autism. ASD male infants and toddlers (age range, 1-4 years) were compared with age- and gender-matched typically developing control subjects. Preliminary results displayed >4,200 differentially expressed genes between ASD and controls (q<0.05) that clustered into 22 modules of co-expressed genes. Permuted AUC-modeling (10,000 iterations) followed by logistic regression with cross-validation analysis of these 22 modules displayed a mono-modal distribution in accuracy with mean around 87%. Specificity and sensitivity distributions were instead bi-modal suggesting the presence of two genetically distinct subgroups of ASD subjects, and mean values ranged from 80 to 90%. Finally, pathway enrichment analysis, using one of the best-performing models included 1970 differentially co-expressed genes and displayed a strong significant enrichment in translation processes. These preliminary RNA-seq findings point to the previously identified genetic signature of ASD using a largely independent set of subjects, and strongly suggest that altered translation in ASD may be a persistent signal in blood useful for early diagnostic identification.

1423W

Integrating expression quantitative brain loci in autism spectrum disorder genome-wide association analyses. *C. Shu[1,2], C. Ladd-Acosta[2,3], K. Benke[1,2], A. Jaffe[4], J.L. Daniels[5], C.J. Newschaffer[6], A.M. Reynolds[7], D.E. Schendel[8], L.A. Schieve[9], M.D. Fallin[1,2].* 1) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 4) Lieber Institute for Brain Development, Baltimore, MD; 5) University of North Carolina, Chapel Hill, NC; 6) A.J. Drexel Autism Institute, Philadelphia, PA; 7) Pediatrics, University of Colorado Denver, Aurora, CO; 8) Aarhus University, Aarhus, Denmark; 9) National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA.

Background: Autism Spectrum Disorder (ASD) is highly heritable and evidence showed genetic variation plays a major role in this variability. However, genome-wide association studies (GWAS) thus far have had limited success. Genetics studies of other psychiatric disorders have shown enrichment for brain expression quantitative trait loci (eQTLs). Limiting genome-wide single nucleotide polymorphism (SNP) analyses to subsets known to be brain eQTLs (denoted "eSNPs"), can reduce the search space allowing important association signals to separate from signals simply due to millions of tests performed. **Objectives:** To perform genome-scale SNP association analyses for ASD, limited to SNPs known to be brain eSNPs or known to be located in genes expressed in early neural development and subsets defined by expression in specific brain regions. **Methods:** Brain eSNPs were obtained from 6 published brain eQTLs studies on 11 different brain tissue types. GWAS was performed on all SNPs, subsets of brain eSNPs, and brain tissue specific eSNPs after LD-based SNP pruning, using SNPs data from the Study to Explore Early Development (SEED) from 584 ASD cases and 725 non-ASD controls drawn from the general population. Comparisons were made by examining the patterns of QQ plots. **Results:** A total of 288,675 brain eSNPs were obtained after LD pruning, along with brain tissue specific eSNPs in cerebellum(3,027), frontal cortex(1,450), hippocampus(764), inferior olivary nucleus(659), occipital cortex(695), pons(440), putamen(425), substantia nigra(359), temporal cortex(1,420), thalamus(636), and intralobular white matter(1,034). GWAS based on brain eQTLs revealed SNPs that separated from expectation in QQ plots, while no separation was observed in the overall GWAS analysis of SEED data. The QQ patterns were also differential by subset in analyses based on eSNPs for specific brain tissues, where only QQ plots of temporal cortex eSNP subsets showed clear positive separation from expectation, and only QQ plots of cerebellum and frontal cortex showed suggestive separation with no apparent deflation and inflation. **Conclusions:** The findings reported here are consistent with literature on the key brain regions involved in ASD etiology, namely cerebellum, frontal cortex and temporal cortex. Subsetting GWAS analysis to brain eSNPs can provide further insight on the ASD common variant signals.

1424T

Exome sequencing of "pure" sporadic cases of autism identifies rare potentially causative mutations. *C.P. Sjaarda[1,2], A.J.M. McNaughton[1,2], C. Henry[1,2], L. Santavy[1,2], M.L. Hudson[1,2], M. Ayub[1,2], A. Guerin[3], X. Liu[1,2].* 1) Department of Psychiatry, Queen's University, Kingston, ON, Canada; 2) Queen's Genomics Lab at Ongwanada Resource Centre, Kingston, Ontario, Canada; 3) Department of Pediatrics, Queen's University, Kingston, ON, Canada.

   Autism Spectrum Disorder (ASD: MIM 209850) is a complex neurodevelopmental disorder affecting 1 in 64 children. Although genetic factors contribute to an estimated >90% of ASD cases and over 100 autism susceptibility genes have been described, only 20-25% of cases have a defined genetic cause. This study reports the analysis of the whole exome sequencing for 21 families with relatively "pure" sporadic cases of ASD based on the hypothesis that *de novo* damaging mutations are most likely the cause for these cases. These 21 families were selected from our database of >1300 families by examination of comprehensive family histories. Families with ASD in the extended family or neurodevelopmental and/or psychiatric disorder in the immediate family were excluded from the study. This rigorous selection process warrants that detected variants are true *de novo* variants. Analysis of exome sequencing data identified 349 rare, *de novo* variants in 289 genes, of which 37 variants in 35 genes have predicted high impact on gene function. Based on a literature search into the function of the genes containing these variants, 8 genes having a role in brain function or immune response were selected for further analysis. PCR and Sanger sequencing confirmed the presence of the SNP in 7 of the 8 variants (87.5%). *In silico* analysis using MutationTaster, MutationAssessor, Polyphen2, HumVar, FATHMM and SIFT showed that each of the 7 variants cause deleterious effects on the function of their respective gene products. This study identifies several *de novo* variants (*APEX1, EMP2, OR6C70, GAD1, TRIM60, XRCC6, ZC3HC1*) that may contribute to the genetic load underlying the development of ASD.

1425F

Metabolomic analysis leading to biomarkers and a potential blood-based test for early detection of autism spectrum disorder. *A.K. Srivastava[1,2], Y. Wang[3], R. Huang[4], C. Skinner[1], T. Thompson[4], L.R. Pollard[4], T. Wood[4], F. Luo[3], R.E. Stevenson[1].* 1) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC; 3) Department of Computer Science, Clemson University, Clemson, SC; 4) Biochemical Genetics Laboratory, Greenwood Genetic Center, Greenwood, SC.

   From the initial descriptions in 1943, autism spectrum disorder (ASD) has emerged as a common form of neurobehavioral disability that affects as many as 1/68 children born in the United States. Despite recent advances, the specific etiology of the majority of cases of ASD is unknown, although numerous genetic/genomic variants and alterations of diverse cellular functions have been reported. The current diagnostic protocol is limited to developmental screening and comprehensive behavioral evaluation that is complex and often difficult to administer to small children. Currently, there are no validated laboratory diagnostic tests or biomarkers for ASD that establish the clinical diagnosis, permit early diagnosis of ASD, or provide efficient screening of individuals with behavioral features suggestive of ASD. To address this, we have investigated whether the metabolomics approach might yield results which could simultaneously lead to a blood-based screening or diagnostic test and to treatment options. We performed nontargeted metabolic profiling, analyzing over 650 plasma metabolites, representing over 70 biochemical pathways, in two large cohorts of patients and controls (discovery cohort: 100 ASD, 32 controls; validation cohort: 83 ASD, 76 controls). The analysis revealed that the plasma metabolome is influenced by both disease status and age, with the effect being more prominent in younger age group (2-10 years of age). We identified more than 25 plasma metabolites that can discriminate children with ASD, as young as 2 years, from children that are developing typically. Furthermore, a quantitative analysis of 18 key discriminant metabolites in 127 ASD and 83 control plasma samples shows that various subsets of these metabolites could be used to detect the ASD samples with a sensitivity and specificity of greater than 85 percent. The metabolites associated with ASD represent several key biochemical pathways representing inflammation and immune dysregulation, increased oxidative stress, and mitochondrial dysfunction. The findings provide new insight by providing the basis for a first-generation blood-based quantitative screening test that can have immediate benefits by identifying individuals with ASD, even at 2 years of age, and thus adding to a better understanding of ASD and improving patient care.

1426W

Autism and obesity: Assessing the prevalence of antipsychotic-induced weight gain SNPs. *Z. Talebizadeh, A. Shah, H. Dai.* Children's Mercy Hosp, Kansas City, MO.

The existing challenge in the reproducibility of genetic findings for autism is, in part, due to its extensive level of heterogeneity. Initiatives such as the Autism Genetic Resource Exchange (AGRE) and Simons Simplex Collection (SSC) provide invaluable resources to the research community by collecting phenotypic data and ongoing generation of massive genetic data from subjects with autism. We designed a study to re-analyze already existing genetic data generated from autistic subjects from AGRE and SSC to address an important and previously unanswered question.   The objective of this project is to better understand the relationship between autism and obesity. It is not clear if obesity is co-occurring with autism or is related to antipsychotic-induced weight gain (AIWG). Weight gain is one of the main side-effects of the commonly used antipsychotics. Since the majority of patients with autism take antipsychotics, a *general assumption* is that the observed elevated rate of obesity in autism (i.e., 40%) is caused by AIWG. Our hypothesis is that the prevalence of known AIWG associated SNPs in obese and non-obese autistic subjects is comparable; thus, AIWG cannot be the only reason for the observed higher rate of obesity. To test this hypothesis, we will re-analyze already existing data (from AGRE and SSC families) by comparing the prevalence of AIWG associated SNPs in obese and non-obese autistic subjects. We will apply equivalence tests on genotyping data to test the presence of equality instead of differences between the two study groups (obese and non-obese). Weight/Height data has been obtained from the AGRE and SSC databases and BMI has been calculated for autistic subjects. We have compiled data on more than 100 AIWG SNPs, which cover all the reported SNPs, so far, and we refer to this list as our Master List in this study. We correlated the SNPs in our Master List with the dbSNP database, a public-domain archive for genetic polymorphisms to collect more detail information, including physical mapping, population data, and microarray platforms. We utilized several tools to obtain Tag SNPs: SNAP, TagSNP, and Tagger. We also searched for LD status of our Master list SNPs.   We downloaded AGRE and SSC genotyping data (combined Affymetrix and Illumina platforms). We are examining this genotyping list for the presence of the AIWG Master List SNPs and identified Tag SNPs. Final data including statistical analyses will be presented at the meeting.

1427T

Whole-genome sequencing provides additional insight into the genetic architecture of autism. *T.N. Turner[1], B.P. Coe[1], B.J. Nelson[1], M.C. Zody[2], F. Hormozdiari[3], Z.N. Kronenberg[1], S.A. McClymont[4], P.W. Hook[4], K. Hoekzema[1], M.H. Duyzend[1], A. Raja[1,5], C. Baker[1], R. Bernier[6], A.S. McCallion[4], R.B. Darnell[2,7], E.E. Eichler[1,5].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) New York Genome Center, New York, NY; 3) Department of Biochemistry and Molecular Medicine, Davis, CA; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Howard Hughes Medical Institute, University of Washington, Seattle, WA; 6) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 7) Howard Hughes Medical Institute, Rockefeller University, New York, NY.

We performed deep (30-fold) Illumina whole-genome sequencing (WGS) on 2,160 genomes from 540 simplex autism families negative for *de novo* likely gene-disruptive (LGD) mutation or large (>100 kbp) copy number variants (CNVs). Using a hybrid cloud / local compute analysis workflow, we processed all genomes in one month. We applied two SNV/indel and four CNV callers to generate a sensitive variant call set of 59 million SNVs/indels and 193 thousand unique CNVs. Extensive, orthogonal experimental validation was undertaken to determine the inheritance status of high-impact variants. WGS analysis recovered ~25% more *de novo*, exonic SNVs/indels and ~85% more gene-disrupting CNVs than previously discovered by whole-exome sequencing (WES) analysis of the same samples. This included *de novo* LGD events in *PCM1*, *MED12L*, and *VARS* and gene-breaking *de novo* CNVs in *APOBEC3D*, *APOBEC3F*, *CHD2*, *DDX43*, *DMD*, *FANCA*, *LINC01347*, *LNPEP*, *MIR3129*, *MUC19*, *PCDHB17*, *PCDHB6*, *TAF1B*, and *ZNF462*. We observed a significant enrichment (p=0.01) of *de novo* missense mutations in children with autism when compared to their unaffected siblings for autism risk genes with known dosage sensitivity. This included proband-specific events in *UBE3C*, *PTEN*, *SUV420H1*, *CREBBP*, *LAMC3*, *GABRB3*, *SYNGAP1*, *NR3C2*, *SRCAP*, *TRIP12*, *UNC45B*, *SCN2A*, *POGZ*, and *TRIO*. We report a modest enrichment (p=0.04) of *de novo* and private disruptive mutations for putative regulatory elements for dosage-sensitive autism genes. We define these as fetal central nervous system (CNS) DNase I hypersensitive sites mapping within 50 kbp of the start and end of the candidate gene transcript. Functional testing of the regions affected by these events confirms that the CNVs enrich in enhancers within the central nervous system. Overall, WGS provides additional insight into the genetic etiology of autism by significantly increasing the yield of gene-disrupting mutations and providing access to noncoding portions of the genome which when deleted adversely affect dosage of autism-risk genes during development.

1428F

Uncovering obsessive-compulsive disorder risk genes in a paediatric cohort by high-resolution analysis of copy number variation. *M. Zarrei[1], M.J. Gazzellone[1], C.L. Burton[2], S. Walker[1], M. Uddin[1], S-M. Shaheen[2,3], J. Coste[2], R. Rajendram[2,4], R.J. Schachter[2], M. Colasanto[2], G.L Hanna[5], D.R. Rosenberg[6,7], N. Soreni[8], K.D. Fitzgerald[5], C.R. Marshall[1], J.A. Buchanan[1], D. Merico[1], P.D. Arnold[2,3,9,10], S.W. Scherer[1,11].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Psychiatry and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Mathison Centre for Mental Health Research and Education and Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 4) Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 5) Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI, United States; 6) Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, United States; 7) The Children's Hospital of Michigan, Detroit, MI, United States; 8) Department of Psychiatry and Behavioural Neurosciences, Faculty of Health Sciences, McMaster University, St. Joseph's Healthcare, Hamilton, Ontario, Canada; 9) Department of Psychiatry and Institute of Medical Science, Toronto, Ontario, Canada; 10) Departments of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 11) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

    Obsessive-compulsive disorder (OCD) is a highly heterogeneous neuropsychiatric condition, thought to have a significant genetic component. When onset occurs in childhood, affected individuals typically experience a more severe manifestation, which can worsen in adulthood. Since neuropsychiatric conditions are associated with copy number variations (CNVs), we considered their potential role in the etiology of OCD. We genotyped 307 unrelated paediatric probands with idiopathic OCD (including 174 that were part of complete parent-child trios) and compared their genotypes with those of 3,861 population controls, in order to identify rare CNVs (<0.5% frequency) of at least 15 kb in size that might contribute to OCD. We uncovered *de novo* CNVs in 4/174 probands (2.3%). Our case cohort was significantly enriched for CNVs in genes that encode targets of the fragile X mental retardation protein ($p=1.85×10^{-03}$), similar to previous findings in autism and schizophrenia. These results identified deletions or duplications of exons in genes involved in neuronal migration (*ASTN2*), synapse formation (*NLGN1* and *PTPRD*), and postsynaptic scaffolding (*DLGAP1* and *DLGAP2*), which may be relevant to the pathogenesis of OCD. Four cases had CNVs involving known genomic disorder loci (1q21.1-21.2, 15q11.2-q13.1, 16p13.11, and 5p11-p12 and 17p12 in the same individual). Further, we identified *BTBD9* as a plausible new candidate gene for OCD. We also sequenced exomes of ten "CNV positive" trios and identified in one an additional clinically relevant mutation: a 13 bp deletion in *DRD4*.

1429W

Polygenic risk score and their clinical traits correlations in schizophrenia: From the first episode of psychosis to chronic patients. *S.I. Belangero[1], M.L. Santoro[1], V.K. Ota[1], S. Jong[2], C. Noto[1], E.S. Gouvea[1,3], P.N. Moretti[1], L.M. Spíndola[1], G. Xavier[1], F. Talarico[1], A. Gadelha[1], Q. Cordeiro[3], R. Bressan[1], G. Breen[2].* 1) Morphology and Genetics, Federal University of Sao Paulo, Sao Paulo, Brazil; 2) Institute of Psychiatry, King's College of London, London, UK; 3) Department of Psychiatry, Santa Casa de Misericórdia, São Paulo, Brazil.

    Background: Schizophrenia (SCZ) is a complex disease, with great influence of the environment but yet a disease with considerable heritability. The polygenic Risk Score (PRS) is an important tool that access the multiple genetic variations and their low effect sizes. Aim: In this study, we aimed to (1) confirm if PRS was able to differentiate cases and controls in two brazilian cohorts of SCZ patients; (2) verify if PRS is correlated to clinical variables of severity and of response to antipsychotics during: first episode of psychosis (FEP) stage, eight weeks after Risperidone treatment and in a chronic schizophrenia patients cohort. Methods: We performed the clinical assessment (SCID, PANSS, CGI and GAF) of 60 antipsychotic naïve patients in their FEP and after eight weeks of Risperidone tratment, 175 chronic schizophrenia patients and 297 healthy controls. After blood collection and DNA isolation, the samples were genotyped using the PsychArray Chips or Omni Express chips. We performed the imputation using the http://imputationserver.sanger.ac.uk and PRSice software to generate the scores and to evaluate the best p-threshold for the PRS in our sample. We used the last available data from PGC2 SCZ as a training sample. Results and discussion: we verified that the PRS works for the Brazilian population but with a moderate explained variance (0.15) and with a best p-threshold lower than that found in PGC (0.05). Furthermore, we found that GAF, CGI, PANSS exc and PANSS Dis are all positively correlated to PRS during the FEP but not after eight weeks of Risperidone treatment or to the chronic schizophrenia cohort. Recently, a study suggested that individuals with higher PRS for schizophrenia are more likely to discontinue longitudinal studies or do not respond questionaries, and thus, they are likely underrepresented in cohort studies. Curiously, we found that those FEP that did not return for the follow-up have even higher symptoms than that who did the follow-up. In the next studies we hope to genotype these individuals and confirm if they have even higher PRS scores. .

1430T

Replicative analysis of SNPs, associated with schizophrenia and its cognitive endophenotypes, in Kazakh patients with schizophrenia. *A. Bocharova[1], A. Marusin[1], K. Saduakasova[2], G. Svyatova[3], G. Berezina[3], V. Stepanov[1,4].* 1) Research Institute of Medical Genetics SB RAMS, Tomsk, Russian Federation; 2) Kazakh National Medical University, Almaty, Kazakhstan; 3) Republican Scientific Centre for Obstetrics, Gynecology and Perinatology, Almaty, Kazakhstan; 4) Tomsk State University, Tomsk, Russian Federation.

   Introduction. Schizophrenia is the most common psychotic disease, with a global prevalence of less than 1%. Patients with schizophrenia commonly experience debilitating social and occupational impairments. Symptom onset is generally between late adolescence and the mid-30s. Schizophrenia is associated with marked executive function disturbances, working memory impairment and thought disorder. These cognitive endophenotypes are frequently found in family and twin studies in clinically unaffected family members. Schizophrenia affects all ethnicities, but a large number of studies for schizophrenia have been conducted mainly in the Caucasoid population. However, it is desirable to research similar studies in different ethnic groups to reaffirm the status of the association identified. The aim of this study was to analyze associations of 45 SNPs reported in GWAS in Kazakh patients with schizophrenia and its cognitive endophenotypes. **Subjects and methods.** 102 patients with schizophrenia and 190 healthy controls, matched to the patients by age, gender, and ethnicity were included in study. 15 SNPs were genotyped by real-time PCR using TaqMan-assay (Applied Biosystems) and 30 SNPs were genotyped by MALDI-TOF mass-spectrometry using MassARRAY Analyzer 4 (Sequenom). Allele-specific ORs and associated p-values were calculated. **Results.** We found 5 significant associations of SNPs with schizophrenia in Kazakh patients: rs2312147 at *VRK2* gene (OR=0.58, p=0.008), rs2247572 at *KCNB2* gene (OR=0.65, p=0.03), rs2252521 at *CPVL* gene (OR=1.46, p=0.037), rs138880 at *BRD1* gene (OR=0.55, p=0.03), and rs6859 at *PVRL2* gene (OR=0.65, p=0.03). **Conclusions.** These genetic markers were previously reported in GWAS associated with cognitive endophenotypes (rs2247572 in *KCNB2* gene, rs2252521 in *CPVL* gene), schizophrenia (rs2312147 in *VRK2* gene, rs138880 in *BRD1* gene) and Alzheimer's disease (rs6859 in *PVRL2* gene).Genetic markers of *VRK2, KCNB2, CPVL, BRD1* and *PVRL2* are associated with schizophrenia but their role in pathogenesis of the disease is not clear. Our findings also demonstrate that genetic variability in schizophrenia, Alzheimer's disease and cognitive endophenotype has overlapping genetic background. This work was supported by the Russian Science Foundation (project # 16-14-00020).

1431F

Genetic risk for schizophrenia associates with living in cities. *L. Colodro Conde[1], B. Couvy Duchesne[1], G. Zhu[1], A. Meyer Lindenberg[2], M. Rietschel[2], S. Medland[1], J. Whitfield[1], N. Martin[1].* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) Central Institute of Mental Health, Mannheim, Germany.

   Background. Urbanization is positively associated with prevalence of schizophrenia and to explain this it has been proposed that the stress of city life may increase the risk of this disease. On the other hand, genetic factors have been shown to have a higher impact on the country v.s city living as people grow older (>40y), while the impact of family background decreases. Our aim was to test the hypothesis that adults with higher genetic risk for schizophrenia are more likely to move to and live in urbanized and populated areas than those with lower risk. **Methods.** Our sample was comprised of 15,253 participants in 7,019 families (63.4% women) over 40 years old (M = 54.43, SD = 10.86) living in Australia. The participants reported their postcode as part of the protocols of several studies on health and wellbeing conducted from QIMR. Participants were genotyped genome-wide and imputed to 1000G v.3. We used three measures of urbanicity; (i) subjects' postcodes were categorized into urban, suburban or rural zones; we also computed (ii) distance to the closest city centre and (iii) the population density of the postcode district. Polygenic risk scores (PRS) for schizophrenia based on summary statistics from the Psychiatric Genomics Consortium (PGC-2) were computed with PLINK 1.90 (version 3) for eight different P-value thresholds. We used linear mixed models to predict the three urbanicity measures from the schizophrenia PRS, controlling for age, sex, genetic principal components, GWAS chip and genetic relationship matrix. Predictions were calculated using GCTA (Genome-wide Complex Trait Analysis, version 1.22). **Results.** Genetic risk for schizophrenia was associated with the place where the participants lived in the expected direction, with increased effects as the risk scores included more SNPs. P-values for PRS at thresholds <0.05 were $7.20*10^{-5}$, $4.75*10^{-2}$ and $9.78*10^{-6}$, for zone, city distance and population density respectively while for threshold P<1 they were correspondingly $4.75*10^{-7}$, $1.2*10^{-3}$ and $1.19*10^{-6}$. **Discussion.** Our results suggest that people with a higher genetic risk for schizophrenia may prefer to live in more urban and populated areas. This study proposes that greater genetic predisposition to schizophrenia is at least one mechanism explaining why this illness is more prevalent in city environments.

1432W

Molecular subtypes of schizophrenia show negative symptom elevation and earlier age of onset: Application of subtyping methods accounting for population stratification and sex to the Wellcome Trust Schizophrenia Case-Control sample. *A.R. Docherty[1,2], D.E. Adkins[1,2], B.P. Reilly[1], A. Corvin[3], A.C. Edwards[1], K.S. Kendler[1], A.H. Fanous[1], S.A. Bacanu[1].* 1) Virginia Commonwealth University, Richmond, VA; 2) University of Utah School of Medicine, Salt Lake City, UT; 3) Trinity College, Dublin, Ireland.

   Background: Identifying genetic subtypes of schizophrenia (SZ) could reduce genetic heterogeneity prior to quantitative phenotypic profiling. Thus far, genetic subtyping analyses of SZ have been largely unsuccessful, mainly due to using all markers in the genome and not controlling for critical covariates such as population stratification and sex. The authors developed, tested, and present here methods to identify putative genetic subtypes of SZ. **Methods:** We used the implicated human biological pathways from the Psychiatric Genomics Consortium (PGC2) SZ GWAS to calculate pathway-based risk scores in the Wellcome Trust case-control sample. Two algorithms, both accounting for population stratification and sex, were considered feasible: a preferred method used model-based clustering in R, and an alternative that used model-based recursive partitioning which constructs its trees using only cuts which are statistically significant after adjusting for multiple testing. **Results:** While hierarchical clustering consistently resulted in only one genetic subtype, recursive partitioning resulted in 4 genetic subtypes with sufficient sample sizes. After testing these methods on a PGC subsample of SZ cases and controls, we then examined symptom dimensions (negative and positive) and age of onset in cases in order to differentiate the genetic subtypes phenotypically. Preliminary results show that individuals in one of the genetic subtypes have significantly elevated negative symptoms. Individuals in another have significantly earlier age of onset of psychosis. Because of its size, the large PGC2 sample provides optimal data for application of the genetic subtyping methods; thus, forthcoming analyses will apply these methods to the largest phenotyped SZ sample to date. **Conclusions:** Previous applications focused on classification algorithms of whole, high dimensional genotypic data, and neglected to account for critical covariates. The application of significance-based recursive partitioning methods to PGC2 biological pathway-based polygenic risk scores, to model genetic subtypes in GWAS, could significantly aid in studying the biological mechanisms of SZ risk. These analyses highlight the power of a careful, stepwise genetic subtyping framework, that is both computationally feasible and accounts for covariates, to provide potential avenues for early prediction and prevention of psychosis.

1433T

Genes influence human sexual partnering behavior. *L.A Farrer[1], R. Sherva[1], R. Koesterer[1], L. Almasy[2], H. Zhao[3], H.R. Kranzler[4], J. Gelernter[5].* 1) Biomedical Genetics, Boston Univ Sch Med, Boston, MA; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) Yale University School of Public Health, New Haven, CT; 4) Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, PA; 5) Departments of Psychiatry, Genetics and Neuroscience, Yale University School of Medicine, New Haven, CT.

   Human sexual partnering is a complex behavior that is influenced by both environmental and biological factors. We investigated the genetic basis of this behavior using a GWAS approach in a sample of 7,311 African- and European-American (AA and EA) subjects ascertained for genetic studies of substance use disorders who were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). Association of the log-transformed response to the SSADDA item "How many sexual partners have you had in your life?" was tested with SNP genotypes imputed to the most current 1000 Genomes reference panel. Analyses were conducted separately in AAs and EAs and also within males and females. Models included covariates for age and principal components of ancestry. Results across groups were combined by meta-analysis. We identified genome-wide significant (GWS) SNPs with *RAD18* (rs141996649, p=2.45x10^{-9}) and *PDSS2* (rs11964253, p=2.56x10^{-8}) in AA females, and *AGBL2* (rs28537909, p=3.24x10^{-9}) in AA males and females. The functional relevance of these genes to sexual partnering behavior is not clear. *RAD18* is involved in post-replicative DNA repair; *PDSS2* is an enzyme that synthesizes the prenyl side-chain of coenzyme Q; and *AGBL2* encodes a glutamate decarboxylase that catalyzes the deglutamylation of polyglutamylated proteins. We observed sub-GWS signals in several biologically interesting genes, including two genes in EA males and females including *OPCML* (rs73604550, 7.9x10^{-8}) which plays an accessory role in opioid receptor function and *CES1* (rs4122231, p=3.57x10^{-7}) which is involved in hydrolysis or transesterification of xenobiotics including cocaine and heroin; and *DLG2* in AA males and females (rs76371861, p=8.77x10^{-7}) which is required for the perception of chronic pain mediated by NMDA receptor signaling. These results suggest a shared genetic liability for substance use and sexual partnering behaviors. We also observed sub-GWS signals in genes involved in hormonal pathways and fertility including *KCNMA1* in EA and AA females (rs72805525, p=1.70x10^{-7}), which is a voltage gated K+ ion channel expressed in brain and smooth muscle and regulated by estrogen, and *NKAIN3* in EA males and females (rs1455569, p=2.59x10^{-7}), which encodes a molecule that interacts with Na/K transporting ATPase (sodium pump). The sodium pump has a role in aspects of steroidogenesis that are crucial for male fertility and reproduction. Replication efforts are underway.

1434F

Mouse brain transcriptome of innate and stress-induced anxiety-like behavior. *I. Hovatta[1], K. Trontti[1], J. Väänänen[1], B. Paranko[1], D. Greco[2]. 1) Department of Biosciences, University of Helsinki, Helsinki, Finland; 2) Institute of Biotechnology, University of Helsinki, Helsinki, Finland.*

Anxiety disorders are complex diseases with largely unknown etiology. Both behavioral and physiological responses to anxiety-provoking situations are evolutionarily conserved and therefore, mouse models allow investigation of both innate and stress-induced anxiety in a controlled environment and genetically homogeneous background. We have taken a cross-species approach to identify genes that regulate anxiety-like behavior using inbred mouse strains. We carried out RNAseq and miRNAseq of frontal cortex and hippocampus, two brain regions central in the regulation of anxiety, in two different mouse models of anxiety: 1) a panel of six inbred mouse strains (DBA/2J, A/J, 129S1/SvImJ, C3H/HeJ, C57BL/6J, and FVB/NJ) that differ in their innate anxiety; and 2) a psychosocial stress-induced anxiety (chronic social defeat stress) in two inbred strains (C57BL/6 and DBA/2). In the innate model, we correlated the gene expression levels with the levels of anxiety-like behavior (measured by the open field and light dark box tests) across the strains to identify genes and gene networks that putatively regulate anxiety. In the hippocampus, the differentially expressed genes were enriched in the vesicle mediated transport, mitotic nuclear division, and lipid metabolism Gene Ontology categories. In the frontal cortex, we observed enrichment in the cell division, vesicle mediated transport, and lipid metabolism categories. We are currently carrying out similar analysis with the psychosocial stress model. These two datasets offer an excellent setup to investigate whether the same biological processes are involved in the regulation of innate anxiety and stress-induced anxiety, i.e. if individuals who are innately more anxious are also more susceptible to stress-induced anxiety and whether these two different types of behaviors are regulated by the same molecular mechanisms. Our long-term aim is to determine whether the pathways identified using mouse models are involved in the etiology of human anxiety disorders.

1435W

***BDNF* methylation and stress response in a clinical population with major depressive disorder.** *C. Lattig[1], V.A. Villamizar[1], A. Gaviria[1], J.D. Angel[2], C. Cañizares[2], E. Ferro[3], A. Arenas[4], Y. Gomez[2]. 1) Dept Biological Sciences, Univ de los Andes, Bogota, Colombia; 2) Dept Psychology, Univ de los Andes, Bogota, Colombia; 3) El Instituto Colombiano del Sistema Nervioso, Clinica Montserrat. Bogota, Colombia; 4) Hermanas Hospitalarias Clínica La Inmaculada. Bogota, Colombia.*

Major Depressive Disorder (MDD [MIM: 608516]) is a multifactorial disease that affects morbidity, mortality and quality of life. It has a reported heritability of approximately 35-40%, indicating an important genetic component. Brain-Derived Neurotrophic Factor Gene (BDNF) Val66Met (rs6265) variant has a well-established role in neuronal plasticity and decreased levels have been observed in individuals with MDD. Additionally, early childhood adversity and stressful life events are recognized environmental risk factors for MDD. Previous studies present evidence that support the idea that vulnerability to environmental stress is mediated by the rs6265 variant. On the other hand, changes in methylation patterns in exon IV of BDNF are emerging as a possible explanation for how stress events can disrupt BDNF expression. However, there is still some controversy on the topic, mainly because most of these studies have been carried out in non-clinical populations. Here we present a case-control study using a clinical population from two psychiatric clinics and matched controls. Informed consent of the study was provided by all individuals. This study was approved by the ethics committee of all the participant institutions. Early stress events where assessed in the clinical questionnaire, the stressful live events questionnaire was applied to evaluate number and response to adverse events. The rs6265 variant was genotyped using real time PCR with Taqman probes from Applied Biosystems 7500/7500. Methylation patterns in exon IV of BDNF were assessed by bisulfite treatment using the EZ-96 DNA Methylation kit (Zymo research) followed by Sanger sequencing and subsequent quantification of methylation percentage at each CpG region. We did not find any association or gene x environment interaction between the rs6265 variant and MDD, however when analyzing methylation patterns we did find an association among percentage of methylation and MDD (p=0.001). We also found association in methylation patterns among controls that had experienced early adverse events (p=0.005), however this association was not observed in cases. Our results suggest that initial alteration of BDNF expression resulting from higher levels of methylation as observed in individuals who experienced early life events, may not necessarily predispose an individual to develop MDD supporting the idea that depression vulnerability and stress response can be mediated by different factors.

1436T

Alcoholism, schizophrenia and Alzheimer's disease susceptibility genes associations with cognitive and mental traits in Russian population. *A.V. Marusin[1], A.N. Kornetov[2], M.G. Swarovskaya[1], K.V. Vagaitseva[1], E.S. Pavlenyuk[2], V.A. Stepanov[1].* 1) Scientific Research Institute of Medical Genetics of Russian Academy of Science, Tomsk, Tomsk area, Russian Federation, 634050; 2) Siberian State Medical University, Tomsk, Tomsk area, Russian Federation, 634050.

    Objectives There is abundant evidence, that personality traits are substantially influenced by the genes. All the genes that determine the intellectual abilities, especially traits of thought and behavior have not been identified yet. The aim of this study was to analyze associations of 12 SNPs polymorphisms involved in the development of severe mental disorders - alcoholism, schizophrenia and Alzheimer's disease with cognitive and mental traits. **Methods** Associations of personality, temperament and character quantitative traits determined by Cattell's (The Sixteen Personality Factor Questionnaire (16PF)), Leonhard's, Spielberger-Khanin's, and Eysenck's (IQ) tests were analyzed with 12 gene polymorphisms. DNA samples from 150 students were genotyped using TaqMan® SNP Genotyping Assays (Applied Biosystems) under condition recommended by the manufacturer. **Results** Paired combinations of *GABRA2 - PICALM*, *PICALM - ADCY3*, *CLU - CBX7* and *CLU - ADCY3* polymorphisms in interallelic nonrandom associations was detected. This may reflect the adaptive nature of natural selection, it is directed on maintaining behavioral homeostasis in the population. A number of statistically significant associations of genetic variation were found: *CLU* with perfectionism (Q3) of 16PF and the exaltation of the Leonhard's tests, *PICALM* with tension (Q4) of 16PF and the imbalance of Leonhard's tests, *DISC1* with vigilance (L) of 16PF, and exaltation, cyclothymia of Leonhard's tests, *ZNF804A* with imbalance by Leonhard's test, *SLC6A4* with reasoning (B) of 16PF test, *ADCY3* with self-reliance (Q2) and extraversion (F2) of 16PF test, *MIR9-2* emotional stability (C), liveliness (C), social boldness (H), extraversion (F2) of 16PF, and dysthymia, hyperthymia of Leonhard's tests, with the personal anxiety of Spielberger-Khanin's test, *CBX7* with vigilance (L), and warmth (A) of 16PF test, *SLC6A3* with IQ by Eysenck's test. For the polymorphism rs13219354, *GAB2*, and *GABRA2* relationship to the the intelligence, temperament and character traits were not found. **Conclusion** Perhaps data of 12 SNPs associations with personal characteristics may be applied for the treatment of severe mental disorders. This work was supported by the Russian Science Foundation (project # 16-14-00020).

1437F

Genome-wide association study of major depressive disorder in European Americans and African Americans. *L. Wen[1], Y. Yao[1], W. Cui[1], T.J. Payne[2], J.Z. Ma[3], M.D. Li[1,4].* 1) State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Zhejiang Universityj, Hangzhou, China; 2) Department of Otolaryngology and Communicative Sciences, University of Mississippi Medical Center, Jackson, MS; 3) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 4) Institute of NeuroImmune Pharmacology, Seton Hall University, South Orange, NJ.

    Major depressive disorder (MDD) is a common complex mental illness, which is one of leading causes of preventable deaths and disabilities throughout the world. To date, genome-wide association study (GWAS) for MDD and its related phenotypes has not yielded consistent results. In this study, we conducted a GWAS for the Center for Epidemiological Studies Depression (CES-D) score, a widely used questionnaire for measuring MDD, in European American (EA) and African American (AA) populations, with age, sex, ancestry-informative principal component scores, and smoking status as covariates. Considering there is no evidence for the presence of heterogeneity between the AA and EA samples, we analyzed the two populations together, giving a sample size of 4,817. Our results revealed a genome-wide significant association between a nonsynonymous SNP rs61753730 in *FZD6* (frizzled class receptor 6; a negative regulator of the canonical Wnt/beta-catenin signaling cascade) that gives rise to a glutamine-to-glutamate substitution and CES-D score ($P = 8.46 \times 10^{-9}$). Such a significant finding has received independent support from molecular study using an animal model where knockdown of *FZD6* expression led to a depressive effect. Because there were a few SNPs included in the Illumina Infinium HumanExome BeadChip for *FZD6*, we performed imputation analysis for the gene with IMPUTE2 using 1000 Genomes as references and then conducted a gene-based association study using SNP-set Kernel Association Test (SKAT) in the pooled samples, which revealed a significant association of *FZD6* with the CES-D score ($P < 0.01$). Additional gene-by-gene interaction analysis using a generalized multifactor dimensionality reduction (GMDR) algorithm showed a significant interaction between *FZD6* and *CREB3L1* ($P = 0.0016$) and between *FZD6* and *CREB3L4* ($P = 0.002$). Finally, we showed that tobacco smoking plays a significant role in the association of *FZD6* with depression, in that significant association of the gene with the CES-D score was detected only in smokers. This study shows that *FZD6* is a plausible candidate gene for MDD, and further study with a large sample is warranted.

1438W

Whole-exome sequencing in bipolar disorder identifies rare variants associated with Parkinson disease and other neuropsychiatric diseases. *X. Jia[1,3], M. Kvale[2,3], L. Shen[4], P.Y. Kwok[2,3], C. Schaefer[4], N. Risch[2,3,4].* 1) Department of Neurology, University of California San Francisco, San Francisco, CA; 2) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 4) Kaiser Permanente Northern California Division of Research, Oakland, CA.

We performed whole-exome sequencing in 483 individuals with bipolar disorder (BD) from the Multi-ethnic Study of Bipolar Disorder and 483 ancestry-matched controls from the Kaiser Permanente Research Program on Genes Environment and Health (RPGEH) using the NimbleGen SeqCap EZ system. Of the 5.8 million called genomic positions, we extracted 1,013,033 positions within the capture target interval with calls made at ≥ 10X coverage. We validated exome sequencing genotypes in 104,719 overlapping SNPs genotyped using the Affymetrix Axiom system, which contained 675,000 to 891,000 SNP markers on four ancestry-specific arrays (non-Hispanic white, African-American, Latino, and Asian). Genotype concordance ranged from 98.1% to 99.3% across the four arrays. Among 966 individuals, we identified a total of 842 rare (minor allele frequency < 0.5%) nonsynonymous exonic and splice-site variants that were predicted to be pathogenic or likely pathogenic by ACMG Guideline classification. Of 290 genes containing a pathogenic or likely pathogenic variant in more than one individual, none reached statistical significance for association with bipolar disorder. However, analysis of top 5% associated genes revealed an enrichment of pathogenic variants in genes previously implicated in bipolar disorder or other neuropsychiatric conditions. *LRRK2*, previously associated with mood disorders in patients with Parkinson Disease, contained exonic mutations in 6 cases and 1 control (OR = 6.1, *p* = 0.13). Mutations in *GBA*, the lysosomal enzyme deficient in Gaucher's disease associated with Parkinson Disease in heterozygotes, was found in 8 cases and 3 controls (OR = 2.7, *p* = 0.22). *ATP7B*, a copper-transporting ATPase implicated in Wilson Disease, contained variants found in 7 cases and 2 controls (OR = 3.5, *p* = 0.18). *DPYD*, a risk locus for schizophrenia, contained variants found in 7 cases and 2 controls (OR = 3.5, *p* = 0.18). These findings suggest a possible pre-morbid link between bipolar disorder and Parkinson disease, as previously observed in *LRRK2*. Additional studies are needed to further investigate the role of rare variants in co-morbid risk for bipolar disorder and other neuropsychiatric conditions.

1439T

The first case of a child with a Joubert Syndrome with obsessive compulsive disorder. *Y. Cengiz[1], M.C. Ergoren[2,4], B. Kaymakamzade[3].* 1) Child and Adolescent Pyshchiatry, School of Medicine, East University, Nicosia, Cyprus; 2) Department of Medical Biology, School of Medicine, East University, Nicosia, Cyprus; 3) Department of Neurology, School of Medicine, East University, Nicosia, Cyprus; 4) Experimental Health Sciences Research Centre,School of Medicine, East University, Nicosia, Cyprus.

Joubert syndrome (JS) is a rare autosomal recessive disorder with partial or complete agenesis of cerebellar vermis. This syndrome is identified mainly by the presence of molar tooth sign in magnetic resonance imaging of the brain since it has a varied phenotypic presentation. Obsessive-compulsive disorder (OCD) is a neuropsychiatric disorder, commonly having an early age of onset, and is characterized by the presence of obsessions (unwanted, intrusive thoughts) and compulsions (repetitive behaviors) that can become incapacitating. Here in, we report a 16 year-old male patient with JS who shows the symptoms of OCD. He was admitted to the neurology department with the chief complaint of inbalance and abnormal eye movements since his birth. He had also delayed motor and cognitive milestones. His mother described neonatal episodic hyperpnea and hypotonia for the early childhood period. On neurological examination he had oculomotor apraxia and truncal ataxia. The brain MRI revealed typical molar-tooth sign. The patient presented to child and adolescent psychiatry department, two years duration of illness characterized by repetitive questions, extreme anxiety and anger when non-answered questions by his mother and worrying about getting sick. Neuropsychological testing performed after clinical interview using the Wechsler Intelligence Scale for Children, revised edition, he demonstrated moderate intellectual disability overall. There is increasing understanding that the interconnectivity between association neocortex, limbic cortex and the cerebellar hemispheres may play a key role in the processing of multimodal information. In turn, disruption of this circuitry may play a role in neuropsychiatric disease. There are many studies that show relation between Joubert syndrome and autistic spectrum disorders suggesting a link between abnormalities of the cerebellar vermis and autism. To the best of our knowledge, this is the first case of Joubert syndrome with OCD. Recent morphological studies demonstrate a potential important role of cerebellum in OCD. Studies focusing on psychiatric symptoms in JS are necessary for clarifying their relation.

1440F

Whole genome sequencing in a multigenerational family with a specific deficit in semantic cognition. *A.K. Tilot[1], K.S. Kucera[1], J. Briscoe[2], D. Skuse[3], S.E. Fisher[1,4].* 1) Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, Gelderland, Netherlands; 2) School of Experimental Psychology, University of Bristol, Bristol, United Kingdom; 3) Behavioural and Brain Sciences, Institute of Child Health, University College London, London, United Kingdom; 4) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Gelderland, Netherlands.

Semantic cognition refers to our ability to connect words with meanings, and make use of those relationships in language. In the JR family, a specific deficit in semantic cognition manifests as difficulty recalling verbally presented information and frequent struggles with the tip-of-the-tongue phenomenon. Intensive phenotyping of this family previously revealed deficits in memory for stories, word learning, listening comprehension and using conceptual knowledge to match words across lists. This constellation of phenotypes was remarkably consistent across affected family members, and the pedigree suggested a dominant and highly penetrant inheritance pattern.   Our goal in this study was to identify a potential genetic explanation for the phenotype, following MRI findings that suggested structural abnormalities in the anterior temporal cortex, an area associated with semantic cognition. For this work, we adopted a monogenic model which assumed that any causative mutations would be rare and follow a dominant inheritance pattern. Initial genotyping of 15 family members (seven affected) identified several suggestive regions of linkage, but none that exceeded the LOD>3 threshold for significance. We next applied whole genome sequencing (WGS) to five samples (three affected). We integrated WGS-based structural variant calls from three programs with copy number variation in the remaining samples, and did not find evidence for a causative structural variant in the family. Filtering smaller variants identified in the WGS data by their inheritance pattern and minor allele frequency (< 0.001) yielded 467 rare variants for further analysis.   To aid variant prioritization, we compared three methods for assessing deleteriousness (Variant Effect Predictor, CADD, and Eigen) and constructed a list of 12 variants categorized as likely deleterious by at least one program. Interestingly, there was only modest correspondence between the three methods, with only four variants considered strongly deleterious by more than one program. Half of the variants were outside of coding regions, and none were linked to related cognitive deficits. These results suggest that this unique neurological phenotype may stem from genetic alterations outside of those currently associated with cognitive phenotypes. Further work will determine if any of the variants in noncoding regions are tied to genes implicated in other cognitive deficits through regulatory networks.

1441W

Gene-environmental interactions between genetic variants in antiviral response genes and prenatal maternal infections in schizophrenia. *V.M. Rajagopal[1], L. Reinert[1], S.R. Palludan[1], J. Grove[1], P.F.R. Nielsen[2], R. Yolken[3], D.M. Hougaard[4], M. Nordentoft[5], T. Werge[6], A. Børglum[1], P.B. Mortensen[2], D. Demontis[1].* 1) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 2) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 3) Stanley Division of Developmental Stanley Division of Developmental Neurovirology, The Johns Hopkins University School of Medicine, Baltimore, MD; 4) Statens Serum Institut, Copenhagen S; 5) Mental Health Centre Copenhagen, Capital Region of Denmark; 6) Mental Health Center Sct. Hans, Capital Region of Denmark.

Background: Several epidemiological studies have found that prenatal maternal infections with herpes simplex virus 1 and 2 (HSV1 and HSV2) and cytomegalovirus (CMV) increase the risk for schizophrenia in the offspring. Multiple mechanisms have been postulated explaining the biological plausibility behind these associations. One among them is activation of the innate immune system leading to release of inflammatory mediators disrupting the neurodevelopment in the embryo. We studied the gene environment interactions between single nucleotide polymorphisms located in the genes (*TLR3, TMEM173, MV21D1, IFNB1, IFI16, TBK1, IRF3* and *IRF7*) involved in innate immune response and prenatal maternal infections with viruses: HSV1, 2 and CMV in schizophrenia. Also, we studied the imputed expression levels of those genes with respect to schizophrenia and maternal infection status. **Methods**: Totally 1806 samples comprising 908 cases (ICD-10: F20) and 898 healthy controls were identified from the Danish Psychiatric Central register. The DNA was isolated from the dried blood spot of the participants stored in the Danish Newborn Screening Biobank. Maternal infection status was established for HSV1, 2 and CMV viruses by measuring the IgG antibodies from the dried blood spot. Genotyping was done using the exome core chip from Illumina. SNPs located within and around 5kb up/downstream of the genes of interest were extracted and analyzed for interaction between maternal infection and offspring genotype using logistic regression. The expression levels of the genes in various brain tissues were imputed using PrediXcan software and compared between the samples with mothers with and without infections among cases and controls. **Results**: We found nominally significant interactions of variants in the analyzed genes of the offspring with maternal infections: *IFI16* with HSV1, *TLR3* with HSV2, and *TMEM173*, *IRF7* and *TBK* with CMV. Also, we found significant differences in the imputed expression patterns of *TLR3*, *IFNB1*, *IRF7*, *MB21D1*, *TBK1*, *TMEM173* and *IFI16* in the brain between the groups. **Conclusion**: In summary, we found indications that the offspring risk of schizophrenia could be influenced by genetic variations in the genes involved in innate immune response interacting with the presence of viral infections (HSV1, 2 and CMV) in the mother during pregnancy. Additionally our results indicate that such interactions could be due to differential expression of the studied genes in the brain.

1442T

Leveraging polygenic risk scores to discover novel biomarkers of disease in a large biobank setting. *L. Davis[1], J. Dennis[1], D. Hucks[1], G. Chen[2].* 1) Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN.

For polygenic traits, genetic risk is distributed throughout the genome and therefore individual single nucleotide polymorphisms (SNPs) explain a very small proportion of trait heritability. Though individually small, the aggregate effects of common SNPs are substantial and can be measured in the form of polygenic risk scores (PRS) defined as the sum of the number of risk alleles weighted by the effect size of each risk allele based on a discovery GWAS. PRS are often used to estimate genetic correlation between complex traits. It follows that quantitative measures (i.e., biomarkers) of a given disease state may also be correlated with the PRS for that disease or disorder. GWAS summary statistics are available for many phenotypes for which there are no reliable biomarkers, such as neuropsychiatric traits. The Vanderbilt University Medical Center Biobank (BioVU) includes genetic data and routinely collected clinical laboratory measurements on over 20,000 people. Lab measures such as complete blood count (CBC) and lipid panels, among many others ($N_{labs} > 200$), can be mined for novel biomarkers of disease through PRS analyses. In a proof of principal experiment, we calculated PRS for triglycerides (PRS_{trigs}) (based on the Global Lipids Genetics Consortium meta-analysis results) in each individual in BioVU genotyped on the Omni-Quad platform (N = 3,422). The best fit PRS_{trigs} (i.e., weighted sum of 3,307 independent triglyceride associated SNPs with p<0.01) accounted for a significant proportion of the variance in mean triglyceride levels ($R^2 = 0.034$; $p = 4.2 \times 10^{-27}$) while adjusting for sex, age, and ancestry. These results provide evidence that PRS derived from GWAS can account for a significant proportion of variance in routinely collected laboratory measurements across a cohort-like hospital population. We will report on our ongoing project to create a panel of PRS representing all complex traits for which GWAS have been performed, for each individual in BioVU with existing genetic data. Associations between PRS of complex traits and available laboratory measurements in BioVU will be presented as will pitfalls and lessons learned during analysis. The application of PRS to laboratory measures collected in a biobank setting provides a unique opportunity to identify biomarkers for notoriously elusive phenotypes such as neuropsychiatric traits, and may substantially improve diagnostics, early interventions, and therapeutics for many diseases and disorders.

1443F

Polygenic risk scores reveal patterns of assortative mating and anticipation in a large severely affected pedigree. *S. de Jong[1], M.J.A. Diniz[2,3], A.C.S. Rodrigues[2,3], A. Gadelha[2], M.L. Santoro[4], V.K. Ota[4], C.S. Noto[4], C. Curtis[1,5], H. Patel[1,5], L. Hall[6], P.F. O'Reilly[1], S.I. Belangero[2,5], R. Bressan[2], G. Breen[1,5].* 1) MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, United Kingdom; 2) Department of Psychiatry, Universidade Federal de São Paulo (UNIFESP/EPM), São Paulo, Brazil; 3) Pax Clinica Psiquiatrica, Rodovia Br153, Santa Luzia - Aparecida de Goiânia/GO CEP: 74922-810; 4) Department of Morphology and Genetics, Universidade Federal de São Paulo (UNIFESP/EPM), São Paulo, Brazil; 5) National Institute of Health Research Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom; 6) Institute of Genetic Medicine, International Centre for Life , Newcastle University, Newcastle upon Tyne, United Kingdom.

The aim of this study is to examine the role that common genetic variation confers on disease risk in a severely affected pedigree that would have traditionally been viewed through the prism of monogenic inheritance only. To achieve this we exploit both a large pedigree (n~300) with high prevalence of mood disorders (30% of family members is affected with major depressive disorder (MDD) or bipolar disorder (BP)) and the recently popular polygenic risk score (PRS) approach. Illumina Psych chip data is available for 243 family members of which 78 individuals are affected with mood disorders (including 36 BP and 38 MDD patients) and 59 unrelated healthy Brazilian controls. Using psychiatric PRS derived from recent large GWAS mega analysis efforts in psychiatric disorders like schizophrenia (SCZ), BP and MDD, we examine patterns of assortative mating and anticipation in the family. First, we examine common variant effect on risk via an association between PRS and case/control status in the pedigree. Next we investigate an increase in PRS value across the generations, seemingly driven by the transmission of elevated PRS to offspring from 11 affected married-in individuals (vs 35 unaffected married-in individuals). This suggests that PRS contributions from married-in individuals, caused by assortative mating on phenotype, may increase risk for psychiatric disorders beyond that conferred by rare variation transmitted from within the family. A joint analysis of both rare and common variation may be the most powerful way to understand the familial genetics of mood and psychiatric disorders.

1444W

Transcriptomic signatures of schizophrenia revealed by dopamine perturbation of lymphoblastoid cell lines. *P.V. Gejman[1,2], J. Duan[1,2], A.R. Sanders[1,2], E. Drigalenko[3], H.H.H. Göring[4].* 1) Ctr Psychiatric Genetics, University of Chicago School of Medicine, Chicago, IL; 2) NorthShore University Health System, Evanston, IL; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, San Antonio, TX.

We performed a dopamine perturbation on lymphoblastoid cell lines from 515 schizophrenia cases and 692 controls and studied the resultant RNA-seq profiles. Dopamine had widespread effects on expression profiles, though these effects were likely not mediated through dopamine receptors. Upon stimulation, of the reliably expressed genes, 3,756 (19%) had expression changes >1 SD at FDR<5%. Of those genes, 539 showed differential response in schizophrenia cases and controls. Pathway analyses uncovered that the set of differentially expressed genes were enriched for two factors: (1) brain expression, and (2) immune and/or apoptotic mechanism involvement. The latter observation is consistent with schizophrenia GWAS findings whose most significant locus is the extended major histocompatibility complex region, which contains numerous genes with fundamental roles in immune response.

1445T

Integrative genomic analysis reveals casual role of genes and networks involved in nicotine consumption using both human and animal data. *P. Huang[1], T. Nesil[2], S. Wang[2], Z. Yang[1], M.D. Li[1,2].* 1) State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; 2) Institute of NeuroImmune Pharmacology, Seton Hall University, South Orange, NJ USA.

Family and twin studies have revealed a moderate-to-high heritability of nicotine dependence (ND), a primary factor maintaining continuous tobacco smoking. However, the molecular mechanisms underlying ND are still largely unknown. In this study, we used the RNA deep-sequencing approach to determine the hippocampus transcriptome profiles of two lines of rats with 9.3-fold difference in nicotine consumption. We detected 1,275 differentially expressed (DE) genes and 1,512 DE transcripts encoded by 1452 genes. Functional enrichment analyses revealed significant overrepresentation of Gene Ontology categories involved in regulating synaptic transmission, plasticity, and the pathways implicated in ND such as calcium signaling, long-term potentiation, and glutamatergic and dopaminergic synapse signaling. Further, weighted gene co-expression network analysis (WGCNA) identified 18 modules whose overall expression profiles were significantly associated with nicotine intake. Follow-up integration analysis using three human GWAS datasets confirmed an important role for a gene module (GM12), where significant enrichment was detected for included genes by containing at least one nominally significantly associated SNP with ND. The GM12 module contained a number of genes known to be important in nicotine addiction, such as *Chrna4*, *Gabrb2*, *Grin3a*, *Grin2d*, and *Slc17a6*. Functional characterization of GM12 revealed their alliances with nicotine addiction, neurotransmitter receptor binding, and gated channel activity. In sum, we provide convergent evidence for the potential casual roles of large-scale gene expression alterations and coordinated differential expression of one ND-associated sub-network in nicotine intake in rats. Given the high similarities between nicotine intake in rats and smoking in humans, our findings provide new insights into the molecular mechanisms underlying nicotine addiction, although further replication and investigation are needed.

1446F

Genetic and phenotypic heterogeneity of mood disorders in a large multigenerational pedigree. *R.L. Kember[1], L. Hou[2], X. Ji[1], L.H. Andersen[3], L.N. Estrella[3], F.J. McMahon[2], C.D. Brown[1], M. Bucan[1].* 1) Department of Genetics, Philadelphia, PA; 2) Human Genetics Branch, National Institute of Mental Health Intramural Research Program, National Institutes of Health, Bethesda, MD; 3) Lancaster General Health/Penn Medicine, University of Pennsylvania Health System, Lancaster, PA.

Bipolar disorder (BD) is a mental disorder characterized by alternating periods of depression and mania. Individuals with BD have higher levels of early mortality than the general population, and a substantial proportion of this may be linked to increased risk for co-morbid diseases. Medical co-morbidity may be a consequence of either chromosomal proximity of a BD-risk gene to a gene underlying non-psychiatric phenotype, or the pleiotropic effect of a risk-allele on both mental illness and other phenotypes. In order to identify the molecular events that underlie BD and related medical co-morbidities, we performed whole genome sequence (WGS) and RNA sequence analysis of lymphoblastoid cell lines from an extended multigenerational Old Order Amish pedigree (400 family members) segregating BD and related disorders. We mapped disease causing variants at known Mendelian loci and performed genomic profiling using polygenic risk scores for several complex diseases, allowing us to genetically 'diagnose' all family members. In addition to known Mendelian mutations, we identified 167 high confidence loss-of-function (LoF) mutations in disease genes. Each individual has on average 31 LoF variants in disease genes with five disease genes predicted to be completely inactivated. Variants in five of these genes (*TMCO1, TPO, SPG21, IDH2, COL18A1*) lead to differential expression between individuals with and without the LoF variant. To explore the contribution of disease genes to BD we performed gene-based and variant-based association tests for BD, and found that Mendelian disease genes are enriched in the top results from both tests. We identified disease causing variants in cardiovascular genes which are found at a higher frequency in individuals with BD compared to unaffected individuals, including a variant in *APOB* associated with hypercholesterolemia. Finally, we examined the region containing the top BD associated linkage peak, and identified a cluster of damaging mutations in human disease genes. In particular, one of the haplotypes associated with BD contains multiple damaging variants in disease genes. Taken together, our results indicate that medical co-morbidity between complex diseases and Mendelian disorders arises as a combination of chromosomal proximity of disease causing variants and pleiotropy of disease genes.

1447W

A second update on susceptibility genes for nicotine dependence identified by genome-wide linkage, candidate gene association, genome-wide association, and targeted sequencing approaches. *M. Li[1,2], J. Yang[1,2], T. Payne[3], J. Ma[4].* 1) Institute of Neuroimmune Pharmacology, Seton Hall University, South Orange, NJ; 2) State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou, China; 3) Department of Otolaryngology and Communicative Sciences, University of Mississippi Medical Center, Jackson, MS; 4) Department of Public Health Sciences, University of Virginia, Charlottesville, VA.

Tobacco smoking is a severe health hazard worldwide, as nearly one-third of the global adult population smokes tobacco products, and these have been associated with numerous serious health problems. This high prevalence of tobacco use highlights the importance of studying the genetic determinants of nicotine dependence (ND). To identify genetic factors for ND, various approaches have been used, including genome-wide linkage, candidate gene-based association, genome-wide association (GWAS), and targeted sequencing analysis. In this study, we systematically analysed the findings from all the abovementioned approaches according to rigorous selection criteria for each included study such as sample size, statistical significance, and independent replication. Our analysis revealed 14 regions nominated by genome-wide linkage analysis and 34 significantly associated loci in 43 genes by candidate gene-based association. The GWAS and meta-GWAS revealed 11 genome-wide significant loci; however, only the loci on chromosomes 8, 15, and 19 have received independent replication. Although it is still in early stages, limited targeted sequencing studies using next-generation techniques have implicated 18 variants in the aetiology of ND. Together, we identified 14 linkage regions and 47 unique loci in 60 genes involved in the development of ND, which forms our current understanding of the susceptibility map for ND. Because almost all of these loci and genes have received replication by independent approaches in different samples, they should be considered high priorities for future functional study of ND.

1448T

Contribution of common genetic variation to risk for severe developmental disorders. *M. Niemi, J. Barrett, The Deciphering Developmental Disorders Study.* Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

Rare and common variants in the genome both contribute substantially to human disease, but these have typically been studied separately from each other. Rare disease variants often have severely deleterious effects on an individual, whereas common variants tend to have relatively small individual effects but can collectively substantially influence a person's disease risk. Recent studies have now conclusively implicated both rare and common variation in common neurodevelopmental conditions, such as autism and schizophrenia.   The Deciphering Developmental Disorders (DDD) study has recruited >13,500 children with severe developmental disorders without a genetic diagnosis from clinical genetics departments around the UK. We are primarily using exome sequencing of the children and their parents to search for rare or *de novo* variants of strong effect, because these individuals are presumed to have Mendelian genetic disorders. Here, we test whether these children also have a higher burden of common variants associated with other neurodevelopmental traits by analysing GWAS data in 2,268 DDD cases and 9,275 healthy controls of European ancestry. We calculated polygenic risk scores (PRS) for a range of diseases from published GWAS summary statistics. For nearly all diseases, we saw no difference between DDD cases and controls (Kolmogorov-Smirnov test) p>0.05, which illustrated that our cases and controls were generally well matched. However, our DDD cases had a significantly higher burden of schizophrenia polygenic risk compared to controls (p=0.0028). Previous evidence of genetic overlap among neurodevelopmental disorders suggests that this PRS might serve as a proxy for neurodevelopment, and is well powered due to the large sample size of the underlying schizophrenia GWAS. These data suggest there is a polygenic component to developmental disorders at least in a subset of patients, and that there may be shared common genetic architecture among a range of neurodevelopmental disorders.

1449F

***MC1R* gene variation and fear related to dental care: Evidence of fear of pain as mediator.** *C.L. Randall[1,2], D.W. McNeil[1,2,3], J.R. Shaffer[2,4], R.J. Crout[2,5], R.J. Weyant[2,6], M.L. Marazita[2,4,7,8,9,10].* 1) Department of Psychology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV; 2) Department of Dental Practice and Rural Health, School of Dentistry, West Virginia University, Morgantown, WV; 3) Center for Oral Health Research in Appalachia; 4) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 5) Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV; 6) Department of Dental Public Health, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 7) Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA; 8) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 9) Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 10) Department of Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA.

Anecdotal evidence in medicine and dentistry has long suggested that individuals with red hair are hypersensitive to pain and more likely to be anxious or fearful. Consistent with this persistent clinical lore, recent investigations have confirmed that variation in the melanocortin-1 receptor (*MC1R* [MIM 155555]) gene, which has approximately 97% prevalence in red-haired and approximately 33% prevalence in dark-haired Caucasians, is associated with general acute pain sensitivity, dental pain sensitivity, and fears related to dental care. The study objective was to clarify genetic influences in the etiology of dental care-related fear, which is ubiquitous and, at high levels, impacts dental treatment-seeking behavior and oral and systemic health. Specifically, the present study aimed to: (1) replicate the finding that *MC1R* variation and dental care-related fear are associated, (2) determine for the first time whether *MC1R* variation is associated with general fear of pain, and (3) determine whether fear of pain plays an intermediary role in the association between *MC1R* variant status and dental care-related fear. Participants were 817 Caucasian adults (62.5% female, *M* age = 34.7 years, *SD* = 8.7) taking part in a larger, cross-sectional project that identified determinants of oral diseases at the community-, family-, and individual levels (Center for Oral Health Research in Appalachia, cohort 1; COHRA1). Six SNPs on *MC1R* (rs1805006, rs11547464, rs1805007, rs1110400, rs1805008, rs1805009) were genotyped for participants who also completed self-report measures of dental care-related fear and fear of pain. Variation at one or more *MC1R* SNPs predicted higher levels of dental care-related fear and fear of pain. Importantly, fear of pain fully mediated the relation between *MC1R* variant status and dental care-related fear, B = 1.60, 95% CI [0.281, 3.056]. *MC1R* variation may influence orofacial/dental pain and, in turn, predispose individuals to develop fears about pain. Such pain-related concerns may contribute to the genesis of fears about dental treatment. This study provides support for genetic influences in the etiologies of fear of pain and dental care-related fear, and offers directions for future research that could identify biomarkers of dental pain sensitivity, fear of pain, and dental care-related fear. R01-DE014899, F31 DE023493.

1450W

Investigating the prevalence of treatable genetic disorders with neuropsychiatric symptoms in schizophrenia and bipolar populations. *V. Sriretnakumar[1,2], K. Mittal[1], R. Harripaul[3,4], J. Kennedy[1,3,5], J. So[1,2,5,6].* 1) Neuropsychiatry Genetics, Center for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada; 2) Laboratory Medicine and Pathobiology (LMP), Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 3) Institute of Medical Science (IMS), Faculty of Medicine, University of Toronto, Ontario, Canada; 4) Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 6) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada.

   Background: Many rare genetic syndromes are known to phenotypically manifest with psychiatric symptoms that can be indistinguishable from primary psychiatric disorders. While the majority of ongoing psychiatric genetic research has been dedicated to the identification and characterization of genes involved in primary psychiatric disorders, little research has been done to determine the extent to which rare genetic variants contribute to the overall psychiatric disease load. Within schizophrenia and bipolar populations, we are conducting the first study of its kind to determine the prevalence of four treatable genetic syndromes (Niemann Pick disease type C (NPC), Wilson disease, acute intermittent porphyria (AIP), and homocystinuria (HOM)) manifesting as primary psychiatric disorders. **Methods:** We are screening 1323 schizophrenia and 1200 bipolar disorder samples, along with 980 sex- and age-matched healthy controls, all with available DNA and extensive phenotype data. We are using a matrix-type pooled targeted deep sequencing of the genes *NPC1, NPC2, ATP7B, HMBS,* and *CBS* to screen for the 4 genetic diseases. Pathogenic variants within the targeted genes will be identified using an in-house analytic pipeline with quality control, variant discovery designed specifically for identifying variants in the matrix pooled targeted sequencing approach, and functionality prediction programs to determine variant pathogenicity. Sanger sequencing will be used to validate identified mutations and decrease false positive calls. **Results:** 1024 schizophrenia samples have been sequenced (average read depth = 468X per sample, average read length=190bp, 98.8% sequencing accuracy). uSequencing of an additional 1024 schizophrenia and bipolar samples, as well as, variant analysis of the already sequenced samples is currently underway. We hypothesize that a significant sub-population of patients with psychiatric disorders have underlying rare genetic conditions. **Conclusions:** Screening for treatable genetic diseases, such as NPC, WD, AIP and HOM, within schizophrenia and bipolar samples could provide a possible explanation for severe treatment resistance and treating the genetic condition can effectively "cure" patients of their otherwise difficult-to-treat psychiatric symptoms. Ultimately, this proof-of-principle study will lead to the development of molecular diagnostic tools for detection of underlying genetic disorders in psychiatric patients and will allow for precision medicine.

1451T

Novel genetic variants in X-linked genes identified in children with apraxia of speech. *C.M. Stein[1,2], E.R. Chan[1,3], B. Truitt[1], H. Voss-Hoynes[1], L. Freebairn[5], J. Tag[5], J. Vick[5], H.G. Taylor[6], B.A. Lewis[5], S.K. Iyengar[1,4].* 1) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Center for Proteomics & Bioinformatics, Case Western Reserve University, Cleveland, OH; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 4) Department of Genetics & Genome Sciences, Case Western Reserve University, Cleveland, OH; 5) Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH; 6) Department of Pediatrics, Case Western Reserve University, Cleveland, OH.

   Childhood apraxia of speech (CAS) is a severe communication disorder that affects 1-2/1000 children in which the precision and consistency of movements underlying speech are impaired in the absence of known neuromuscular deficits. We have found that cases of CAS are predominantly sporadic in families, that is, only one child diagnosed with CAS per nuclear pedigree. Thus, we hypothesized that such a rare and severe disorder could be caused by novel genetic variants, so we conducted whole genome sequencing of 21 cases of CAS and 21 unaffected family members from 21 CAS families. Whole genome sequencing was performed using DNA isolated from blood using PCR-Free sample preparation with Illumina HiSeq X™ Ten v2 chemistry run to 30X Mean Coverage. Reads were aligned to hg19 and variants called using the GATK pipeline. Initial filtering for exonic variants without rs numbers in dbSNP only present in CAS cases and not in unaffected siblings led to 105 novel variants occurring in 79 unique genes. Of these 12 were highly deleterious and non-synonymous and occurred in brain/neural pathways or morphological development, thus having potential relevance to CAS. Five of these genes, *FLNA, RAI2, HUWE1, PLXNB3,* and *VCX3B*, are X-linked, which may explain the sex ratio distortion seen in CAS and other communication disorders. Deficits in *FLNA* have been associated with syndromes manifested by hearing loss and problems in development of the palate. Similarly, deficits in *RAI2* have been associated with syndromes manifested by visual impairment, intellectual deficit, and psychomotor delay. *HUWE1* has been associated with mental retardation and intellectual disability. *PLXNB3* plays a role in axon guidance and cell migration, and has been associated with verbal performance. *VCX3B* has also been associated with intellectual disability and psychomotor delay. Multiple novel variants also occurred in *PCDHA10*, a protocadherin involved in cell-cell connections in the brain. Unexpectedly, there were no known or novel variants in CAS cases in *FOXP2*, a gene previously identified for dyspraxia of speech and presumed to be the most common genetic cause of CAS. Our findings illustrate the genetic heterogeneity underlying CAS, reflecting its phenotypic complexity, thus posing challenges for clinical diagnosis and classification. Furthermore, these findings broaden our knowledge of the genetic basis of human speech development.

1452F

ADHD, IQ, and inherited deletions in Canadian children with 22q11.2 deletion syndrome. *E. Chow[1,2], T. Leung[2], D. Young[2], A. Rideout[3], S. Dyack[3,4].* 1) Clinical Gen Service, Ctr Addiction & Mental Hlth, Toronto, ON, Canada; 2) University of Toronto, ON, Canada; 3) IWK Health Centre, Halifax, NS, Canada; 4) Dalhousie University, Halifax, NS, Canada.

Background: Attention Deficit Hyperactivity Disorder (ADHD) is common in children with 22q11.2 deletion syndrome (22q11DS), with reported prevalence rates between 31% to 55%. ADHD may be more common in the general population in individuals with intellectual disability. We aim to examine the prevalence of ADHD in a group of Canadian children with 22q11DS and to report on their IQ, medication usage and outcomes on follow-up. Methods: 41 children with 22q11DS (19M, 22F) and between 8 to 12 years old (mean=10.1 years, SD =1.8 years) were recruited from Canadian medical genetic clinics in Toronto and Halifax. All subjects have had FISH or microarray studies confirming a microdeletion in the 22q11.2 region. The 22q11.2 deletion was known to be inherited in 10 of these subjects. All subjects were interviewed with a parent using K-SADS-PL to assess for DSM-IV ADHD. They also had their IQ assessed using WISC-III/IV. Subjects diagnosed with ADHD were re-assessed approximately 5 years later. Results: The baseline prevalence of ADHD in this cohort was 31.7% (n=13). The mean full-scale IQ (FSIQ) for the whole group was 67.2 (SD=11.7). Average FSIQ for subjects with ADHD was non-significantly lower than that for subjects without ADHD. Subjects with inherited deletions were more likely to have ADHD than the rest of the group (46.2% vs 14.3%, p=0.049), although their mean FSIQ did not differ significantly. The most common ADHD subtype was the combined type (n=8, 61.5%). In 6 of the 13 subjects (46.2%), ADHD was a new diagnosis, made during this research assessment. Subjects previously diagnosed with ADHD were treated with stimulants (n=5), risperidone (n=2), atomoxetine (n=1), and no medications (n=1) at time of initial assessment. Ten of the 13 subjects with ADHD at initial assessment were reassessed at an average of 5.1 years later. ADHD Not Otherwise Specified was the most common type (n=4, 40%) at reassessment. Only 4 out of 10 subjects were on medications (all stimulants) at reassessment. Reasons for not accepting medications included side effects (n=2), medication felt not to be needed (n=2), psychosis (n=1) and parental refusal (n=1). Conclusions: ADHD is common in 22q11DS in Canada, but it is not always diagnosed in a timely fashion nor treated with medications. Medical geneticists can assist in prompt identification and management of ADHD in 22q11DS by having a higher index of suspicion, especially in children with inherited deletions.

1453W

Finding causal variants in the GWAS-implicated *MIR137* locus for schizophrenia. *J. Szatkiewicz[1], X. Zheng[1], S. Williams[1], K. Sakamoto[1], N. Ancalade[1], P. Giusti-Rodríguez[1], M. de Ussel[1], A. Collins[1], Y. Li[1], P. Sullivan[1,2].* 1) University of North Carolina , Chapel Hill, NC; 2) Karolinska Institutet, Stockholm, Sweden.

Schizophrenia (SCZ) is a common and devastating neuropsychiatric illness. Finding causal variants is essential for understanding SCZ etiology. The largest GWAS to date successfully identified common risk variants in 108 loci, but the identification of causal variants remains elusive. One of the strongest association loci spans *MIR137*, a microRNA gene important for neuronal function. We conducted a set of fine-mapping experiments to identify both novel and known variants that are the best candidates for functionally affecting MIR137 or other genes in the same locus, and thereby most likely contributing to SCZ risk. Our discovery sample included 280 SCZ cases and 280 controls from the Swedish Schizophrenia Study (S3). First, we deeply sequenced the 750kb LD region surrounding *MIR137* in the discovery sample to generate high-quality variant calls for SNPs, Indels, and CNVs. Using the same discovery sample, we also genotyped a known variable number tandem repeat (VNTR), 6bp upstream of the pre-*MIR137* gene. We integrated SNPs, Indels, CNVs, and VNTR to construct phased reference haplotypes. Next, we imputed the reference haplotypes into the full S3 sample, carried out experimental validation of the imputed data, performed association testing using imputed dosages, and conducted functional annotation using public databases and our own functional data on the region. Our resequencing identified 3061 novel SNPs not present in the 1000 Genomes database. Association testing based on SNPs identified two refined associations beyond the original GWAS in gene *DPYD* (rs61787785, rs75641275) and *MIR137* (rs1198577, rs10875125), both reported previously. Association testing based on VNTR suggests that this variant is unlikely to alter SCZ risk. Our deep resequencing data, imputation reference, and VNTR imputation method are useful resources for the community.

1454T

Refining the association between *de novo* **copy number variation and ASD risk.** *E.M. Wigdor[1,2], J.A. Kosmicki[1,2,3], D.J. Weiner[1,2], K.E. Samocha[1,2,4], S.J. Sanders[5], M.J. Daly[1,2], E.B. Robinson[1,2].* 1) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Program in Bioinformatics and Integrative Genomics, Harvard University, Cambridge, MA; 4) Program in Genetics and Genomics, Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 5) University of California, San Francisco, San Francisco, CA.

De novo copy number variants (dnCNVs) are significantly associated with risk for autism spectrum disorders (ASDs; Sanders et al., 2015). The purpose of this study was to identify and interpret subgroups of dnCNVs, particularly de novo deletions, which are responsible for the majority of this association. Using data from the Simons Simplex Collection (SSC), we found that deletions including a gene intolerant of heterozygous loss of function variation (from here: constrained genes; Lek et al., bioRxiv 2015) are seen more than 4 times as frequently in ASD cases compared to controls (Carrier Ratio (CR)=4.12, p=5.20e-07). After accounting for the presence of a constrained gene, deletion size and case status are no longer associated (p=0.72). Large (>500kb) deletions lacking a constrained gene are similarly overrepresented in cases compared to controls (CR=6.50, p=0.05), but are rare as most (>80%) large deletions include a constrained gene. These two categories form our strong acting deletions group (total CR=4.3, p= 9.13e-08) and can be found in approximately 2.6% of ASD cases and 0.6% of controls. The remaining deletions (not in one of these two categories) can be found in 1.9% of cases and 1.5% of controls. These show no association with ASD risk (CR=1.22, p=0.43). The associations are not meaningfully different in models controlling for sex or parental age. Highly damaging de novo variants associated with ASD risk are consistently more common in ASD cases with low IQ. We see a strong association between ASD proband IQ and rate of de novo deletions in the strong acting category (p=0.002), but no association between proband IQ and the rate of remaining de novo deletions (p=0.53). As previously shown with de novo protein truncating variants (Kosmicki et al., biorRxiv 2016), strongly contributing de novo deletions are overrepresented in female ASD cases (p<0.05). This is not true for the remaining de novo deletions. In contrast to the deletion findings, we do not see evidence that constrained genes are any more associated with ASD risk when duplicated than other genes (p=0.92). This finding suggests that duplication of highly constrained genes is generally well tolerated. These analyses suggest that risk conferred by de novo deletions primarily reflects the loss of function of a constrained gene. This finding will improve interpretation and functional follow up studies of risk-conferring copy number variants.

1455F

Genome wide interaction between SNPs and global methylation in conferring risk for suicide in schizophrenia. *V. De Luca, A. Bani Fatemi, R. Raymond, J. Nobrega.* CAMH, Toronto, Canada.

Background: Suicide attempt in schizophrenia is an important clinical issue. We performed a genome-wide association study to identify genetic markers, which increase the risk for suicide attempt in schizophrenia. **Methods:** Suicide attempt lifetime was assessed in 433 schizophrenia patients and defined by the means of the Columbia Suicide Severity Rating Scale and the Beck Scale for Suicidal Ideation. Genotype distribution of 1,205,383 single nucleotide polymorphisms (SNPs) in patients with suicide attempt lifetime was compared to that in patients without any suicide attempt lifetime. The same SNPs were analyzed in interaction with global DNA methylation in white blood cells. **Results:** None of the variants reached genome-wide significance, the best marker in the main genetic model was the SNP rs12895203 (p=0.00001) and the top SNP interacting with global methylation was the marker rs7897059 (p=0.00005). The odd-ratio of the top SNP in the main-genetic effect model was 3.91 and in the gene-methylation interaction model was 1.13. We also found an interesting locus effect on chromosome 15 in controlling the global methylation variation. **Conclusions:** Although our data need to be interpreted carefully owing to the small numbers in this cohort, they suggest that a combination of genetic markers and global DNA methylation might be used to identify patients at risk for suicide attempt.

1456W

De novo germline and nodular heterotopia-associated postzygotic mutations of STXBP1 in an epilepsy patient successfully treated with resective surgery. *M. Uddin[1,2], C. Boelman[3], M. Woodbury-Smith[4], A. Chan[1,2], L. Brunga[2], S. Lamoureux[1,2], D. Stavropoulos[5], J. Drake[6], C. Hahn[7], C. Hawkins[8], A. Shlien[2], B. Minassian[2,6], S.W. Scherer[1,2,9,10].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, CanadaOntario, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Division of Neurology, BC Children's Hospital, Vancouver, BC, Canada; 4) Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, Ontario, Canada; 5) Genome Diagnostics, Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Division of Neurosurgery, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Division of Neurology, The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Division of Pathology, The Hospital for Sick Children, Toronto, Ontario, Canada; 9) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 10) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

Syntaxin-binding protein 1 (STXBP1) mutations have been reported across a spectrum of neurodevelopmental disorders, including Otahara syndrome and other epileptic encephalopathies. In order to delineate the phenotypic complexity of cases, it is necessary to analyze the STXBP1 mutation with consideration for the sequence context. Here, we report the first case of a successfully treated epilepsy surgery patient with a heterozygous STXBP1 germline mutation that was mosaic within a surgically resected area populated with dysplastic cells. Patients undergoing resective epilepsy surgery for refractory focal epilepsy that was associated with suspected cortical dysplasia were recruited prospectively for genetic study. For the current patient, a targeted sequencing approach was applied to identify rare pathogenic mutations. DNA from peripheral blood was collected from the nuclear family. DNA was also collected from resected left lateral temporal cortex tissue with dysplastic neurons. We applied droplet digital PCR (ddPCR) technology to validate the copy number variation (CNVs). We have identified a *de novo* 4.9Kb deletion impacting exon 3-4 of the STXBP1 gene in a 6-year old proband (male) diagnosed with refractory focal-onset epilepsy with initial infantile spasms, global developmental disorder, autism spectrum disorder and MRI imaging suspicious for focal cortical dysplasia type 1 in the left anterior temporal cortex. The patient underwent left anterior temporal lesionectomy including mesial structures. Surgical pathology demonstrated nodular heterotopias in the superior temporal gyrus and deep temporal white matter with collections of cells with small round nuclei and clear cytoplasm. Hippocampus, para-hippocampal gyrus and lateral temporal cortex were normal. The ddPCR result revealed heterozygous state of the deletion in blood, whereas, evidence of the mosaic presence of heterozygous and homozygous deletions was observed within the dysplastic tissue cell population of the proband. The patient has been seizure free since surgery. We report a novel *de novo* mutation within STXBP1 in a patient with a complex neurodevelopmental phenotype and focal epilepsy with areas of dysplastic neurons. This is the first case of STXBP1 showing formation of mosaic copy number variations within dysplastic brain cells. These findings suggest a role for STXBP1 in malformations of cortical development and a role for epilepsy surgery in the management of STXBP1-associated epilepsy.

1457T

TROVE2 alleles are linked to aversive memory: A whole exome sequencing study in phenotypic extremes. *A. Heck[1,2], A. Milnik[1,2], V. Vukojevic[1,3], J. Petrovska[1], T. Egli[1], J. Singer[4], P. Escobar[5], T. Sengstag[5], D. Coynel[6], V. Freytag[1], P. Demougin[3], E. Loos[6], F. Hartmann[1], N. Schicktanz[6], C. Vogler[1,2], C. Beisel[7], N. Beerenwinkel[4], D. de Quervain[2,3], A. Papassotiropoulos[1,2,3].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Basel, Switzerland; 2) University of Basel, Psychiatric University Clinics, Basel, Switzerland; 3) University of Basel, Department Biozentrum, Life Sciences Training Facility, Basel, Switzerland; 4) ETH Zurich, Department of Biosystems .Science and Engineering, Basel, Switzerland; 5) University of Basel, sciCORE Center for Scientific Computing, Basel, Switzerland; 6) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Basel, Switzerland; 7) ETH Zurich, Quantitive Genomics Facility, Basel, Switzerland.

Over the last decade, whole exome sequencing (WES) has become a standard approach for the identification of rare variants implicated in monogenetic disease. However, also for complex traits, WES gained popularity in the last years[1,2]. By sequencing only individuals at the extreme ends of an empiric phenotypic distribution, WES offers a cost-effective still statistically powerful method, based on the assumption that rare alleles are enriched in the extreme end of the distribution. We followed this approach and sequenced 94 individuals with extremes in aversive memory performance. WES was performed using the SureSelectXT Human All Exon V5+UTR kit (Agilent) on an Illumina HiSeq 2500 machine. After alignment (BWA) and de-dupping (Picard), further preprocessing was done with GATK. The mean sample coverage was 111X, with 94% of the target regions having a coverage equal or greater than 30X. Aversive memory was initially tested in 2'739 healthy young individuals with European background with a picture-based memory test. After matching for sex, age, smoking and genetic background, 47 pairs (i.e. 47 high and 47 low performers) were sequenced. We performed burden tests with PlinkSeq and restricted our analysis to variants showing a minor allele frequency < 0.125 in the population of phenotypic extremes. We identified one Bonferroni-corrected association between *TROVE2* and aversive memory performance (p=2.0E-6). One SNP, rs72740218, drives the association, with 10 minor allele carriers in the high-performer group versus 2 minor-allele carriers in the low-performer group. This association could be validated in an independent sample (n=2'590). In an fMRI study in a subsample of probands (n=1'258) the minor allele of SNP rs727403218 was positively correlated with brain activation in the frontal cortex (t=5.4, P_{FWE}=0.0015) during the encoding of aversive pictures. Next, we analyzed the public available data of the BRAINEAC project. In these data, SNP rs727403218 SNP was associated with expression of *TROVE2* transcript 2372955 in post-mortem frontal cortex (p=0.005). By combining WES with an extreme phenotype approach, we provide first evidence that *TROVE2* plays a role in aversive memory performance. [1] Frazer, K.A., *et al.* (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10, 241-251 [2] Cirulli, E.T. and Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11, 415-425.

1458F

Large Scandinavian exome sequencing study of schizophrenia. *F. Lescai[1,2,3], F.K. Satterstrom[4,5], T. Als[1,2,3], J. Grove[1,2,3,6], J.B. Maller[4,5], J. Grauholm[1,7], C. Stevens[4,5], M. Mattheisen[1,2,3], R. Walters[4,5], J.I. Goldstein[4,5], S.A. McCarroll[4,5], D.M. Hougaard[1,7], T.M. Werge[1,8,9], B.M. Neale[4,5,10], M.J. Daly[4,5,10], A.D. Børglum[1,2,3], iPSYCH-Broad Consortium.* 1) Biomedicine, Aarhus University, Aarhus, Denmark; 2) iPSYCH, The Lundbeck Initiative for Integrative Psychiatric Research; 3) iSEQ, Centre for Integrative Sequencing; 4) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 6) BiRC, Bioinformatics Research Centre, Aarhus University, Denmark; 7) Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 8) Mental Health Centre Sct. Hans, Institute for Biological Psychiatry, Capital Region of Denmark, Roskilde, Denmark; 9) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 10) Department of Medicine, Harvard Medical School, Boston, MA, USA.

The allelic architecture of complex traits is likely to be underlined by multiple common and rare variants. Large-scale consortia GWAS meta-analyses have successfully been applied in the search for common variants affecting the risk of developing psychiatric disorders, in particular schizophrenia. However, these studies are designed to examining only "the common variant" proportion of the genomic landscape, while next-generation-sequencing allows direct examination of both common and rare alleles. Schizophrenia is one of the most disabling mental disorders. It has a lifetime prevalence of around 1 % and a heritability estimated as high as 80%. Although around one third of the heritability can be attributed to common variants, it is likely that rare variants also contribute substantially to the etiology. A collaboration between iPSYCH (The Lundbeck Initiative for Integrative Psychiatric Research, Denmark) and the Broad Institute, Boston, has been established to perform large scale exome sequencing of major psychiatric disorders, including schizophrenia. The DNA is extracted from blood samples existing in the Danish Neonatal Screening Biobank, which has collected dried blood spots (DBS) for every child born in Denmark since 1981. The samples available (totaling more than 2.2 million) can be matched to a wide range of phenotypic information available in Danish registries, thus empowering the analysis of genetic data with a wealth of data from medical history as well as social and cultural environment. We present here an analysis of a large collection of schizophrenia individuals recruited in clinical centers (1,686) and from DBS samples (3,545): a total of 5,231 cases and 7,258 controls from Denmark. We follow-up in a similarly sized sample from Sweden; a related Scandinavian population. The combined analysis represents the largest exome sequencing study in schizophrenia to date. In a preliminary data freeze consisting of 3,200 exomes, we have found evidence suggesting a higher per-person rate of ultra-rare loss-of-function variants located in constrained genes among schizophrenia cases compared to controls. Likewise in the full data set, particular focus is devoted to the analysis of the polygenic burden of rare and ultra-rare variation. The results from burden and enrichment analysis, as well as ranking of mutational load, will be discussed in the larger perspective of findings from GWAS, and other exome analyses in schizophrenia.

1459W

Interim results from the GSCAN exome chip project. *D. Brazel[1], Z. Liu[1], D. Liu[2], S. Bertelsen[18], Y.L. Chou[3], S. David[4], J. Faul[5], J. Gong[6], A. Hammerschlag[7], C. Hsu[8], D. Irons[9], A. Jackson[10], A. Loukola[11], M. Mangino[12], G. Pistis[13], R. Rhode[14], Y. Shao[14], M. Steri[13], H. Stringham[10], M. Keller[1], V. Turcot[15], L. Wetherill[17], W. Zhao[16], S. Vrieze[1], CHD Exome+ Cons., Cons. for Gen. of Smoking Behav., GWAS & Seq. Cons. of Alcohol and Nicotine.* 1) Institute for Behavioral Genetics, University of Colorado - Boulder, Boulder, CO; 2) College of Medicine, Penn State University, State College, PA; 3) Department of Psychiatry, Washington University, St. Louis, MO; 4) Stanford University Medical Center, Stanford University, Stanford, CA; 5) Institute for Social Research, University of Michigan, Ann Arbor, MI; 6) Fred Hutchinson Cancer Center, Seattle, WA; 7) Center for Neurogenomics & Cognitive Research, VU University of Amsterdam, Amsterdam, Netherlands; 8) Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 9) Psychology Department, University of Minnesota, Minneapolis, MN; 10) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 11) Department of Public Health, University of Helsinki, Helsinki, Finland; 12) Twin Research & Genetic Epidemiology Unit, King's College London, London, UK; 13) Institute of Neurogenetics and Neuropharmacology, Consiglio Nazionale delle Ricerche, Rome, Italy; 14) Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 15) Montreal Heart Institute, Montreal; 16) School of Public Health, University of Michigan, Ann Arbor, MI; 17) Department of Medical and Molecular Genetics, Indiana University-Purdue University at Indianapolis, Indianapolis, IN; 18) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, Manhattan, NY.

The use and abuse of alcohol and nicotine has a significant impact on public health. Twin and family studies show that these behaviors have a significant genetic component. Genetic association studies have discovered common variants associated with alcohol and nicotine. The exome chip portion of the GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) was formed to examine the association between rare nonsynonymous variants and alcohol and nicotine use, by aggregating and meta-analyzing studies using exome sequencing or genotyping arrays with substantial rare exonic content. Recently, we have added the UK Biobank cohort. The UK Biobank has collected genetic and phenotypic data from 500,000 volunteers in the United Kingdom. 92% of the sample was genotyped on a custom platform, the UK Biobank Axiom array, which contains 111,162 rare nonsynonymous variants and 280,838 additional rare variants. The remainder were genotyped on the UK BiLEVE array which has a high degree of marker overlap with the Axiom array. Genetic data from 150,000 individuals is currently available with the remainder expected by the end of 2016. We performed extensive quality control and phenotype standardization, followed by rare variant association analyses, separately in the BiLEVE and Axiom subsamples, for five phenotypes: smoking initiation (total N=112,811), cigarettes per day (total N=40,882), age of initation of smoking (total N=41,634), pack years (total N=40,624), and drinks per week (total N=90,213). Genome-wide significant rare variant associations were found in the classic *CYP2A6* locus for cigarettes per day and in the alcohol metabolism genes *ADH1B* and *ADH1C*. A meta-analysis was then performed between BiLEVE and Axiom. Although these data are not available at this time, results will be presented from a meta-analysis of the UK Biobank sample with the other samples in the GSCAN exome chip consortium, adding approximately 100,000 individuals to our analysis and including both single variant and gene based tests.

1460T

Conduct disorder GWAS and extension to corticolimbic function. *C.E. Carey[1], N. Vakkalagadda[1], A. Agrawal[2], E. Drabant Conley[3], A.R. Hariri[4], E.C. Nelson[2], R. Bogdan[1].* 1) Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO; 2) Psychiatry, Washington University School of Medicine, St. Louis, MO; 3) 23andMe, Mountain View, CA; 4) Psychology and Neuroscience, Duke University, Durham, NC.

Conduct disorder (CD), a heritable childhood psychiatric illness characterized by high levels of aggression and rule-violating behavior, is a strong predictor of violence in adulthood. However, little is known about the individual genetic variants conferring risk for CD, or the neural mechanisms that might mediate associations between genetic risk and antisocial behavior. To examine this, we performed a genomewide association of retrospectively reported CD in the Comorbidity and Trauma Study (n_{cases} = 680, $n_{controls}$ = 995), which yielded one genomewide significant hit: rs12536973 in an intergenic region on chromosome 7. We then probed the effects of this genetic risk on Self Report of Psychopathy Short Form (SRP-SF) scores and corticolimbic function during an emotional face-matching task among 404 non-Hispanic Caucasian young adult participants in the Duke Neurogenetics Study. rs12536973 genotype was associated with the SRP-SF Lifestyle subscale, but not with corticolimbic function. Genomewide polygenic risk scores predicted greater SRP-SF Antisocial scores, as well as decreased insula reactivity to emotional faces. These results provide insight into the neurogenetic underpinnings of CD and antisocial behavior and may, in the future, help to inspire interventions to prevent children with CD from committing violent acts as adults.

1461F

Genome-wide association study of the traits defining the three-factor model of post-traumatic stress disorder symptomology reveals novel risk variants. *R. Sherva[1], H.R. Kranzler[2], R.H. Pietrzak[3,8], L.A. Farrer[1,4,5,6,7], J. Gelernter[8,9,10,11].* 1) Biomedical Genetics Department, Boston University School of Medicine, Boston, MA; 2) Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania and VISN 4 MIRECC, Philadelphia VAMC, Philadelphia, PA; 3) Translational Psychiatric Epidemiology Laboratory, Clinical Neurosciences Division, U.S. Department of Veterans Affairs National Center for PTSD, New Haven, CT; 4) Department of Neurology, Boston University School of Medicine, Boston, MA; 5) Department of Ophthalmology, Boston University School of Medicine, Boston, MA; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 7) Department of Epidemiology, Boston University School of Public Health, Boston, MA; 8) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 9) VA Cooperative Studies Program Coordinating Center, West Haven, CT; 10) VA CT Healthcare Center, West Haven, CT; 11) Department of Neurobiology, Yale School of Medicine, New Haven, CT.

Posttraumatic stress disorder (PTSD) is a potentially serious psychiatric disorder that can develop after a traumatic event. Individuals vary in their susceptibility to PTSD following similar exposures, and in addition to demographic factors and other environmental risk factors, approximately 40% of the variability in risk is attributable to genetic risk factors. Several genome-wide association studies have been performed using PTSD diagnosis as a trait. The results have been mixed and inconsistently replicated. We performed a GWAS for each of three major domains of DSM-IV PTSD symptoms: 1) re-experiencing the event (defined as the sum of the number of symptoms related to experiencing intrusive thoughts, nightmares, flashbacks, and reactivity to reminders of the trauma); 2) avoidance and emotional numbing (the sum of symptoms of avoidance of stimuli associated with the trauma and anhedonia); and 3) hyperarousal (the sum of symptoms associated with anger, hypervigilance, exaggerated startle response, and sleep disruption). The sample included 2,561 trauma-exposed African Americans and European Americans recruited for studies of addiction, including 1,198 with DSM-IV PTSD diagnoses. We identified 3 novel regions with genome-wide significant SNP associations for the re-experiencing score, one for avoidance/numbing, and 7 for hyperarousal. The SNPs associated with re-experiencing symptoms were located in an overlapping pseudogene and antisense transcript (*RN7SKP83* and *AC105053.3*, respectively, p=4.78×10^{-8}), cilia and flagella associated protein 77 (*CFAP77*, p=3.67×10^{-8}), and an intergenic region between an antisense transcript and pancreatic ribonuclease *(RP11-219E7.1* and *RNASE1*, respectively, p=4.56×10^{-8}). The SNPs associated with avoidance/numbing symptoms were located between two processed pseudogenes (*RP11-550A9.1* and *RP11-448K10.1*, p=2.69×10^{-8}). The SNPs associated with hyperarousal include variants in *RP5-952N6.1*, *DNAJC13*, *ANXA3*, *GDPD4*, and *USP3*; between *snoU13* and *CTD-2037L6.2*; and a missense variant in *ZNF217*. Two of the signals are significant after correcting for testing three outcomes (*USP3*, p=9.16×10^{-9} and *ZNF217*, p=1.49×10^{-11}). These findings require replication but illustrate: 1) the utility of PTSD symptom cluster scores for genetic analysis; 2) hyperarousal symptoms may have the strongest genetic influence; and 3) genetic risk factors for PTSD symptomology may be a combination of coding and regulatory variants in poorly understood pathways.

1462W

Genetic influences on normative and problematic alcohol use in a population-based sample of college students. *B.T. Webb[1,2], A.E. Edwards[1,2], A.R Woolen[4], J.E. Salvatore[1,5], F. Aliev[6,7], B.P. Riley[1,2,3], C. Sun[1,2], V.S. Williamson[1], J.N. Kitchens[1], K. Pedersen[2], A. Adkins[5,8], M.E. Cooke[1], J.E. Savage[1], Z. Neale[5], S.B. Cho[6], D.M. Dick[3,5,6,7], K.S. Kendler[1,2].* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US; 2) Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US; 3) Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, US; 4) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA, US; 5) Department of Psychology, Virginia Commonwealth University, Richmond, VA, US; 6) Department of African-American Studies, Virginia Commonwealth University, Richmond, VA, US; 7) Faculty of Business, Karabuk University, Turkey; 8) College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, VA, US.

Background: Genetic factors impact alcohol use behaviors and these factors may become increasingly evident during emerging adulthood. Examination of the effects of individual variants as well as aggregate genetic variation can clarify mechanisms underlying risk. Methods: We conducted genome-wide association studies (GWAS) in an ethnically diverse sample of college students for three quantitative outcomes including typical monthly alcohol consumption, alcohol problems, and maximum number of drinks in 24 hours. While these outcomes are moderately heritable in older adults, their heritability is known to change with age. Therefore, heritability based on common genetic variants (h^2_{SNP}) was assessed for this age group. We also evaluated whether risk variants in aggregate were associated with alcohol use outcomes in an independent sample. Results: Two genome-wide significant markers were observed: rs11201929 in *GRID1* for maximum drinks in 24 hours, with supportive evidence across all ancestry groups; and rs73317305 in *SAMD12* (alcohol problems), tested only in the African ancestry group. The h^2_{SNP} estimate was 0.19 (SE=0.11) for consumption, and was non-significant for other outcomes. Genome-wide polygenic scores were significantly associated with alcohol outcomes in an independent sample. Conclusions: These results robustly identify genetic risk for alcohol use outcomes at the variant level and in aggregate. We confirm prior evidence that genetic variation in *GRID1* impacts alcohol use, and identify novel loci of interest for multiple alcohol outcomes in emerging adults. These findings indicate that genetic variation influencing normative and problematic alcohol use is, to some extent, convergent across ancestry groups, and is relevant though modest during emerging adulthood. Studying college populations represents a promising avenue by which to obtain large, diverse samples for gene identification.

1463T

Anorexia nervosa genetic risk score. *Z. Yilmaz[1], S.C. Zerwas[1], L.M. Thornton[1], J. Euesden[2], L.E. Duncan[3], N. Micali[4], L. Huckins[4], J.H. Baker[1], M.A. Munn-Chernoff[1], J.J. Crowley[5], PGC Anorexia Nervosa Working Group[6], E. Zeggini[7], P.F. Sullivan[1,5,8], G. Breen[9], C.M. Bulik[1,8,10].* 1) Department of Psychiatry, University of North Carolina, Chapel Hill, NC; 2) Social Genetic & Developmental Psychiatry, King's College London, London, United Kingdom; 3) Department of Psychiatry, Stanford University, Palo Alto, CA; 4) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Genetics, University of North Carolina, Chapel Hill, NC; 6) Psychiatric Genomics Consortium; 7) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 9) Institute of Psychiatry, King's College London, London, United Kingdom; 10) Department of Nutrition, University of North Carolina, Chapel Hill, NC.

Anorexia nervosa (AN) is a serious eating disorder which is heritable (~50-60%) and has the highest mortality rate of any psychiatric disorder. As genetic research on AN progresses, investigating the effects of a genetic risk score (GRS)—calculated as the weighted sum of common genetic risk variants per individual—on developmental and clinical characteristics is a logical next step. Thus, we calculated GRS in a discovery dataset of 2,464 female AN cases and 7,355 female controls from the Genetic Consortium for Anorexia Nervosa (GCAN)/Wellcome Trust Case Control Consortium 3 (WTCCC3). The resulting AN GRS was then applied to a target dataset of 1,031 AN cases from the Price Foundation/Children's Hospital study and 3,627 pediatric controls recruited by Children's Hospital of Philadelphia. GRS was calculated using PRSice (http://prsice.info), with case-control significant principal components as covariates. The best-fit p-value threshold was determined to be 0.1384, and the AN GRS set at this threshold included 30,005 SNPs. The GRS model fit (Nagelkerke's R^2) from a logistic regression of AN case/control status was significant (p = 2.9 x 10^{-8}), and individuals with AN GRS in the highest decile had a significantly higher risk of AN compared to the individuals with AN GRS in the lowest decile (OR=2.1; 95% CI 1.6, 2.9). These results provide evidence for polygenicity in the etiology of AN and suggest that AN GRS has the potential to be an important tool in the development of multi-factorial risk prediction models in the future. However, larger samples with greater statistical power are required to increase the utility of AN GRS in contributing to the prediction of risk or illness course. As the Anorexia Nervosa Working Group of the Psychiatric Genomics Consortium adds a significant number of genotyped samples over the next two years (~20,000), AN GRS will become more robust and its utility will increase.

1464F

Postmortem transcriptomic profiling of three psychiatric disorders reveals widespread dysregulation of cell-type associated transcripts and novel disease-related transcription changes. *B.N. Lasseigne[1], K.M. Bowling[1], R.C. Ramaker[1,2], M.H. Hagenauer[3], N.S. Davis[1], J. Gertz[1], P.M. Cartagena[4], D.M. Walsh[4], M.P. Vawter[4], A.F. Schatzberg[5], J.D. Barchas[6], S.J. Watson[3], B.G. Bunney[4], H. Akil[3], W.E. Bunney[4], J.Z. Li[7], S.J. Cooper[1], R.M. Myers[1]. 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Department of Genetics, The University of Alabama at Birmingham, Birmingham, AL; 3) Mental Health Research Institute, University of Michigan, Ann Arbor, MI; 4) Department of Psychiatry and Human Behavior, College of Medicine, University of California, Irvine, Irvine, Ca; 5) Department of Psychiatry, Stanford University School of Medicine, Stanford, CA; 6) Psychiatry, Weill Cornell Medical College, New York, NY; 7) Department of Human Genetics, University of Michigan, Ann Arbor, MI.*

Psychiatric disorders are multigenic diseases with complex etiology contributing significantly to human morbidity and mortality. We compared molecular signatures across brain regions and disorders in the transcriptomes of postmortem human brain samples (Pritzker Neuropsychiatric Disorders Research Consortium). RNA sequencing was performed on tissue from the anterior cingulate cortex, dorsolateral prefrontal cortex, and nucleus accumbens from three groups of 24 patients each diagnosed with schizophrenia, bipolar disorder, or major depressive disorder, and from 24 control subjects. We validated 129 gene expression changes (FDR<0.05) with an independent cohort provided by the Stanley Neuropathology Consortium Integrative Database Array Collection. The most significant disease differences were observed in the anterior cingulate cortex of schizophrenia samples compared to controls and biochemical consequences of gene expression changes were assessed with untargeted metabolomic profiling. We also detected significant heterogeneity within schizophrenia and bipolar disorder samples with a subset of patients exhibiting gene expression profiles significantly different from both the larger subset of schizophrenia and bipolar disorder patients and from the major depressive disorder and control cohorts. We find no evidence this disease heterogeneity is linked to any known clinical phenotypes or technical variables, however virtual microdissection reveals high correlation with transcripts previously identified as specific to astrocyte- (p-value = 6.90x10-04) and endothelial-specific expression (p-value = 4.99x10-07), but depleted of neuron-specific expression (p-value = 3.77x10-09). We identify *NPAS4*, *IGKC*, and *IGHG1* as a refined list of putative cell-type independent transcripts associated with psychiatric disease. Chromatin immunopreciptation with massively parallel DNA sequencing of NPAS4 in human neuronal cell lines demonstrates enriched binding near GWAS-implicated psychiatric disorder associated genes (p-value = 6.66x10-16). Furthermore, knockdown of *NPAS4* resulted in reduced expression of a subset of those genes (*ABCB9*, *CHRNA3*, and *NDUFA6*), indicating cis-regulation by NPAS4. Here we provide a set of high confidence, independently validated genes differentially expressed between schizophrenia and control patients and propose a refined list of putative cell-type independent transcripts that may particularly warrant further investigation.

1465W

CNV deletions may affect schizophrenia risks through lncRNA genes that regulate protein-coding genes. *C. Liu[1,2], Q. Meng[2], K. Wang[2], Y. Xia[2], C. Jiao[2], C. Chen[2]. 1) Dept Psychiatry, Univ Illinois, Chicago, Chicago, IL., USA; 2) The State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China.*

Schizophrenia is a complex psychiatric disorder with strong genetic background. Studies demonstrated that rare copy number variations (CNVs) contribute to the risk of schizophrenia. However, the exact mechanisms remain unclear. More attention has been put to protein-coding genes residing in those CNVs. In this study, we explored the potential roles of long non-coding RNAs (lncRNAs) inside CNV deletions (CNV-lncRNAs) in schizophrenia risk. We retrieved lncRNAs mapped to the CNV deletion regions known to increase risks of developing schizophrenia. Seven such regions were repeatedly reported by large case-control studies, including 1q21.1, 3q29, 15q11.2, 15q13.3, 17p12, 17q12, and 22q11.2. We carried out weighted gene co-expression network analysis (WGCNA) using RNA-seq data from Genotype-Tissue Expression (GTEx) and BrainSpan projects to look for coexpression modules that harbor CNV-lncRNAs. We identified one male reproduction-related module in male individuals, one neuronal functions-related module both in male and female individuals. Protein-coding genes inside the neuronal functions enriched modules involved several ion channel activities including calcium and potassium channel activities, which were known to be related to schizophrenia. Pathway analysis of these two modules further suggested that CNV-lncRNAs involved in the olfactory transduction, neuroactive ligand-receptor and calcium signaling pathways. The CNV-lncRNA co-expression patterns were preserved through the different development and aging time points, and across different tissue types of human body. Our results suggest that lncRNAs inside those rare CNVs may play significant temporal and spatial roles in regulating other protein-coding genes, and subsequently contribute to schizophrenia risk. Future studies of those CNVs should pay more attention to the regulatory cascades instead of the few protein-coding genes inside them to fully appreciate the impact of those CNVs.

1466T

Exploring association of the mitochondrial genome with antipsychotic-induced weight gain. *K. Mittal[1,2], V. Goncalves[1,2], A.K. Tiwari[1,2], D.J. Mueller[1,2], M.P. Vawter[3], J.L. Kennedy[1,2].* 1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Department of Psychiatry, University of Toronto, ON, Canada; 3) Functional Genomics Laboratory, Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA.

Schizophrenia (SCZ) is a complex disorder characterized by psychosis with a lifetime prevalence of 0.4-1.0 %, and is ranked among top 10 causes of disability among people in developed countries. Antipsychotic medication treatment of schizophrenia often results in excessive weight gain, which can then lead to diabetes and cardiovascular diseases. Mitochondria play a vital role in brain metabolism since they are the main source of aerobic energy for cellular functioning. There is increasing evidence that antipsychotics such as clozapine and olanzapine modulate mitochondrial function, for example, by inhibiting activity of the oxidative phosphorylation pathway and altering its physiology. This study assessed the role of mitochondrial DNA (mtDNA) variants in conferring risk for Anti-Psychotic Induced Weight Gain (AIWG) in Schizophrenia patients. We analyzed 101 individuals from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). Only individuals of European ancestry were used for association analysis. The mitochondrial genome was sequenced in these samples and association analysis was conducted to assess 161 common variants and 502 rare variants. We analyzed coding region (577-16023 bp) as well as the control region (16024-576 bp) SNPs. The sequencing protocol was validated to work reliably on mtDNA, capturing all types of variation present in the mitochondrial genome, including common SNPs, small insertions/deletions, and heteroplasmy. The sequencing coverage was $\geq$500x average with >98% confidence in variant calling. Association between each common mitochondrial SNP and AIWG was tested using linear regression assuming an additive model and using treatment duration as covariate. We found nominal significance (p<0.05) for six mitochondrial SNPs, one in CO1 gene, one in ND3, and four in the mtDNA control region with anti-psychotic induced weight gain. The significance did not change after inclusion in the model of treatment duration as a covariate. The rare variants were analyzed using the SKAT protocol. No significant association with rare SNPs was detected. To our knowledge, this is the first study to address the role of mtDNA in anti-psychotic induced weight gain. Analysis in a larger sample is required and will allow a better understanding of the role of mitochondria in anti-psychotic induced weight gain.

1467F

Increased mitochondrial content in schizophrenia and functional implications. *L. Olde Loohuis[1], A. Ori[1], T. Wu[1], R. Kahn[2], R. Ophoff[1,3].* 1) Center for Neurobehavioral Genetics, University California Los Angeles, Los Angeles, CA; 2) 2)Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, The Netherlands; 3) 3)Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, California, USA.

Changes in the amount of mitochondrial DNA (mtDNA) have been linked to stress and inflammation, and increased levels of mtDNA have, among others, been reported in patients with major depressive disorder. While mitochondrial dysfunction has been suggested to play a role in the pathophysiology of schizophrenia, it is unknown whether patients also have increased mtDNA content. Here, we aim to test this hypothesis, and relate mtDNA content to transcriptome-wide expression patterns. Using whole exome sequencing (WES) data from whole blood of 96 cases and 96 matched controls, we quantify relative mtDNA content using the ratio of reads mapping to the mitochondrial reference genome and autosomal reads. In addition, we quantify mtDNA content in a (partially overlapping) sample of 96 cases and 96 controls using multiplex RT-qPCR. We observe a highly robust and significant increase in mtDNA content in cases (42% increase, P=5.7e-09, t-test on residuals). This difference is independent of age and gender and not due to technical artifacts. Using RT-qPCR, we measure a similar increase (P=1.8e-15), and observe a strong correlation between the two measurements in the overlapping sample of mainly controls (r2=0.43, P=2.5e7, n=48 controls, 7 cases), thus both validating and replicating this result. mtDNA content does not correlate with cell-type proportions as estimated from methylation data (n=52). Preliminary gene expression analysis using RNA sequencing (RNAseq) of the discovery sample shows that mtDNA levels do not correlate with increased expression of the any gene on the mitochondrial genome. However, expression of several autosomal genes is associated with mtDNA content, including MTND1P23 (β=0.06, P=3.3e-39) and IGLV7-43 (β=0.05, P=4.1e-12). We are the first to show blood-derived increased levels of mtDNA in schizophrenia, and are able to relate this finding to gene expression differences using RNAseq. The strong correlation between the two quantification methods (based on WES and RT-qPCR) highlights the value of WES to study mtDNA content across groups. The absence of a direct link between mitochondrial expression and mtDNA content is surprising, and may point to decreased functioning of the mitochondria. While we continue to explore the functional underpinnings of mtDNA content at the transcription level, to gain insight in the relation between mtDNA content and disease, we also study connections with genetic risk, disease severity and medication use.

1468W

Identifying functionally relevant mental health hotspots and polymorphisms in African Americans. *C. Cross[1], L. Jackson[2], M. Shestov[3].* 1) Anatomy Dept, College of Medicine, Howard University, Washington DC; 2) National Human Genome Center, College of Medicine, Washington DC; 3) Biomedical Engineering, Sciences and Health System, Drexel University, Philadelphia, PA.

   BACKGROUND: African Americans are 20 percent more likely to report serious psychological distress than European Americans. Recent work by our group suggests that genes sitting at the intersection of addiction and mental health form genomic hotspots that are functionally cohesive with drugable targets. This lead us to question how the mental health genes underlying schizophrenia, bipolar disorder, post-traumatic stress disorder, and depression form functional hotspots in the genome, whether they have unique allele frequencies in African ancestry populations, and whether those hotspots had drugable targets.   **METHODS:** Mental health associated gene sets were obtained from NCBI Gene and were projected onto gene ontology categories and cellular pathways to draw a bioinformatics portrayal of mental health disorders. Single nucleotide polymorphism datasets (HapMap, Human Genome Diversity Panel) were used identify sequence variants that were significantly different in African Americans from other populations. Finally we annotated the drug binding sites lying in each hotspot to predict the effects of these drugs on human populations.   **RESULTS:** Mapping addiction genes onto human genome resulted in eight gene clusters, with at least 20 mental health genes (Range: 21-46 genes/ hotspot) in a 4Mb distance along DNA. Hotspot genes were involved in neurological transmission, responses to organic substances, and cell-cell signaling. Analysis of hotspot drug binding sites show that mental health has a treatment overlaps with the functional annotations   **CONCLUSION:** This approach identified 10 key putative variants for mental health in individuals of African Ancestry, found functionally cohesive hotspots. Drugable targets annotation suggest that treatment could be better targeted (*in progress*).

1469T

Cascade TF-miRNA-mRNA regulations identified by co-expression network modules in brains of patients with schizophrenia and bipolar disorder. *C. Chen[1], L. Cheng[2], P. Gunaratne[3], C. Zhang[4], Q. Meng[1], Y. Xia[1], J. Badner[5], E. Gershon[6], C. Liu[1,2], The PsychENCODE Consortium.* 1) State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; 2) Psychiatry Department, University of Illinois at Chicago, Chicago, IL, USA; 3) Biology and Biochemistry Department, University of Houston, Houston, TX, USA; 4) Center for Research Informatics, The University of Chicago, Chicago, IL, USA; 5) Department of Psychiatry, Rush University Medical Center, Chicago, IL, USA; 6) Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL, USA.

   Schizophrenia and bipolar disorder are complex mental disorders, with risks contributed by multiple genes. Recently, genome-wide systemic approaches have been used to reveal the associations of hundreds of SNPs with those disorders. Disregulation of gene expression has been implied, but little is known about such regulation systems in the human brain. By integrating gene expression data of mRNA and microRNA (miRNA) obtained from the same brain collection of 51 cases and 24 controls using the weighted gene co-expression network analysis (WGCNA), we built mRNA-miRNA co-expression networks and detected modules differentially expressed between controls and patients of schizophrenia and bipolar disorder. In the disorder-associated modules, we identified the potential regulators, including transcription factors (TFs) and miRNAs, and their co-expressed protein-coding genes. In silico predicted binding relationships were used to validate the putative regulations suggested by co-expression patterns. Using SNP genotype data, we resolved some causal relationships among the regulators and their targets. We further validated the predicted regulations using RNA interference knockdown experimentally. We identified a module differentially expressed between cases and controls. This module contained five miRNAs and 501 mRNA genes. Six TFs also served as hub genes in these modules. The co-expression-suggested regulatory relationships were consistent with the binding relationships predicted by databases. Focusing on those regulations containing TFs and miRNAs, we resolved a regulation cascade from SNP variants (rs16853832) to TF (POU2F1) to miRNA (hsa-mir-320e) to target genes (NR2E1) and ultimately, to disease risks. This study showed that we can utilize multi-dimensional data to construct co-expression networks, which were enriched for regulatory relationships. Causal relationships can be resolved among SNPs, regulatory molecules and their downstream target genes through data integration. Novel genes and their corresponding regulations underlying the disease risks could be revealed.

1470F

Association of norepinephrine transporter gene (*SLC6A2*) variant with remission to venlafaxine in older adults with depression. *V. Marshe[1,2], M. Maciukiewicz[2], S. Rej[3], A.K. Arun K. Tiwari[2,4], E. Sibille[2,5], D. Blumberger[2,4], J. Karp[6,7], J.L. Kennedy[1,2,4], E.J. Lenze[8], B.H. Mulsant[1,2,4], C.F. Reynolds, III[6], D.J. Müller[1,2,4].* 1) Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 2) Campbell Family Mental Health Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 5) Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada; 6) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA; 7) Geriatric Research Education and Clinical Center, VA Pittsburgh Health System, Pittsburgh, PA, USA; 8) Healthy Mind Lab, Department of Psychiatry, Washington University, St. Louis, MO, USA.

   Background: Finding predictive biomarkers for antidepressant treatment response is imperative in the geriatric population to optimize response and minimize side effects. We investigated the putatively functional variant rs2242446/T-182C of the norepinephrine transporter (*SLC6A2*, *NET*) and 5-HTTLPR of the serotonin transporter (*SLC6A4*, *SERT*) in association with remission to the dual serotonin-norepinephrine reuptake inhibitor, venlafaxine. As an exploratory analysis, we also investigated 20 variants of other serotonergic system genes (*HTR1A*, *HTR2A*, *HTR1B*, *TPH1* and *TPH2*). **Methods:** Our sample comprised of 350 participants (≥60 years) assessing efficacy of venlafaxine treatment using a randomised, double-blind, placebo-controlled design. Participants diagnosed with major depression (MADRS≥15) were prospectively assessed for approximately 12 weeks. Associations with remission status (MADRS≤10) were conducted using multivariate binary logistic regression. We conducted mixed-models analysis for response trajectories as well as investigated time-to remission using Kaplan-Meier survival and log rank (Mantel-Cox) analyses. **Results:** The *NET* variant rs2242446, but not the *SERT* 5-HTTLPR, was significantly associated with remission after adjusting for covariates and correcting for number of tests (OR=1.68, *95% C.I.* [1.15, 2.45], p=0.007). Individuals with the C/C genotype (73.1% remitters) were more likely to remit than those with either the C/T (51.5%) or T/T genotypes (48.3%). Mixed-models analysis also revealed a significant difference in response trajectories across eight time-points (χ^2=10.47, p=0.005). Individuals with the C/C genotype reached remission faster (M=8.13±4.63 weeks) than those with the C/T (M=9.98±3.90 weeks) or T/T (M=10.25±4.34 weeks) genotypes (Mantel Cox, χ^2=7.88, p=0.019). **Conclusion:** Our findings suggest that *NET* rs2242446/T-182C mediates response to venlafaxine in late-life depression.

1471W

Association of *HTR1B* genetic polymorphisms with alcohol dependence, using the Alcohol Use Disorders Identification Test (AUDIT). *J. Kim[1], I. Choi[2], H. Shin[1,3].* 1) Research Institute for Basic Science, Sogang University, Seoul, Seoul, South Korea; 2) Department of Neuropsychiatry, Hallym University Kangnam Sacred Heart Hospital, Seoul, South Korea; 3) Department of Life Science, Sogang University, Seoul, South Korea.

   Several studies have suggested *5-hydroxytryptamine (serotonin) receptor 1B* (*HTR1B*) as a risk gene for alcohol use disorders. In addition, the serotonergic pathway has been considered as a modulator that plays an important role in numerous neuropsychiatric disorders including alcohol and drug dependence. To investigate the association between *HTR1B* genetic polymorphisms and alcohol dependence based on The Alcohol Use Disorders Identification Test (AUDIT) as a reliable and widely used screening scale, we genotyped a total of 7 common single nucleotide polymorphisms (SNPs) in 459 alcoholic patients and 455 nonalcoholic controls. Although *HTR1B* polymorphisms showed nominal associations with alcohol dependence based on DSM-IV in the case–control study, further linear regression analyses based on the AUDIT scores showed relatively significant associations between *HTR1B* polymorphisms and haplotype with alcohol dependence (minimum *P* = 0.006). Despite the needs for replications in other populations and further functional evaluations, our replication study suggests that *HTR1B* polymorphisms might be associated with alcohol dependence.

1472T

Integrating gene expression, GWAS, and eQTL results in schizophrenia.
C. Armoskus, V. Spitsyna, T. Souaiaia, C. Walker, J. Nguyen, A. Camarena, J. Herstein, K. Wang, J.A. Knowles, O.V. Evgrafov. Zilkha Neurogenetic Institute, USC, Los Angeles, CA.

   Genome-Wide Association Studies (GWAS) on schizophrenia (SCZ) have been successful in identifying over 100 regions containing variants believed to be associated with the disorder. However, given that most of the variants identified are non-coding and many loci are in intergenic regions, the mechanism connecting genomic variation to disease risk is not immediately clear. One hypothesis is that genetic differences result in differences in gene expression that then increase or decrease risk of developing SCZ. To identify such expression Quantitative Trait Loci (eQTLs) we collected gene expression data from cultured neural progenitor cells derived from olfactory neuroepithelium (CNON) from 91 patients with SCZ as well as 99 controls by RNA-Seq. Genotype data was also gathered on the patients. Comparison of CNON gene expression data to that provided by the GTEx project confirmed that CNON cultures show the closest matches to brain tissue, of the various tissues examined, and comparison to Brainspan data shows the cells are most similar to fetal brain, rather than child/adult brain tissue. Using identified eQTLs, we looked for cases where a GWAS risk variant is connected to a gene differentially expressed in SZ (DE).   SHAPEIT and Impute2 were used to impute genotyping against the 1000G reference panel. Associations between genetic variants and DESeq2-normalized gene expression were analyzed by Matrix-EQTL, which identified 4,847 cis-eQTLs (max 1 Mb between gene and variant) matching 1,224 genes, as well as 4,316 trans-eQTLs matching 1,064 genes. In total 11% of genes expressed in these samples had at least 1 eQTL identified for them. Of the 1,224 genes with cis-eQTLs, 6 had SNVs that showed significant association in the PGC SCZ GWAS, including *AS3MT*, *ITIH4*, and *SRR*, which have previous reports implicating them in SCZ pathology. 14 genes out of 121 found to be differentially expressed between SCZ and control (DE) were found to have associated eQTLs. The most significant eQTL SNV for a DE gene in the GWAS results was rs62227018 regulating *EFCAB6* which has a p-value of 0.005 for association with SCZ. Using GTEx eQTL data, we can find connections of greater significance. For example, *PRSS16* was found to be and to be connected by cis-eQTL to a nearby indel (rs201260192; raw p = 2.6e-11) which shows significant association with SCZ (raw p = 4.2e-10). In conclusion, using eQTL data allows us to bridge the gap between variants identified by GWAS and the effected genes.

1473F

Whole genome sequencing of multiply-affected schizophrenia and bipolar disorder families from the Azores and Madeira. *T.B. Bigdeli[1], S.A. Bacanu[1], K.S. Benke[2], B.S. Maher[2], J.A. Knowles[3], S. McCarroll[4], M. Pato[5], C.N. Pato[5], A.H. Fanous[6].* 1) Virginia Inst Psychiatric & Behavioral Gen, VCU, Richmond, VA; 2) Dept of Mental Health, Johns Hopkins University, Baltimore, MD; 3) Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) Department of Genetics, Harvard Medical School, Cambridge, MA; 5) Department of Psychiatry, SUNY Downstate Medical Center, Brooklyn, New York; 6) Mental Health Service Line, Washington VA Medical Center.

   Schizophrenia (SZ) and bipolar disorder (BP) are debilitating neuropsychiatric disorders, each affecting 0.5-1% of the world's population, and for which a shared polygenic liability has been demonstrated. Here, we consider recent progress in an ongoing, whole genome sequencing (WGS) study of multiply affected SZ and BP families from the Portuguese archipelagos of the Azores and Madeira. Using the Illumina HiSeq2000 platform, we obtained deep, WGS data (~30X) for 51 SCZ and 39 BP cases, and 72 unaffected relatives from 40 families; an additional 16 relatives meeting criteria for some other psychiatric diagnosis (e.g., major depression or alcoholism) and 7 of unknown phenotype; and 30 unrelated ethnically-matched controls.   Using a basic prioritization strategy for filtering putatively functional SNPs and INDELs, based on population frequency and sharing between affected relatives, we identified a disruptive INDEL in *SERPINA1* carried by affected members of 10 families, and a disruptive INDEL in *PCDHGA9* shared between all affected relatives in a pedigree which contributed substantially to a published linkage finding at 5q31-34. Individual burden scores of rare disruptive variants are markedly greater among affected subjects (*P*= 0.00076), as well as their unaffected relatives (*P*=.00034), recapitulating the pattern of results observed for common polygenic risk scores constructed from PGC2 results, for which the comparisons of affected and unaffected relatives to controls were also significant (*P*=0.01 and *P*=0.02, respectively). Next, we compared lists of genes carrying family-specific variants shared among affected relatives to curated, etiologically relevant gene-sets. Following correction for multiple tests, we observed a significant overlap of genes carrying disruptive or non-synonymous variants in the PIC with genes enriched for *de novo* non-synonymous variants (*P*=5.9E-5), associated SCZ GWAS intervals (*P*=3.48E-4), and human post-synaptic density genes (*P*=9.14E-4)**.** Genes with disruptive or non-synonymous variants also exhibited significant overlap with *de novo* findings for autism and intellectual disability, autism PPI networks, and FMRP targets. Ongoing analyses include prioritization of variants leveraged by observed population structure of these geographically isolated island populations, and attempted replication of prioritized variants an independent sample of 220 SZ cases and 185 controls from the Genomic Psychiatry Cohort (GPC).

1474W

Targeted sequencing of *FKBP5* in suicide attempters with bipolar disorder. *M. Breen[1], E. Monson[1], S. Gaynor[1], K. de Klerk[1], M. Parsons[1], T. Braun[2,3], A. DeLuca[2], P. Zandi[4], J. Potash[1], V. Willour[1].* 1) Dept. of Psychiatry, University of Iowa, Iowa City, IA; 2) Dept. of Ophthalmology and Visual Sciences, University of Iowa Carver College of Medicine, Iowa City, IA; 3) Dept. of Biomedical Engineering, University of Iowa College of Engineering, Iowa City, IA; 4) Dept. of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Background: FK506 Binding Protein 5 (FKBP5) is a critical component of the Hypothalamic-Pituitary-Adrenal (HPA) axis, a system which regulates our response to stress. FKBP5 forms part of a complex of chaperones, which inhibits binding of cortisol and glucocorticoid receptor translocation to the nucleus. Variations in both the HPA axis and *FKBP5* have been associated with suicidal behavior. **Methods:** We developed a systematic, targeted sequencing approach to investigate both coding and regulatory regions in or near *FKBP5* in 476 bipolar disorder (BP) suicide attempters and 473 BP non-attempters. This allowed us to cover more functional areas of *FKBP5* and therefore interrogate the gene more fully than our previous genome-wide association study could. Following stringent quality control checks, we performed single variant, gene burden and haplotype tests on the resulting 481 common and rare variants. Secondary analyses were performed to investigate whether sex-specific variations in *FKBP5* increased the risk of attempted suicide. **Results:** One variant, rs141713011, remained statistically significant following correction for multiple testing (P-value=7.5 x 10^{-4}, Corrected P-value=0.038, Odds Ratio=6.65). The minor allele for this variant was present in 1.7% of attempters and 0.2% of non-attempters. Three female-specific and four male-specific variants of nominal significance were also identified (P<0.05). The gene burden and haplotype association tests did not produce any significant results (P≤0.05). **Conclusions:** This comprehensive study of common and rare variants in *FKBP5* is the first to employ a novel sequencing approach that focused on both regulatory and coding regions in relation to attempted suicide. One rare variant, residing in an intron and the 3' untranslated region of multiple *FKBP5* transcripts, remained significant following permutation testing. Further investigation is required in larger sample sets to fully elucidate its association with suicidal behavior.

1475T

Genetic architecture of mental disorders: A family study based on national register data. *A. Buil[1], C. Fan[2], D. Helenius[1], A. Rosengreen[1], W. Thompson[1], Y. Wang[1,3], C.B. Pedersen[4], P.B. Mortensen[4], T. Werge[1], IP-SYCH-BROAD Researchers.* 1) Mental Health Centre Sct. Hans, Roskilde, Denmark; 2) University of San Diego, California, US; 3) NORMENT, institute of clinical medicine, University of Oslo, Norway; 4) National Centre for Register-based Research, Aarhus University, Denmark.

Recent efforts have been made to describe the genetic architecture of different mental disorders and the overlap of genetic effects affecting several disorders at the same time. However, these results are based on SNP chip data that only capture common genetic effects. Classical quantitative genetics methods applied to family related samples give estimates of heritability, genetic correlation and individual breeding values using only the family relationships among individuals. This approach informs on a larger part of the underlying heritability than SNP based studies. In this work we use classical quantitative genetic methods to describe the genetic architecture of mental disorders in the Danish population and compare these results with the SNP-based counterparts commonly used nowadays. The iPSYCH project consists in 80000 Danish individuals born between 1981 and 2005. 30000 individuals were sampled randomly from the population (~2% of subjects born in the period) while the remaining 50,000 subjects have a diagnosis of schizophrenia, bipolar disorder, attention deficit disorder, major depression or autism (100% of diseased subjects born in the period). All 80,000 subjects have been genotyped on the PsychChip. We retrieved information from the national civil and health registers on all first degree relatives, including familiar relationships and diagnosis, expanding the sample to 500000 individuals. Variance components (VC) models allow the estimation of disease heritability of diseases as well as genetic correlation between diseases. Cox linear mixed models allow the estimation of heritability taking into account the age of onset of the disease and produces breeding values (genetic load related to the disease) for each individual. Here we present genealogy-based estimates of heritability for the five major mental disorders and compare them with current SNP-based estimates of heritability in the same population. We also calculate genealogy-based pairwise estimates of genetic correlations and compare them to their SNP-based counterparts. Finally, we will estimate individual breeding values for each disease and we will compare them with the polygenic risk scores obtained from SNP data and the burden of putative high-risk variants identified through SNP and sequence based analyses. In summary, we show the utility of classical quantitative genetics methods applied to national based register data to describe the genetic architecture of mental disorders.

1476F

Multiple rare mutations of postsynaptic protein genes associated with schizophrenia. *M. Cheng, S. Hsu, H. Tsai, S. Tsai.* Department of Psychiatry, Yuli Branch, Taipei Veterans General Hospital, Hualien County, Taiwan.

Schizophrenia is a chronic debilitating mental disorder with a high genetic component in its etiology. Mounting studies have shown that schizophrenia is a neurobiological disorder with aberrant synaptic connectivity and synaptogenesis. Gene involved in the formation and functional integrity of synapse can be considered as the candidate gene of schizophrenia. We resequenced the exon regions of 18 genes encoding the synaptic proteins that belong to DLGAP, NRXN, NLGN, SHANK, and HOMER family in 50 schizophrenia patients using targeted-sequencing method for searching potential genetic variants related to schizophrenia. We identified 23 missense mutations from schizophrenic patients. Bioinformatic analysis showed some of these mutations were damaging or pathological to the protein function. Of note, we detected one putative damaging missense mutation, p.R309Q, of the *NLGN2* gene in one patient but not in 549 controls. This patient-specific mutant significantly reduced the reporter gene activity compared with NLGN2 wild-type, which is in line with our previous findings (Hum Mol Genet 2011 Aug 1;20(15):3042-51), supporting that rare missense mutations of the *NLGN2* gene are associated with schizophrenia. In addition, we found four missense mutations of SHANK genes in schizophrenia patients, including SHANK1:p.2074C, SHANK2:p.R443H, SHANK3:p.P209A, and SHANK3:p.T1236I. These missense mutations were not identified in 526 controls. Taken together, multiple rare missense variants of the gene encoding postsynaptic protein contribute to the pathogenesis of schizophrenia in some patients, but gene functional assays are needed to verify their relevance to the pathogenesis of schizophrenia.

1477W

A pilot study for association of tryptophan metabolome and major depressive disorder in Taiwan. *L. Chuang[1], Y. Chung[2], P. Kuo[2].* 1) Department of Nursing, Cardinal Tien Junior College of Healthcare & Management, I-LAN, Taiwan; 2) Department of Public Health & Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan.

Objectives: Major depressive disorder (MDD) is one the most common psychiatric disease. Serious emotional disturbance, insomnia and suicide attempt of these patients then cause serious problems at work and social life. Many studies have indicated the significant association between MDD episode and interaction of life stress and serotonin transportation. Serotonin is metabolized from tryptophan. Tryptophan has two major metabolic pathways, serotonin and kynurenine pathway. The regulation of restriction enzyme for tryptophan metabolism may influence the balance between amount of serotonin and kynurenine in the blood. Recent studies indicated cytokine interferon-γ of inflammation can regulate the restriction enzyme of tryptophan metabolism. The present study explored the possible interaction between inflammation cytokine, tryptophan metabolome, and Taiwan MDD. **Methods:** We used one-to-one conditional case-control study design. We recruited 38 patients of MDD and 38 subjects for healthy control group, and gender and age gap that within 5 years were matched between case and control subjects. Enzyme-linked immunosorbent assay was used to detect the amount of serotonin in the serum. The amount of tryptophan metabolome was detected by liquid chromatography-mass spectrometry. **Results:** After adjusted the frozen period of serum sample, subjects cannot detect serotonin in the peripheral blood were associated with an increased risk of the MDD showing odds ratios (ORs) with 95% confidence intervals (CI) of 8.9 (2.3-46.8) (p-value=0.0036). Compared with participant who can detect serotonin and low kynurenine expression, our data shown non-detectable serotonin and high kynurenine expression plus non-detectable serotonin had increased risks of the MDD showing multivariate ORs (95% CIs) of 18.6 (2.74-384.3)(p-value=0.0011) and 21.5 (2.4-511.1) (p-value=0.0159), respectively. Experimental validation is underway for the individual genotype of rs2430561 in gene interferon-γ to evaluate the possible hypothesis of regulation for restriction enzyme of tryptophan metabolism. **Conclusion:** The expression of tryptophan metabolome has significant association for MDD in Taiwanese population. .

1478T

Contribution of rare copy number variants to the development of autism spectrum disorder in high-risk siblings. *L. D'Abate[1,2], S. Walker[1], R.K.C. Yuen[1], K. Tammimies[1,3], M.J. Gazzellone[1,2], B. Thiruvahindrapuram[1], J. Howe[1], J. Brian[4], S. Bryson[5], K. Dobkins[6], R. Landa[7], D. Messinger[8], S. Ozonoff[9], I. Smith[5], W. Stone[10], S. Warren[11], G. Young[9], L. Zwaigenbaum[12], S.W. Scherer[1,2], Infant Siblings Consortium.* 1) Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; 4) Autism Research Centre, Bloorview Research Institute, Toronto, Ontario, Canada; 5) IWK Health Centre, Dalhousie University, Halifax, Nova Scotia, Canada; 6) Department of Psychology, UC San Diego, La Jolla, California, USA; 7) Hugo Moser Research Institute, Kennedy Krieger Institute, Baltimore, Maryland, USA; 8) Department of Psychology, University of Miami, Coral Gables, Florida, USA; 9) MIND Institute, UC Davis, Davis, California, USA; 10) Autism Centre, University of Washington, Seattle, Washington, USA; 11) Vanderbilt Kennedy Center Treatment and Research Institute for Autism Spectrum Disorders, Vanderbilt Kennedy Centre, Nashville, Tennessee, USA; 12) Autism Research Centre, University of Alberta, Edmonton, Alberta, Canada.

   Background: Studies conducted on trios and quartet families have implicated rare *de novo* and inherited copy number variants (CNVs) in the complex etiology of Autism Spectrum Disorder (ASD), however no study to date has investigated whether genetic testing in young children can inform of ASD status prior to the typical age of onset of symptoms. Here, we correlated our genetic findings with longitudinal clinical data available on infant siblings registered with the Infant Siblings Consortium, allowing us to assess the predictive value of early genetic testing. **Methods:** A total of 441 individuals (190 probands, 84 siblings with ASD, 167 unaffected siblings) from 199 families (126 SPX, 73 MPX) were genotyped on the high-resolution Affymetrix CytoScan HD platform, with high-confidence rare CNVs (<0.1% of control population) being ascertained. All infant siblings were clinically assessed periodically and given an ASD diagnosis, if appropriate, at 3 yrs. of age. **Results:** Of 199 families analyzed, we observed *de novo* CNV rates of 5.3% in affected individuals and 3.5% in unaffected siblings. Two of the non-ASD siblings harbouring *de novo* CNVs displayed a broader autism phenotype (BAP). Presumed pathogenic rare CNVs were found in 15/199 families (7.5%). In 7 of these families, the index case and one or more siblings carried the presumed pathogenic CNV; in 6 of these cases the carrier individual either had ASD (4) or had a subclinical form of ASD (2). One such example was a maternally-inherited 90 Kb deletion of the third exon of *PTCHD1-AS* in a male proband and male sibling. Another notable genetic finding was a 200 KB deletion spanning the 5'-end of *NRXN1* in a male proband and non-ASD female sibling. Close examination of the latter sibling's phenotype revealed global developmental delay and BAP. There were also 4 additional families in which CNV data would have been informative for early diagnosis. **Conclusions:** Our results suggest CNVs may serve as early biological markers for ASD, information particularly relevant when there is a family history.

1479F

Co-localization of eQTL and GWAS signatures in schizophrenia. *A. Dobbyn[1,4], L.M. Huckins[1], S.K. Sieberts[6], B. Devlin[5], P. Roussos[1,2,3], C. Giambartolomei[1], P. Sklar[1,2,3,4], E.A. Stahl[1,2], CommonMind Consortium.* 1) Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029; 5) Department of Psychiatry, University of Pittsburgh, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213; 6) Systems Biology, Sage Bionetworks, 1100 Fairview Ave N, Seattle, WA, 98109, USA.

   INTRODUCTION: GWA studies have identified numerous Schizophrenia (SCZ) associated loci. However, the utility of GWAS for SCZ and other complex, polygenic disorders has been limited by a lack of mechanistic hypotheses linking associated variants to risk. Expression quantitative trait loci (eQTLs), variants that are associated with changes in gene expression, represent the best known example of a putative functional role for genetic associations. Using summary statistics from the latest SCZ GWAS and eQTLs from DLPFC, we characterized the overlap between GWAS loci and eQTL association signals, for primary (most significant) and secondary (significant after conditioning on the primary) eQTLs. Co-localization of secondary but not primary eQTLs with GWAS signal is consistent with regulation of expression underlying genetic risk in cell types, developmental time points, or stimulation conditions that are only partly represented by available eQTL data. METHODS: We ran MatrixEQTL on genotype and RNA-seq data from CommonMind Consortium post-mortem DLPFC samples (467 EA cases and controls) to identify primary and secondary (conditional) cis-eQTLs (FDR <=0.05). We tabulated the number of primary and secondary eQTLs falling within 108 PGC SCZ GWAS (36,989/113,075 cases/controls) loci defined as r^2 >=0.6 LD-range, resulting in 158 genes with overlapping eQTLS. For each of these genes, we ran a set of conditional analyses to determine the association signal of each conditional cis-eQTL in isolation. We assessed concordance between SCZ GWAS and eQTL signals by (1) COLOC (PMID: 24830394), and (2) inspection of SCZ GWAS versus eQTL p-values. RESULTS: The stepwise forward selection analysis identified ~29,000 primary and/or secondary eQTLs in ~13,000 genes, with a mean of 2.24 eQTLs per gene. COLOC results show 42/158 genes to have a maximum posterior probability for one common causal GWAS association/eQTL: 37/42 of these genes correspond to the primary eQTL and 5/42 (including DOC2A, IREB2, and NEK1) to secondary. Manual inspection of plots supports these findings. CONCLUSIONS: There is evidence for overlap between GWAS association and eQTL signature in 42/158 SCZ-associated genes. The relationship between eQTL and GWAS signature in the remaining 116/158 genes is unclear; for these loci, eQTLs do not appear to explain the regional GWAS signal.

1480W

Investigating the genetic relationship between Alzheimer's disease and cancer using GWAS summary statistics. *Y. Feng[1], J. Cormick[3], K. Cho[3,5], J. Driver[3,4,5], L. Liang[1,2], IGAP consortium, GAME-ON consortium.* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Massachusetts Veterans Epidemiology Resource and Information Center, Boston VA Medical Center; 4) Geriatric Research Education and Clinical Center and Boston VA Medical Center, Boston, MA; 5) Division of Aging, Brigham and Women's Hospital, Boston, MA.

Epidemiological cohort studies and transcriptomic meta-analysis suggest an unusual inverse association between Alzheimer's disease (AD) and cancers. We adopted a genome-wide approach, examining the genetic correlation between AD and a variety of cancer types using GWAS summary statistics from studies with a sample size of up to 20,000-30,000 individuals and >5 million SNPs imputed from the 1000 Genomes European reference panel. Our main results, estimated by cross-trait LD Score regression, showed an overall positive genetic correlation between AD and any of the five cancers combined (colon, breast, prostate, ovarian, and lung; r_g = 0.17, P = 0.04); specifically, we found a significant positive genome-wide relationship of AD with breast cancer (ER-negative and overall; r_g = 0.21 and 0.18, minP = 0.03) and with lung cancer (adenocarcinoma, squamous cell carcinoma and overall; r_g = 0.31, 0.38 and 0.30, minP = 0.01) while a negative genetic correlation of AD with prostate cancer (aggressive and overall; r_g = -0.07 and -0.09, minP = 0.20). This indicated that the two diseases share some genetic variants involved in biological processes that modulate disease risk in the same or opposite direction. Potential cross-phenotype associations were detected between AD and breast, prostate, or lung cancer at specific SNPs, also pointing to common genetic pathways shared by the two traits. Partitioning genetic covariance between AD and each cancer type into functional categories via stratified LD Score regression showed a mixture of positive and negative signals, with stronger signals appearing at enhancers, histone marks (H3K4me1 and H3K4me3), and DNase hypersensitivity sites (DHSs), suggesting that these functional elements may contribute more to the co-heritability between the two diseases than others across the genome. Gene expression data in the most relevant tissues is needed for comprehensive investigation of patterns of shared disease mechanisms and connecting the association observed at the genetic level to that at the phenotypic level. Our study is the first to our knowledge that examines the genetic overlap between AD and specific cancer types using large-scale GWAS meta-analysis results where no individual genotype data is required. We anticipate that our findings will be incorporated into future functional studies that can provide insight into the development of novel therapies for both AD and cancer.

1481T

Transcriptome sequencing in post-mortem brain samples implicates a role for the complement system in schizophrenia. *L. Feuk[1,2], E. Lindholm Carlström[1,2], J. Halvardson[1,2], M. Etemadikhah[1], J. Brännström[1], G. Rajkowska[3], C.A. Stockmeier[3], B. Nilsson[1], P.U. Magnusson[1].* 1) Dept. of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala, Sweden; 3) University of Mississippi Medical Center, Jackson, MS, USA.

Transcriptome sequencing is a robust and powerful method to estimate gene expression changes. Here we present findings from transcriptome sequencing in post mortem brain tissue in a cohort of 65 schizophrenia patients and 43 age and sex matched control samples. This represents a sample size twice as large as previously published brain tissue analyses in schizophrenia. Fresh frozen tissue samples were obtained from frontal cortex (Brodmann area 8 and 9) and inferior frontal gyrus. Each sample was cryo-sectioned twice, and then sequenced in duplicate using independent libraries. Expression analysis yielded 42 significant differentially expressed genes between the patient and control groups (adjusted p-value < 0.05). The set of differentially expressed genes showed a general enrichment of immune system genes, with a striking up-regulation of several genes in the complement system in patients. Clustering based on differentially expressed genes yielded three clear clusters, with one cluster containing >80% patients. This group of samples all showed strong complement system activation. Several previous studies point to an important role for the classical complement cascade in neuron differentiation and synaptic pruning. While we do find significant differential expression for the recently published C4 gene (adj. p-value=0.003), we find equal or more highly significant expression differences for several additional complement genes, including C1R, CFB and C7. The overexpression of complement factors in patients was validated with qrtPCR. Using immunohistochemistry we also show that complement factor differences can be detected at the level of the protein in brain sections. Taken together our findings support an important role for C4, but in addition implicate activation of many other genes in the complement cascade as a crucial pathway in schizophrenia pathology.

1482F

Targeted sequencing and functional assessment of the 2p25 region in attempted suicide. *S.C. Gaynor[1], M.E. Breen[1], E.T. Monson[1], K. de Klerk[1], M. Parsons[1], A.P. DeLuca[2], T.E. Scheetz[2,3], P.P. Zandi[4], J.B. Potash[1], V.L. Willour[1].* 1) Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA; 2) Department of Ophthalmology and Visual Sciences, University of Iowa Carver College of Medicine, Iowa City, IA; 3) Department of Biomedical Engineering, University of Iowa College of Engineering, Iowa City, IA; 4) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Suicidal behavior is a complex phenotype with an estimated heritability of 30-50%. This heritability partly depends on the presence of a comorbid psychiatric disorder. However, there is thought to be an independent heritable factor that is specific to suicide. We previously conducted a genome-wide association study (GWAS) of the attempted suicide phenotype with the goal of identifying this independent factor (Willour et al., 2012). This GWAS implicated common variation in 2p25, a 350kb region encompassing four genes (*FAM110C, SH3YL1, ACP1, FAM150B*). In the current study, we have performed a targeted next-generation sequencing assessment of the entire 2p25 region in an effort to identify both common and rare variation that may contribute to the risk for suicidal behavior. We assessed 476 bipolar subjects with a history of attempted suicide (224 males and 252 females) and 473 bipolar subjects with no history of an attempt (222 males and 251 females). Our top gene-level finding from this study was *FAM150B* (p = 0.022), but this result did not remain significant following correction for multiple testing. Our top individual-variant result was rs300799, an intergenic variant between *FAM110C* and *SH3YL1*. This variant was significantly associated with the attempted suicide phenotype in males, with the minor allele present in 22.3% of attempters and 12.3% of non-attempters (p = 4.8×10^{-5}, corrected p = 0.035, odds ratio = 2.13). Nearly all of our top variants localized to an 80kb linkage disequilibrium block within 2p25. As part of the current study, we are also performing a functional assessment of this 80kb region using the CRISPR-Cas genome-editing system to look for the presence of regulatory elements affecting gene expression or cellular morphology. This study provides further support for a potential role of the 2p25 region in the attempted suicide phenotype.

1483W

Polygenic risk score reveals a genetic relationship between SCZ and OCD. *W. Guo[1], C. Mathews[2], E.M. Derks[3], D.Y. Yu[4], D. Schijven[5], J.J. Luykx[5], J. Scharf[6,7,8], S.E. Stewart[9], J. Samuels[10], J.A. Knowles[11], G. Nestadt[10], Y.Y. Shugart[1], PGC-OCD working group.* 1) Unit of Statistical Genetics, NIMH, Bethesda, MD; 2) College of Medicine, University of Florida, Gainesville, Florida, United States of America; 3) Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 4) Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Department of Psychiatry, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, United States of America; 5) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America; 7) Division of Cognitive and Behavioral Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America; 8) Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, United States of America; 9) British Columbia Mental Health and Addictions Research Institute, University of British Columbia, Vancouver, British Columbia, Canada; 10) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America; 11) Department of Psychiatry and Behavioral Sciences Keck School of Medicine at the University of Southern California, Los Angeles, California, United States of America.

It is known that obsessive-compulsive disorder (OCD) and schizophrenia (SCZ) are both highly heritable neurodevelopmental disorders and seem to have notable levels of overlap on the psychopathological level. OCD is fairly common in SCZ, having approximately 13.6% prevalence. In this study, we sought to explore the genetic relationship within and across the two disorders by conducting a standard polygenic score approach. Polygenic score is defined as the sum of multiple single-nucleotide polymorphism alleles associated with the trait for an individual, and the sum is typically weighted by effect sizes estimated from a genome-wide association study. In this study, we use the summary statistics obtained from the PGC-GWAS analysis to predict OCD susceptibility. The first OCD (OCGAS) dataset includes 2,064 individuals in nuclear families. The second OCD (IOEU) dataset includes 3,859 unrelated individuals. The SCZ dataset include 36,989 cases and 113,075 controls. Together, the analyses conducted identified a statistically significant polygenic component of SCZ, predicting approximately 1% of the phenotypic variance in two independent OCD datasets. Our results suggested the presence of polygenic sharing between SCZ and OCD. Yet would like to note that we made the assumption that all genetic markers have effects that are included in the polygenic scores. The merits and limitations of using polygenic score analyses for the predication of psychiatric diseases need further evaluations. .

1484T

A comprehensive survey of protein truncating variants in schizophrenia. *H. Huang[1,2], M. Rivas[1,2], G. Genovese[2], T. Singh[3], D. Howrigan[1,2], F.K. Satterstrom[1,2], S.A. McCarroll[2,4], B.M. Neale[1,2], M.J. Daly[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Sanger Institute, Hinton, UK; 4) Department of Genetics, Harvard Medical School, Boston, MA, USA.

Schizophrenia is a severe and disabling psychiatric disorder affecting one percent of the general population, with the overall cost of care in the U.S. exceeding $60 billion per year. There have been limited breakthroughs in the past 60 years in developing new treatments for schizophrenia, partially due to limited understanding of how the disease develops. It has been recognized that schizophrenia has a strong hereditary component, with a broad-sense heritability of up to 81%, indicating that genetics can be extremely helpful in providing new insights and advancing our knowledge of the mechanism of disease. Studies on common variants associated with schizophrenia have been successful with more than 100 susceptible loci identified to date. It is also important to find disease-associated rare variants, especially those that are functionally critical, as they generally have a large effect and are more easily translated into models of the disease mechanism. In addition, independent findings in common and rare variants have been shown to converge at a broad functional level, including genes encoding calcium channels or involved in glutamatergic neurotransmission and synaptic plasticity. Here we report a comprehensive survey of protein truncating variants (PTVs), a class of highly deleterious variants, in schizophrenia by combining large-scale exome-sequencing datasets. We aggregated 9,672 cases and 19,017 controls from various sources including the UK10K, Swedish, Finland, Ashkenazi Jewish and Taiwan datasets. After QC, we kept 64,938 PTVs that were not observed in the Exome Aggregation Consortium 'nopsych' subset (release 0.3, 45,376 samples excluding psychiatric cohorts). For each gene, we counted the numbers of PTVs in cases and controls, respectively, and tested for association using Fisher's exact test. Initial analysis identified several genes with suggestive associations with schizophrenia. As additional cohorts are added to this study, we expect the power will be improved such that some genes may reach genome-wide significance. All results, including the number of PTVs in cases and controls per cohort, are presented in an interactive browser for search and visualization. In addition to PTVs, we included all exonic variants sequenced in this study and their meta-analysis odds ratio and *P*-value.

1485F

Variants in mitochondrial metabolic enzymes are associated with major depressive disorder in the Japanese population. *D.D. Ikeda[1], K. Okuda[2], S. Nagasaka[3], T. Ono[1], M. Masuda[3], T. Fujiwara[3], H. Haretsugu[1], K. Hattori[4], H. Kunugi[4].* 1) Biomedical Technology Research Center, Otsuka pharmaceutical Co., Ltd., Tokushima, Japan; 2) Department of Clinical Research and Development, Otsuka Pharmaceutical Co.,Ltd., Osaka, Japan; 3) Clinical Pharmacology, Headquarters of Clinical Development, Otsuka pharmaceutical Co., Ltd., Osaka, Japan; 4) Department of Mental Disorder Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry (NCNP), Tokyo, Japan.

Major depressive disorder (MDD) is a complex and chronic disease with a partly genetic etiology. Although a number of genome-wide association studies (GWASs) mainly in European large cohorts have been conducted so far, there were few replicated variants which achieved the genome-wide significance. Here we report a MDD-GWAS in the Japanese small population, which includes 103 cases and 3,229 healthy controls consisting partly of the JPDSC (the Japan PGx Data Science Consortium) database. After whole-genome imputation utilizing the 1000 genomes reference data, we gained over seven million high-quality (Rsq > 0.7) SNPs. We found that multiple candidate SNPs were identified in the two gene regions, *ME3* and *DHTKD1* ($p < 10^{-6}$). Several SNPs of them in the *ME3* gene region exceeded the genome-wide significance ($p < 5 \times 10^{-8}$). Notably, both the two genes encode metabolic enzymes localized in mitochondria. Furthermore, cis-eQTL analyses based on GTEx database (V6) revealed that four strongly disease-associated SNPs in the *ME3* gene region were also associated with the corresponding gene expressions in brain caudate ($p < 0.05$). These findings imply that metabolic imbalances in caudatal mitochondria could contribute to the disease state of MDD under specific genetic backgrounds or circumstances.

1486W

Mutations in *CADPS* identified in patients with bipolar disorder affect protein function and animal behavior. *S. Jamain[1,2,3], J. Sitbon[1,2,3], C. Kappeler[1,2,3], A. Nicolas[1,2,3], D. Nestvogel[4], A. Henrion[1,2,3], B. Etain[1,2,3,5], M. Leboyer[1,2,3,5], J.S. Rhee[4].* 1) Psychiatrie Translationnelle, INSERM U955, Créteil, France; 2) Université Paris Est, Faculté de Médecine, Créteil, France; 3) Fondation FondaMental, Créteil, France; 4) Department of Molecular Neurobiology, Max Planck Institute of Experimental Medicine, Göttingen, Germany; 5) AP-HP, Hôpital H. Mondor – A. Chenevier, Pôle de Psychiatrie, Créteil, France.

   With a prevalence of 1% in general population, bipolar disorder is one of the most severe and common psychiatric disorders. Many studies suggest a preponderant role for genetic factors in bipolar disorder, mainly in early-onset form of the disease. However molecular mechanisms underlying this disorder remain unclear. We identified missense variations and one deletion in a gene encoding the calcium-dependent activator protein for secretion (*CADPS*) in patients with early-onset bipolar disorder. CADPS is an essential regulator of synaptic and large dense core vesicles exocytosis in mammalian neurons and neuroendocrine cells, respectively. Moreover, CADPS promotes vesicular catecholamine uptake and storage mediated by vesicular monoamine transporters. In this study, we showed a higher mutation frequency in patients with early-onset bipolar disorder than in unaffected controls. In cellular models, the CADPS functions were affected by the mutations identified in patients. Some mutations resulted in a decreased expression level of the protein. Others impaired the ability of CADPS to promote the vesicular monoamine uptake *in vitro*. At last, some mutations affected the ability of CADPS to increase the calcium-dependent neurotransmitter release in PC12 cells. Behavioral studies of mutant mice for *Cadps* exhibited modification in anxiety-related behaviors as well as in forced swim test and tail suspension test, suggesting an altered sensitivity for these animals to mild acute stress, as compared to wild-type littermates. Altogether, our results suggest that impairment in CADPS functions may affect the behavior of mutant subjects. Such mutations may thus increase the vulnerability to early-onset bipolar disorders in humans.

1487T

Integrative analysis of *de novo* mutations and common variants in schizophrenia. *P. Jia[1], Z. Zhao[1,2].* 1) School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.

   Recent studies of schizophrenia (SCZ) have increasingly supported that genetic components susceptible to SCZ involve a wide spectrum of risk factors, including common variants (CVs), rare variants, and de novo mutations (DNMs) with effect sizes ranging from small to large. However, few overlapping genes were found both impacted by CVs associated with SCZ and by extremely rare DNMs occurred in only SCZ probands. Considering that it is unlikely these two types of variants work independently in unrelated biological processes and lead to the same disease phenotype, we hypothesized that they could congregate on common biological pathways and processes to cause SCZ. To this end, we referred to these variants as network-attacking variants and set to distinguish their interaction perturbations in networks by studying their residing genes. We mapped genes with DNMs (DNMgenes) and genes with CVs (CVgenes) onto the human protein-protein interaction (PPI) network as well as its sub-networks derived by spatial and temporal expression profiles in brain. Our results showed DNMgenes had a pattern of node attacking, where DNMs lead to node removal from the network and destroyed all the interactions of the DNMgenes. In contrast, CVgenes appeared to have a pattern of edge rewire, where some, but not all, of their interactions were impacted, likely due to dysregulated expression by CVs with regulatory roles. Both patterns were replicated in spatiotemporal sub-networks of brain development, especially in the frontal cortex and sub-cortical regions in fetal developmental stage. In addition, DNMgenes and CVgenes were found to be more accessible to each other than to control genes, indicating they were likely involved in common biological processes. We then developed a network-assisted method to link DNMgenes and CVgenes in frontal cortex during the fetal brain developmental stage and built a SCZ-specific module that was enriched with both categories of genes. We found the resultant SCZ-specific module featured with major groups of genes functional in immune, chromosome modification, and neuronal pathways, which were in line with previous studies on the largest GWAS of SCZ. Strikingly, both DNMgenes and CVgenes contributed to these functional pathways, suggesting a core set of common pathways and networks underlying SCZ. In summary, we conducted a systematic investigation of common variants and de novo mutations and revealed their linking roles underlying SCZ.

1488F

No evidence that the most studied candidate genes for schizophrenia are more relevant to schizophrenia than random sets of genes. *E.C. Johnson[1,2], W.E. Melroy-Greif[3], R. Border[1,2], C. de Leeuw[4], M.A. Ehringer[2,5], M.C. Keller[1,2].* 1) Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO; 2) Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO; 3) Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA; 4) Department of Neurosciences-CNCR, VU University Amsterdam, Amsterdam, Netherlands; 5) Department of Integrative Physiology, University of Colorado at Boulder, Boulder, CO.

A recent analysis of 25 historical candidate gene polymorphisms for schizophrenia found no clear evidence that these polymorphisms were associated at levels above chance in the largest GWAS study conducted to date. However, this study focused on specific polymorphisms, leaving open the question of whether the classic schizophrenia candidate genes themselves, and not just the most studied polymorphisms, are more enriched for lower GWAS p-values than random genes. To answer this question, we used a gene set analysis to examine whether there is evidence that these 25 previously studied candidate genes have stronger associations with schizophrenia than would be expected by chance. Based on our findings, there is no evidence that these classic candidate genes are any more associated with schizophrenia than sets of randomly chosen genes. Candidate gene research continues to be conducted on other phenotypes, but the candidate gene literature from schizophrenia should serve as a cautionary tale. Our results suggest that hypothesis-driven candidate genes need to be validated by GWAS data before further scientific effort is devoted to them.

1489W

Sex-specific characterization of the genetic architecture of obsessive-compulsive disorder. *E.A. Khramtsova[1,2], R. Heldman[3], D. Yu[4], E.M. Derks[5], B.E. Stranger[1,2], L.K. Davis[6] on behalf of the OCD PGC Workgroup.* 1) Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL; 3) University of California, Berkeley, California; 4) Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Department of Psychiatry, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, United States of America; 5) Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 6) Vanderbilt Genetics Institute, Department of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN.

Although sexually dimorphic human phenotypes and disease characteristics are widespread, the etiology of those differences remains poorly understood. Such dichotomy is exemplified by higher prevalence of many autoimmune diseases in females and many of the neurodevelopmental phenotypes in males. These differences are often attributed to varying sex-hormones or to the sex chromosomes. Recent genetic studies have implicated a contribution to sex-biased phenotypes from autosomal genetic variation, subsequently motivating the characterization of the sex-specific genetic architecture of complex traits. Obsessive-compulsive disorder (OCD) demonstrates a sex-biased age of onset, with males displaying earlier age of onset peaking in adolescence and females displaying a later age of onset peaking in adulthood, suggesting a possible difference in genetic architecture between the sexes. We sought to characterize the sex-specific genetic architecture of OCD by first performing a sex-stratified genome-wide association meta-analysis on over 10,000 samples available through the OCD-PGC (N_{female} = 6030; N_{male} = 4130). We did not identify any genome-wide significant genetic associations for either sex. Next, we performed a sex-stratified estimation of narrow-sense heritability (h^2) using LD score regression method, and found that the heritability of OCD in males (h^2 = 0.15, SE = 0.08) is lower than in females (h^2 = 0.24, SE = 0.08). Based on the genetic correlation between the sexes (rg = 1.0, SE = 0.49, p=0.03), we find that this difference does not reflect variable genetic mechanisms, but may suggest a difference in liability thresholds between the sexes. Additionally, we report results from ongoing analyses including the contribution of sex-biased expression quantitative trait loci to OCD, global analysis of heterogeneity (I^2) between the sexes, and an assessment of the polygenic contribution of the X chromosome. Finally, to further characterize the genetic correlation between OCD and common sex-biased comorbidities (such as Tourette syndrome and attention deficit and hyperactivity disorder), we provide results of a polygenic risk score analysis performed on our sex-stratified sample. In addition, we provide an analytic pipeline and share pitfalls and best practices for sex-specific genetic analyses. Our analysis provides insight into the sex-specific polygenic architecture of OCD.

1490T

Whole genome sequencing analysis of a large pedigree with psychiatric disorders. *J. Lihm[1], P. Thomson[2], S. McCarthy[1], K.L. Evans[2], N.M. Ryan[2], M. Kramer[1], E. Ghiban[1], J.K. Millar[2], A. McIntosh[2], D. Blackwood[2], D. Porteous[2], W.R. McCombie[1].* 1) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, New York, USA; 2) University of Edinburgh, Edinburgh, UK.

The study of complex diseases in an extended pedigree design has several advantages. Family based design avoids population stratification and reduces genetic heterogeneity, narrowing the search for rare and novel variants that segregate with a phenotype. However, there are a number of challenges in collecting and analyzing the data, particularly when the pedigree is large and multigenerational. Here we will present a comprehensive analysis pipeline that addresses measures of data quality control for pedigrees, and detects and prioritizes various types of variants including single nucleotide variants, small InDels, copy number variations, and structural variations. We constructed the pipeline with published and in-house software tools such as GATK, Scalpel, LUMPY, and ANNOVAR. The analysis pipeline was applied to a multigenerational pedigree where major mental illness segregates with a balanced translocation t(1;11), between chromosomes 1 and 11. Forty-nine samples were available for whole genome sequencing (20 carriers, 29 non-carriers of the translocation). We implemented quality measures from sequencing and variant calling, gender confirmation, and measures to confirm pairwise relations as marked in the pedigree. Next, for a fast and straightforward scan of SNVs and CNVs, we prioritized possible risk variants by odds ratios and allele frequency. Genome-wide analysis identified rare and novel SNVs in chromosome 1 with high odds ratios that segregate with the translocation located within 8.4Mb of the translocation breakpoint in chromosome 1. Around the t(1;11) breakpoint on chromosome 11, we identified relatively common (MAF≤0.08) intronic and intergenic SNVs within 3Mb of the breakpoint and relatively rare (MAF 0.02) single copy deletion in a TENM4 intron within 11Mb of the breakpoint, all of which segregate with the translocation. We would like to extend this approach to small InDels and structural vairations, incorporating annotations for functional class including the study of the transmission pattern of common and rare variants with different penetrance values. We will use the pipeline to study the effect of shared variants in the affected members and their interaction with the DISC1 pathway.

1491F

Genetic variation and DNA methylation of *SKA2* in obsessive-compulsive disorder. *A.J. Lisoway[1,2], C.C. Zai[1,3], G. Zai[1,2,3,4], A. Nair[2,5], S. Ebrahimi[2,5], A.K. Tiwari[1,3], V. Labrie[2,3,5], A. Petronis[2,3,5], Z.A. Kaminsky[6], J.L. Kennedy[1,2,3], M.A. Richter[1,2,3,4].* 1) Neurogenetics Section, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON; 2) Institute of Medical Science, University of Toronto, Toronto, ON; 3) Department of Psychiatry, University of Toronto, Toronto, ON; 4) Frederick W. Thompson Anxiety Disorders Centre, Sunnybrook Health Sciences Centre, Toronto, ON; 5) Krembil Family Epigenetics Laboratory, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON; 6) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD.

Background: Obsessive-Compulsive Disorder (OCD) has a strong genetic component, is highly comorbid with Major Depressive Disorder (MDD), and is often triggered by environmental stress. Epigenetic modifications are thought to play a significant role in complex psychiatric disorders and may impact the relationship between genes and the environment via the hypothalamic-pituitary-adrenal (HPA) axis. The *SKA2* gene interacts with the glucocorticoid receptor, is implicated in mediating HPA axis function, but has yet to be examined in OCD. We hypothesized that genetic and epigenetic variation in *SKA2* may play a role in OCD disease risk and symptom severity. **Methods:** A sample of n=64 OCD patients, divided equally by presence/absence of comorbid MDD, and age and gender matched, were selected. Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) score was used to measure disease severity. Genotyping for the *SKA2* rs7208505 marker was performed using the Taqman allele specific assay method and DNA methylation levels were quantified using bisulfite pyrosequencing. Methylation percentage was corrected for genotype using previously established methods. Genotypes for rs7208505 were compared to controls of European ancestry (n=365; 1000 Genomes Project Consortium) using Pearson's chi-squared test. The relationship among the rs7208505 variant, methylation corrected, symptom severity, and MDD status was modeled using linear regression. **Results:** The *SKA2* rs7208505 genotype distributions significantly differed between OCD patients and controls (χ^2=8.66, p=0.013). Genotypes were associated with symptom severity (p=0.011), in that patients with the CC genotype had lower Y-BOCS scores than the CT/TT genotypes. While OCD patients diagnosed with MDD had more severe Y-BOCS scores (p=0.023), neither genetic nor epigenetic variation of rs7208505 was associated with comorbid MDD status. Methylation corrected for genotype was not significant in the regression model (p>0.05). **Conclusions:** These results provide some evidence that *SKA2* genetic variation may be associated with OCD risk and symptom severity. Methylation corrected for genotype was not associated with OCD symptom severity; however, patients with TT and CT genotypes had more severe Y-BOCS scores than the CC genotype carriers. Further work will include replication in larger samples and interrogating additional genetic and epigenetic variations in *SKA2*.

1492W

Objective ranking of neurocognitive endophenotypes for schizophrenia in the Western Australian Family Study of Schizophrenia (WAFSS) using measures of heritability and genetic correlation. *N.S. McCarthy[1], M.L. Clark[2], E.E.M Knowles[3], G. Cadby[1], D. Glahn[3], J. Blangero[4], M. Dragović[2], P.E. Melton[1], E.K. Moses[1], J.C. Badcock[2], A. Jablensky[2].* 1) Centre for the Genetic Origins of Health and Disease, Faculty of Health Sciences, Curtin University and Faculty of Medicine, Dentistry & Health Sciences, University of Western Australia; 2) Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, University of Western Australia; 3) Department of Psychiatry, Yale University School of Medicine; 4) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley.

Schizophrenia (SZ) is a clinically heterogeneous disorder with multifactorial causes including a significant genetic contribution. Identifying endophenotypes of SZ will reduce heterogeneity and improve power for disease gene discovery. *Glahn et al* (2012) have described the *Endophenotype Ranking Value (ERV)* using family-based observations to assess heritability of the endophenotype and its genetic correlation with disease liability. W*e used the ERV to rank neurocognitive measures (candidate endophenotypes of SZ) in the Western Australian Family Study of Schizophrenia (WAFSS) and tested them for association with the Polygenic Risk Score for SZ.* Our cohort comprised 157 families (n=657, including 161 SZ cases). We tested cognitive measures across 6 domains: *General cognitive ability* (National Adult Reading Test, NART & Shipley Institute of Living Scale, SILS IQ); *verbal learning and memory* (Rey Auditory Verbal Learning Test, RAVLT); *sustained attention* (Continuous Performance Task, degraded stimulus, CPT-DS and identical pairs, CPT-IP); *personality factors* (Schizotypal Personality Questionnaire, SPQ & Temperament and Character Inventory, TCI); *speed of information processing* (inspection time, IT); *executive function* (Controlled Word Association Task, COWAT); plus two composite measures, cognitively deficit (CD) and cognitively spared (CS) (Hallmayer et al, 2005). Heritability (h_{CP}) and genetic correlation with SZ (ρ_g) was calculated for each measure in SOLAR and used to calculate ERV=$| \sqrt{h_{SZ}}\sqrt{h_{CP}}\rho_g |$ (h_{SZ} = heritability of SZ). PRS was calculated for each individual using genotypes at 102,636 SNPs (Psychiatric Genomics Consortium). The IQ measures were the most heritable; h^2= 0.56 (SILS) and 0.68 (NART) ($P<10^{-16}$). Genetic correlation was high (>0.5) for most traits. The highest ERVs were for SILS (0.67) and NART (0.55) IQ, RAVLT immediate (0.58) and delayed (0.54), and the CPT-IP (0.51). For all traits there was a trend towards increased neurocognitive deficit with increasing Polygenic Risk Score – especially for CPT-IP (P0.008) and SPQ (P0.02). Overall, measures of *general cognitive ability, verbal learning and memory and sustained attention* are the most promising neurocognitive endophenotypes for SZ among those tested here. Genetic analysis of these endophenotypes may identify latent variants previously missed due to the heterogeneity of the neurobiological disorders subsumed under the clinical diagnosis of SZ.

1493T

Genome-wide association studies and polygenic score analyses of suicide attempt in mood disorders. *N. Mullins[1], R.A. Power[1], A. McQuillin[2], E. Stahl[3], D. Mehta[4], C.M. Lewis[1,5], Major Depressive Disorder and Bipolar Disorder Working Groups of the Psychiatric Genomics Consortium.* 1) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience,, King's College London, United Kingdom; 2) Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, United Kingdom; 3) Department of Psychiatry, Mount Sinai School of Medicine, New York, USA; 4) Queensland Brain Institute, The University of Queensland, Brisbane, Australia; 5) Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, United Kingdom.

Background: Suicide is a global public health problem and the 10th leading cause of death in the U.S. Over 90% of suicide attempters or victims have a psychiatric diagnosis, particularly bipolar disorder (BPD) and major depressive disorder (MDD). However, twin and family studies have recognized a genetic component to suicidal behavior, which is independent of psychiatric disorders. Genome-wide association studies (GWAS) to date have failed to identify replicable genetic associations with suicide attempt. These studies have indicated that suicide attempt is a polygenic trait, arising from the combined effect of many risk alleles, each with small effect sizes. Here, we conduct GWAS and polygenic score analyses on suicide attempt using the largest available cohorts of patients with mood disorders, in the Psychiatric Genomics Consortium. **Methods:** Samples comprise 1075 suicide attempters and 7081 non-attempters with MDD and 2021 suicide attempters and 3253 non-attempters with BPD, as well as 18215 healthy controls from initial PGC1 studies. Two GWAS were performed in the MDD and BPD cohorts: a within-case analysis of suicide attempter cases vs. non-attempter cases and an analysis of suicide attempter cases vs. healthy controls. To further explore the genetic architecture of suicide attempt, polygenic risk scoring was used to test whether suicide attempt in MDD and BPD have a shared genetic etiology. **Results:** A genome-wide significant association was found on chromosome 14 comparing suicide attempters with MDD vs. healthy controls (rs8013144, P=8.60x10^{-11}, OR=2.2). No significant associations were found in the BPD sample or in meta-analysis of the cohorts. Polygenic risk scores for suicide attempt in BPD were associated with suicide attempt in MDD, indicating genetic overlap between suicide attempt in the mood disorders (P=0.01). We are currently incorporating new PGC2 data which will increase this sample size to over 4500 suicide attempters, 15500 non-attempters and 50000 controls. **Discussion:** This GWAS on suicide attempt is the largest to date. Polygenic risk scoring is a useful approach to investigating the genetics of a complex trait and should further our understanding of the genetic architecture of suicidality and its relationship with psychiatric disorders.

1494F

Integrated analysis of multiple types of rare variants to infer risk genes for schizophrenia. *T. Nguyen[1], D. Ruderfer[1,2], M. Fromer[1,5], P. Sklar[1,2], S. Purcell[1,2], X. He[3], P. Sullivan[4], E. Stahl[1,2].* 1) Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 2) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 3) Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA; 4) Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA; 5) Verily Life Sciences.

Integrating rare variation from family and case/control studies has successfully implicated specific genes contributing to risk of autism spectrum disorder (ASD). In schizophrenia, however, while sets of genes have been implicated through study of rare variation, very few individual risk genes have been identified. Here, we apply hierarchical Bayesian modeling of rare variation in schizophrenia and infer the proportion of risk genes and distribution of risk variant effect sizes across multiple variant annotation categories. Briefly, we employed the same model used previously in ASD studies. To simplify the complexity of the model, an approximation for the case-control model is used in which case variants are conditional on total counts. All classes of de novo and case-control variants are used jointly to infer genetic parameters using Markov Chain Monte Carlo. We applied this method to 1,024 trios and 4,954 cases/6,239 controls. We defined four variant annotation categories: disruptive (nonsense, frameshift, essential splice site mutations) and missense damaging *de novos* (predicting damaging by seven algorithms), and disruptive and missense damaging case/control singletons. We estimated that 8.4% of 19,441 genes are risk genes (95% credible interval 3.5-16%), with mean effect sizes (95% CIs) of 14.21 (5.04- 25.65) for disruptive *de novos*, 1.99 (1-3.99) for missense damaging *de novos*, 1.79 (1-2.94) for disruptive case/control singletons, and 1.56 (1-2.46) for missense damaging case/control singletons. Our analysis identified only three genes with FDR<0.1: SETD1A, TAF13 (both with FDR<0.05) and RB1CC1. We further analyzed the top 100 genes, with FDR<=0.496, for enrichment in several candidate gene sets. Significant results are observed in gene sets previously implicated in schizophrenia (including in a subset of these data): targets of the fragile X mental retardation protein genes (FMRP), the RNA binding proteins genes (Rbfox2/13), missense constrained genes, genes with ASD *de novo* mutations (all sets p < 7.8x10[-4]), and synaptic genes (p = 1.3x10[-3]). Overall, our results replicate previous studies for known gene sets as well as the SETD1A gene, indicating the robustness of the approach. We anticipate this approach will improve our power to detect schizophrenia risk genes as more data are included.

1495W

An in-depth characterization of common and rare coding variants in schizophrenia risk loci using targeted sequencing. *N. Norgren[1], J. Kang[1], M. Lam[2], J. Foo[1], E. Shyong Tai[3], J. Lee[2], J. Liu[1].* 1) Genome Institute of Singapore, Singapore; 2) Institute of Mental Health, Singapore; 3) Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

Genome-wide association studies (GWAS) have revealed more and more loci associated with schizophrenia, the largest and most recent study being from the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) study. They reported 108 loci, some previously implicated in the disease, and some novel, as associated with SCZ exceeding genome-wide significance. Our purpose with this study is to characterize the variant burden in the Asian schizophrenia population, and identify rare variants in genes increasing disease risk. For achieving this we apply multiplex targeted sequencing using Molecular Inversion Probes (MIPs), of 375 genes within the reported GWAS loci from the PGC study, to deep sequence the Singapore-Chinese schizophrenia cohort. 1585 cases and 2202 controls were sequenced to an average depth of 159 and 124X respectively. 12,205 variants passed quality controls, of which 58% were singletons. Single variant analysis revealed a nonsynonymous SNP on chr12, rs36121382, displaying a suggestive level of significance. Meta analysis with an additional 153 cases and 1900 controls increased the significance to P = 2.38e-05, OR = 1.23. This variant is much more frequent in the Asian population (maf = 0.3) than in the European (maf = 0.05). Rs36121382 is a nonsynonymous variant causing an amino acid change in the MPHOSPH9 gene, and also overlaps with an H3K27Ac mark, which are commonly found near regulatory sites. The rs36121382 variant might induce a dysregulation of the expression of MPHOSPH9 by altering its regulatory regions Burden tests were used to assess the burden of rare variants in cases for each gene. Some of the genes that displayed an increased burden (P < 0.05) are: AMBRA1, regulator of autophagy and nervous system development, DRD2, the D2 subtype of the dopamine receptor, and PCDHA6, a gene that may be involved in establishment and maintenance of specific neuronal connections in the brain. However, none of these genes could withstand Bonferroni correction, probably due to the relatively small sample size. Further sequencing on larger sample sizes are needed to confirm these associations. In conclusion, we found a nonsynonymous variant on chr12 associated with schizophrenia, which could potentially alter MPHOSPH9 expression. Further we discovered nominally significant genes involved in regulation of the nervous system with a higher burden of rare variants in cases.

1496T

Polygenic risk score associated with chronotype is associated with sleep problems and obesity in Chinese women from the CONVERGE study of major depressive disorder. *R.E. Peterson[1], T.B. Bigdeli[1], M. Lind[1], S.A. Bacanu[1], J. Flint[2], K.S. Kendler[1], CONVERGE Consortium.* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA.

Our innate, individual partiality for mornings or evenings, known as chronotype, is determined by circadian rhythms, and has relevant implications for numerous health-related outcomes. For example, chronotype has been associated with sleep problems, psychiatric disorders, and obesity. Twin and family studies indicate a partial genetic basis of chronotype, with estimates of heritability ranging from 12 to 42%. Recently, meta-analysis of two genome-wide association studies (GWAS) of European descent from 23andMe and UK Biobank yielded 15 genome-wide significant single-nucleotide polymorphisms (SNPs) associated with chronotype, including 7 loci near known circadian genes. The authors also reported phenotypic associations between chronotype and depression, body mass index (BMI), and educational attainment. Here, we sought to investigate the relationship between chronotype-associated loci and major depressive disorder (MD) in the CONVERGE (**C**hina, **O**xford a**n**d **V**CU **E**xperimental **R**esearch on **G**enetic **E**pidemiology) study of 5278 Han Chinese women with recurrent MD and 5,196 screened controls. Our strategy consisted of (1) constructing polygenic risk scores (PRS) from the 23andMe and UKBiobank meta-analysis of chronotype; (2) testing for associations between chronotype-PRS and MD, MD subtypes, MD symptoms (including sleep problems), BMI, and educational attainment; and (3) testing for enrichment of chronotype-associated genes in MD. Of the 15 reported GWAS loci for chronotype, only 1 was associated with MD in CONVERGE (rs3972456 in *FAM185A*, p=0.007). However, several top SNPs from the meta-analysis were associated with sleep problems among MD cases (i.e; rs12995069 in *VSNL1,* p=0.0001; rs2749951 in *BANF2,* p=0.002). A chronotype-PRS constructed from 456 SNPs (p<10-4 in meta-analysis) was not associated with MD but showed a nominal association in the expected direction with the melancholic MD subtype (p=0.020). Within MD cases, this PRS was nominally positively associated with trouble sleeping during MD episodes (p=0.034) but not with early-waking or "sleeping too much". The PRS was not associated with BMI but was nominally positively associated with obesity (p=0.042), which is consistent with the UKBiobank analysis. No association was found between chronotype-PRS and educational attainment. Preliminary results suggest a genetic link between chronotype-associated loci and MD subtypes--adding support for the role of circadian rhythmicity in psychological health.

1497F

NCAM1 interactions pathway association with depressive symptoms and their DTI correlates in healthy individuals: A gene set-based approach. *J. Petrovska[1], D. Coynel[2], M. Fastenrath[2], A. Milnik[1,4], B. Auschra[1], T. Egli[1], L. Gschwind[1], F. Hartmann[1], E. Loos[2], K. Spalek[1], C. Vogler[1], D.F. de Quervain[2,4], A. Papassotiropoulos[1,3,4], A. Heck[1,4].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Basel, Switzerland; 2) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Basel, Switzerland; 3) University of Basel, Department Biozentrum, Life Sciences Training Facility, Basel, Switzerland; 4) University of Basel, Psychiatric University Clinics, Basel, Switzerland.

Depressive symptoms exist on a continuum, the far end of which is found in depressive disorders. Therefore, utilizing the continuous spectrum of depressive symptoms may contribute to the investigation of the biological underpinnings of depression. Gene set enrichment analysis (GSEA) is a potent tool for identifying gene sets underlying complex traits, and is used in the present study to dissect the genetic background of depression scores and their brain-level structural correlates in two large samples of healthy young individuals. Robust enrichment was identified for two gene sets: NCAM1 Interactions and Collagen Formation, in association with depression scores. Depression scores were also associated with fractional anisotropy (FA) decrease within the forceps minor and the left (temporal) superior longitudinal fasciculus. The mean FA value for depression score-associated voxels within each of these tracts was used as a phenotype in a post-hoc GSEA. The NCAM1 Interactions gene set was significantly enriched across all tracts. By associating the NCAM1 Interactions gene set with depression scores and their imaging correlates in healthy participants, the current study suggests its relevance for depression, when operationalized as a continuous phenotype in a nonclinical population.

1498W

Investigation of *SHANK3* in schizophrenia suggests the variant G1011V as a rare risk factor shared with autism spectrum disorder. *G. Rappold[1,2], A. de Sena Cortabitarte[1], F. Degenhardt[3], J. Strohmaier[4], M. Lang[4], B. Weiss[1], R. Roeth[1], I. Giegling[5], S. Heilmann-Heimbach[3], A. Hofmann[3], D. Rujescu[5], C. Fischer[6], M. Rietschel[4,7], M. Nöthen[3,8], S. Berkel[1].* 1) Human Molecular Genetics, Institute for Human Genetics, Heidelberg, Germany; 2) Interdisciplinary Center for Neurosciences (IZN), University of Heidelberg, Germany; 3) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 4) Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health Mannheim, Medical Faculty Mannheim, University of Heidelberg, Germany; 5) Klinik für Psychiatrie, Psychotherapie und Psychosomatik, Universitätsklinikum Halle (Saale),Halle, Germany; 6) Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; 7) Department of Psychiatry, University of Bonn, Bonn, Germany; 8) Institute of Human Genetics, University of Bonn, Bonn, Germany.

The postsynaptic scaffolding protein SHANK3 is essential for the normal function of glutamatergic synapses in the brain. Emerging evidence suggests that impaired plasticity of glutamatergic synapses is partly responsible for the pathology of schizophrenia (SCZ). To investigate whether variants in the *SHANK3* gene contribute to the etiology of SCZ, we sequenced *SHANK3* in 500 affected individuals (cohort C1). In total, we identified 48 variants and compared them to European controls from the 1000 Genomes Project and the Exome Variant Server. The five most promising variants were followed up in an independent cohort (C2) comprising 993 SCZ patients and 932 German controls. We could not find an association for three of these variants (rs140201628, rs1557620 and rs61729471). Two rare variants with predicted functional relevance were identified in further SCZ individuals: c.3032G>T (p. G1011V) and c.*27C>T. The 3´UTR variant was found in one additional SCZ individual and the G1011V variant was identified in two additional SCZ individuals from cohort C2. The G1011V variant was the most interesting variant in our study; together with previous studies this variant has been identified in 4 out of 1524 SCZ patients and in 4 out of 2147 individuals with autism spectrum disorder (ASD), but not in 2468 European Sanger-sequenced controls. Therefore, we conclude that this variant may be a risk factor for both disorders. A comparison of deleterious *SHANK3* mutations comprising frameshift, nonsense, missense and splice variants between ASD and SCZ individuals revealed a higher frequency in individuals with ASD compared to SCZ.

1499T

Genetic mapping of mental disorder in the Greenlandic population. *A. Rosengreen[1], Y. Wang[1,3], C.B. Pedersen[2], P.B. Mortensen[2], J. Grauholm[4], D. Houggard[4], T. Hansen[5], I. Molkte[6], A. Albretchsen[6], T. Werge[1], W. Thomsom[1], A. Buil[1], IPSYCH-BROAD Researchers.* 1) Mental Health Centre Sct. Hans, Roskilde, Denmark; 2) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 3) NORMENT, institute of clinical medicine, University of Oslo Norway; 4) Statens Serum Institut, Copenhagen, Denmark; 5) The Novo Nordisk Foundation Center for Basic Metabolic Research, Copenhagen University, Denmark; 6) The Bioinformatics Centre at Copenhagen University, Demark.

Identifying the etiology and pathology of mental disorders is one of the more challenging problems in modern biomedical sciences. Even that some of them, like schizophrenia, bipolar disorder or autism have high heritabilities (>=70) little is known about the underlying genetic causes of these diseases. The inhabitants of Greenland constitute an isolated and admixed population with a unique genetic history that resulted in a unique allele frequency distribution that enhances the power to map genes related to complex diseases. Recent studies in type II diabetes found genome-wide significant results in a Greenlandic sample with only 223 cases and about 2000 controls. We hypothesize that we can find similar results for mental disorder traits and aim to find genes related to mental disorders in the Greenlandic population. We use a sample of 1708 individuals born in Greenland or with at least one parent born in Greenland. 905 of them have at least one diagnose for schizophrenia, bipolar disorder, autism, attention deficit disorder or depression and the rest are healthy controls. All subjects have been genotyped on the PsychChip. We use two complementary approaches to identify genetic loci related to disease: 1) Genome-wide ancestry-specific association-mapping that takes population admixture as well as family relationships into account. 2) Genomic segments Identical by Descent (IBD) were identified using BEAGLE and EMI and linear mixed models to test the association of genomic loci with mental diseases. Preliminary analysis GWAS analysis of 1213 samples shows a significant hit on chromosome 5 for autism spectrum disorder. Preliminary IBD analysis of 595 samples shows a suggestive signal for mood disorders in chromosome 3. In summary, the isolated population of Greenland improves the power to map genetic loci related to mental disorders. .

1500F

Identification of rare variants conferring risk for psychiatric disorders by whole exome sequencing of dried bloodspots. *F.K. Satterstrom[1,2], F. Lescai[3,4,5], C.R. Stevens[1], J.B. Maller[1,2], D.M. Hougaard[3,6], T.M. Werge[3,7,8], B.M. Neale[1,2,9], A.D. Børglum[3,4,5], M.J. Daly[1,2,9], iPSYCH-Broad Consortium.* 1) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; 4) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 5) iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; 6) Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 7) Mental Health Centre Sct. Hans, Institute for Biological Psychiatry, Capital Region of Denmark, Roskilde, Denmark; 8) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 9) Department of Medicine, Harvard Medical School, Boston, MA, USA.

To uncover rare genetic variants conferring risk for psychiatric disorders, the iPSYCH-Broad Consortium has begun a large-scale exome sequencing effort, drawing upon a wealth of existing biological and phenotypic data in Denmark. The Danish Neonatal Screening Biobank has stored dried neonatal blood samples for virtually every child born in the country since 1981, and methods have been developed by Denmark's Statens Serum Institut to recover sufficient DNA for high-throughput DNA sequencing from these samples. Using phenotypes from the Danish Psychiatric Central Registry, our consortium seeks to obtain whole exome sequences from the archived bloodspots of individuals who have been diagnosed with psychiatric disorders. Previous comparisons of sequences obtained from archived bloodspots with sequences obtained from fresh samples of whole blood (presented at this meeting last year) have demonstrated high quality (e.g. 99.9% concordance of heterozygote calls between the two sample sources), leading us to pursue a much larger study.　We present results from the combined analysis of approximately 20,000 bloodspot-derived exome sequences, including over 5,000 autism cases and 6,000 controls, as well as over 3,500 ADHD cases, 3,500 schizophrenia cases, and 1,000 bipolar cases. In a preliminary data freeze consisting of 8,000 exomes, we have identified significantly higher per-person rates of deleterious rare variants (not present in the Exome Aggregation Consortium database and causing truncation, frameshift, or essential splice site loss in proteins which have been identified as intolerant of loss-of-function mutation) in autism and ADHD compared to controls. As a further positive control, the genes with the most rare loss-of-function variants observed in autism cases include known autism- and intellectual disability-associated genes identified in *de novo* mutation studies, such as *SCN2A*, *SYNGAP1*, and *SHANK3*. While the presence of these variants in our data is likely due to *de novo* mutation, the scale of our data enables identification of new genes not previously flagged by *de novo* mutation analysis. Specifically, genes with lower penetrance, and therefore lower selective pressure, likely present with excesses of loss-of-function variants that will be inherited rather than *de novo*; for such variants, a very large sample size, such as that brought by this project, will be necessary to identify them as risk factors for psychiatric disorder.

1501W

GWAS of longitudinally collected brain MRI data in ADHD enriched samples. *T. Schwantes-An[1], E. Székely[2], C.M. Justice[1], R. Muetzel[3], P.R. Jansen[3], H. Tiemeier[3], T.J.H. White[3], P. Shaw[2], A.F. Wilson[1].* 1) Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, Baltimore, MD; 2) Neurobehavioral Clinical Research Section, Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) The Generation R Study Group, Erasmus University Medical Center, Rotterdam, the Netherlands.

In vivo magnetic resonance imaging (MRI) of the developing child has tied the course of childhood psychopathology to the trajectories of structural brain development. Several groups have associated attention-deficit/hyperactivity disorder (ADHD) with atypical neuroanatomical growth trajectories of the cerebral cortex, cerebellum and striatum. Using repeated MRI measures of brain volume from two independent cohorts, we performed a genome wide association study (GWAS) of brain growth during childhood and adolescence. We focused on six key regions of the brain: cerebellum, cerebral cortex, left and right lateral prefrontal cortex, basal ganglia and cortical white matter. The GWAS was conducted on 458 individuals, of whom 119 had a diagnosis of ADHD, using EMMAX (Efficient Mixed-Model Association eXpedited) to control for population stratification and sample relatedness. In the second stage, an independent cohort of 260 individuals also with an *in vivo* brain growth phenotype was tested with EMMAX. Meta-analysis was then performed with the rMETA package in R, focusing on the 288,218 SNPs that were genotyped in both cohorts. No significant associations were found after Bonferroni correction (2.89×10^{-08}). However, there were 141 suggestive hits at a critical value of 1×10^{-05} across the six brain compartment measures. The most significant association was between rs1424450 (chr7:134,209,317), an intergenic variant, and white matter with a p-value of 3.06×10^{-08}. Rs7089041 (chr10:125,003,630) showed suggestive associations with left (p-value = 5.26×10^{-06}) and right (9.46×10^{-05}) prefrontal cortex and cerebral (2.74×10^{-05}) cortex traits, and is located near GPR26, a gene reported to be expressed in mouse cerebral cortex. Overall, we report 141 suggestive associations between SNPs and brain growth in a sample enriched for behavioral problems characteristic of ADHD. The lack of significance at the Bonferroni level is likely due to the small sample size, a reflection of the scarcity of longitudinal MRI measures. Further work in prioritizing the nominally associated variants and implicated biological pathways is in progress. Additionally, we will explore the genetic overlap between our brain growth phenotypes and the diagnosis of ADHD using polygenic risk score analysis.

1502T

Schizophrenia polygenic burden is significantly associated with cognitive performance in nonpsychotic individuals. *R. Shafee[1,3], P. Nanda[4], N. Tandon[5,6], J. Padmanabhan[5,7], N. Alliey-Rodriguez[8], E. Gershon[8], B. Clementz[9], G. Pearlson[10], C. Tamminga[11], M. Keshavan[2,6], S. McCarroll[1,3].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Psychiatry, Harvard Medical School, Boston, MA; 3) Stanley Center, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Neurological Surgery, Columbia University Medical Center, New York, NY; 5) Psychiatry, Beth Israel Deaconess medical Center, Boston, MA; 6) Baylor College of Medicine, Houston, TX; 7) McLean Hospital, Belmont, MA; 8) Psychiatry and Human Genetics, University of Chicago, Chicago, IL; 9) Psychology, University of Georgia, Athens, GA; 10) Psychiatry, Yale University School of Medicine, New Haven, CT; 11) Psychiatry, UT Southwestern Medical School, Dallas, TX.

   Psychiatric disorders such as schizophrenia are frequently accompanied by cognitive deficits. Probands as well as their unaffected relatives show decreased functioning in many cognitive domains. The Brief Assessment of Cognition in Schizophrenia (BACS) is a cognitive battery designed to capture the deficits that are seen frequently in schizophrenia. We used data from the Bipolar-Schizophrenia Network for Intermediate Phenotypes (BSNIP) consortium to investigate the effect of the polygenic burden of schizophrenia common variants (PRS_SZ) on BACS performance as well as on Wide Range Achievement Test (WRAT, often used as a proxy for premorbid IQ) in two groups: psychotic probands (N = 314) and nonpsychotic individuals consisting of healthy controls and non-psychotic relatives (N = 438). Our BACS score consisted of six components: verbal memory, verbal fluency, digit sequencing, digit symbol coding, token motor task and Tower of London. We found that PRS_SZ is significantly negatively correlated with digit sequencing (DS) and digit symbol (DSC) coding scores in the nonpsychotic group (combined r = -0.21, $p<10^{-4}$) but not in the psychotic group (p>0.05). PRS_SZ showed no significant correlation with WRAT in either group. Additionally, the correlation between BACS and PRS_SZ in the nonpsychotic group remained unchanged after regressing out WRAT. We noted that while DS and DSC BACS scores differed significantly between the psychotic and nonpsychotic groups ($p<10^{-10}$), WRAT was only nominally different (p~0.05 after controlling for years of education). Additionally, a polygenic score for educational attainment showed significant positive correlation with WRAT ($p<10^{-3}$) but not with BACS (p>0.05). Together these results indicate that in our sample schizophrenia polygenic burden captures the variability in cognitive functions affected in schizophrenia much better than the variability in overall intelligence. The absence of significant correlation between PRS_SZ and BACS components in the probands could be due to the effects of medication or due to disease pathology altering these cognitive functions in a way that could not be captured by genetic predisposition to schizophrenia. .

1503F

Schizophrenia patients have an increased burden of rare disruptive variants in genes shared with neurodevelopmental disorders. *T. Singh[1], J.T.R. Walters[2], M. Johnstone[3], E. Rees[2], J. Suvisaari[4], M. Torniainen[4], D. Curtis[5], A.M. McInstosh[3], D. Blackwood[3], G.K. Kirov[2], A. Palotie[6], M.C. O'Donovan[2], M.J. Owen[2], J.C. Barrett[1] on behalf of the UK10K consortium.* 1) Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK; 2) MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, UK; 3) Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK; 4) National Institute for Health and Welfare (THL), Helsinki, Finland; 5) University College London (UCL) Genetics Institute, London, UK; 6) The Broad Institute of MIT and Harvard, Cambridge, USA.

   Substantial progress has been made using whole-exome sequencing to identify genes in which damaging rare variation or *de novo* mutations confer very high risk for autism, intellectual disability, and severe developmental disorders. Indeed, these studies, which have largely been conducted independently, have revealed that many of the same genes are disrupted in patients with a wide range of diagnoses and presentations. In a recent meta-analysis of schizophrenia exomes, we identified at genome-wide significance the first such gene (*SETD1A*) for which loss of function (LoF) variants confer substantial risk for schizophrenia, an adult onset neuropsychiatric disorder where the genetic connections to autism and developmental disorders are less clear. Intriguingly, LoF variants in the same gene were also found to confer risk for severe developmental disorders.   Here, we describe a series of analyses based on large-scale genetic datasets that further explores the potential overlap of genetic risk between schizophrenia and broader developmental disorders. We jointly analyzed data from 1,077 schizophrenia trios, 4,264 case and 9,343 control exomes, and array-based CNV calls from 6,882 cases and 11,255 controls. We first show that schizophrenia individuals carry a higher burden of rare damaging variants in a subset of 3,230 genes with near-complete depletion of truncating variants ($P < 1x10^{-14}$). Rare variant enrichment analyses in 1,776 gene sets demonstrate that this burden is most strongly enriched in known autism risk genes ($P = 1.6x10^{-5}$), genes diagnostic of severe developmental disorders ($P = 3.5x10^{-5}$), and the autism-implicated sets of promoter targets of *CHD8* ($P = 5.7x10^{-7}$), and splice targets of *RBFOX* ($P = 3.3x10^{-4}$). Furthermore, in a subset of 119 schizophrenia patients with learning disability, we show this burden is even stronger than in the general schizophrenia population ($P = 7.3x10^{-5}$), mirroring previous results comparing low and high IQ autism probands. Notably, a general enrichment of rare variants still exists in the larger schizophrenia population, including in patients without cognitive impairment. Combined, our results demonstrate that schizophrenia risk loci of large effect implicate a common set of genes shared with broader neurodevelopmental disorders, further supporting a neurodevelopmental etiology to the pathogenesis of schizophrenia.

1504W

Induced pluripotent stem cells as a tool to analyze the molecular effects of the 15q13.3 microdeletion. *S. Zhang, C. Purmann, J. Hallmayer, A. Urban.* Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA.

The 1.5 Mb microdeletion on human chromosome 15q13.3 is strongly associated with a number of neurodevelopmental symptoms, including developmental disability, seizures, speech problems, schizophrenia and autism spectrum disorder. However, there are at least eight genes affected by the microdeletion, and quite possibly in a manner specific to tissue and developmental time point, and novel approaches are needed to analyze the deletion's effects on the molecular and cellular level. Using the Sendai virus approach, we have created a cohort of induced pluripotent stem cells (iPSCs), from fibroblasts from three patients with the 15q13.3 deletion, and from matching controls. The iPSCs' genomes have been characterized with digital droplet PCR and whole-genome sequencing. The cell lines were characterized by staining for the pluripotency markers Nanog, Tra-1-60, and SSEA-4. RNA-Seq analysis of gene expression patterns was carried out for both the fibroblast and iPSC stage and showed mostly decreased gene expression levels, in a cell type specific manner, for the genes within the deletion boundaries, as well as genome wide effects on the transcriptome. Furthermore, we induced the iPSCs into a neuronal cell type (induced Neurons, iNs), using the protocol described in Zhang et al 2013. In summary, we have created a cohort of iPSCs from patients with the 15q13.3 microdeletion, the lines are well characterized on the genome level and on the level of gene expression patterns, and they show distinct potential to differentiate into a neuronal cellular phenotype. We expect this cohort to be a useful tool in the analysis of the molecular effects to the 15q13.3 microdeletion.

1505T

Genome-wide association study of post-cannabis dependence anxiety. *H. Zhou[1], H.R. Kranzler[2,3], R. Sherva[4], B.Z. Yang[1], J.L. Montalvo-Ortiz[1], L. Farrer[4,5,6,7,8], H. Zhao[9,10], J. Gelernter[1,10,11,12].* 1) Division of Human Genetics, Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 2) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 3) VISN 4 MIRECC, Crescenz VA Medical Center, Philadelphia, PA, USA; 4) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 5) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 6) Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA; 7) Department of Genetics and Genomics, Boston University School of Medicine, Boston, MA, USA; 8) Department of Epidemiology and Biostatistics, Boston University School of Public Health, Boston, MA, USA; 9) Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA; 10) Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; 11) VA CT Healthcare Center, Department of Psychiatry, West Haven, CT, USA; 12) Department of Neurobiology, Yale University School of Medicine, New Haven, CT, USA.

Cannabis use disorder and anxiety disorders often co-occur. Cannabis use is associated with new onset of panic disorder and can worsen anxiety symptoms. Thus, anxiety can be viewed as an adverse consequence of cannabis abuse. We examined genetic factors influencing cannabis-associated anxiety disorder onset by comparing cannabis dependence (CaD) subjects with anxiety onset after CaD onset (case group) to all other CaD subjects (i.e., those whose onset of anxiety disorder was before that of CaD or who had no anxiety disorder (control group)). Subjects were recruited from 2000 to 2013 as part of the Yale-Penn study and were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), with CaD diagnosed according to DSM-IV criteria. Anxiety disorder onset was defined as the earliest time of diagnosis of any one of the following anxiety disorders: generalized anxiety disorder, panic disorder, agoraphobia and social phobia. Subjects included 1,185 CaD African Americans (AAs) and 1,249 CaD European Americans (EAs) in two datasets. The first dataset comprised 61 AA cases, 806 AA controls, 90 EA cases and 662 EA controls, which were genotyped using the Illumina HumanOmni1-Quad array (~890,000 SNPs). The second dataset comprised 21 AA cases, 297 AA controls, 64 EA cases and 433 EA controls, which were genotyped using the Illumina HumanCore Exome array (~500,000 SNPs). Additional genotypes were imputed with the 1000 Genomes Project panel. We conducted genome-wide association studies (GWAS) and meta-analysis in each population. A short deletion, rs201796339, was significantly associated with anxiety onset after CaD in EAs ($P = 2.09 \times 10^{-8}$ in dataset 1, $P = 2.83 \times 10^{-3}$ in dataset 2, and $P = 4.37 \times 10^{-10}$ for meta-analysis). This variant is ~60kb downstream of *NUDT12* (nudix hydrolase 12), which is sparsely annotated and has been demonstrated to play a role in NAD metabolism. This study provides insight into the genetics of anxiety disorder onset in the context of CaD.

1506F

Genetic polymorphisms and serum concentrations of adiponectin and resistin in anorexia nervosa and healthy controls. *K. Ziora-Jakutowicz[1], J.G. Zimowski[1], F. Rybakowski[2], M. Bednarska-Makaruk[1], J. Zaremba[1], M. Pawelec[1], D. Narowska[1], K. Ziora[3].* 1) Department of Genetics, Institute of Psychiatry and Neurology, Warsaw, Poland; 2) Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Neurology, Warsaw, Poland; 3) Department of Paediatrics, School of Medicine with the Division of Dentistry in Zabrze, SUM in Katowice, Poland.

   Introduction- Anorexia nervosa (AN) is a heritable disorder characterized by low body weight and disturbances of appetite. A possible role of adipokines in the regulation of body mass has been proposed. Polymorphisms in genes encoding adiponectin and resistin in AN have not been frequently assessed. **Aims of the study-**1. To assess the frequency of selected polymorphisms in adiponectin gene: *ADP*c.45 T>G and *ADP* c.276G>T and in resistin gene: *RETN*c.62 G>A and *RETN*c.-180C>G in AN patients and healthy controls (HC); 2. To assess the correlation between specific polymorphisms with adiponectin and resistin plasma levels. **Material and methods-** We examined 67 AN girls (aged 11-18) and 38 controls (aged 11-18). Analyses of polymorphisms in adiponectin (*ADP*) and resistin (*RETN*) genes were performed using RFLP method, after the DNA was isolated from frozen samples of peripheral blood. Adiponectin and resistin serum levels were examined with the use of commercial kits for ELISA [Bio-Vendor, LLC (USA)]. **Results-** We observed a statistically significant differences in allele frequency in polymorphic sites *ADP* c.276 and *RETN*c.-180 between AN patients and HC. In AN subjects, TT genotype in polymorphism *ADP* c.276 was significantly more frequent than in HC (p=0.005; contingence index C=0.31). Also, GG genotype of *RETN*c.-180 in AN patients was significantly more frequent, than in HC (p<0.0001; C=0.47), in whom CC alleles were predominant. In *ADP*c.45 polymorphic site, TT alleles were the most frequent in AN (93.8%), as well as in HC (78.9%). Alleles TG in this site were more frequent in HC (21.1%) than in AN (6.2%), but this difference was not statistically significant (p=0.051; C=0.19). In *RETN* c.62 alleles distribution of GA and GG did not differ significantly in both groups (p=0.591; C=0.05). The most frequently observed genotype was GG in both AN (86.57%) and HC (92.11%). The mean adiponectin serum level in AN (43.8 ± 5.3 µg/ml) was significantly higher than in control group (6.52 ± 1,3 µg/ml; p< 0.0001). The mean resistin serum level in AN (2.75 ± 0.60 ng/ml) was significantly lower than in control group (9.15 ± 1.38 ng/ml).There were no statistically significant relationships between mean adiponectin and resistin serum levels and allele frequency in polymorphisms *ADP*c.276 and *RETN*c.-180 in neither group. **Conclusion-** Differences in genotype frequencies of *ADP*c.276 and *RETN*c.-180 suggest a need for studies on a larger cohort.

1507W

Comprehensive genomic characterization of long non-coding rnas of amygdala tissues in schizophrenia. *Z. Wei[1], T. Tian[1], X. Chang[2], Y. Liu[2], P. Sleiman[2,3], H. Hakonarson[2,3].* 1) New Jersey Institute of Technology, Newark, NJ; 2) The Children's Hospital of Philadelphia, Philadelphia, PA; 3) The Perelman School of Medicine University of Pennsylvania, Philadelphia, PA.

   Schizophrenia is a chronic and severe mental disorder with which people will exhibit abnormal social behavior and/or fail to differentiate real from unreal. The amygdala brain region has been implicated in the pathophysiology of schizophrenia through emotion processing. Long noncoding RNAs (lncRNAs) have gained widespread attention in recent years as a potentially new and crucial layer of biological regulation. lncRNAs of all kinds have been implicated in a range of developmental processes and diseases, but very few, if any, studies have investigated long Non-coding RNAs (lncRNAs) in the amygdala of schizophrenia. To delineate genome-wide lncRNA expression, we curated 22 schizophrenia patients and 24 non-psychiatric controls through deep RNA-sequencing (over 100 M reads on average per sample). We applied *ab initio* assembly methodology to this data set, and identified a consensus transcriptome of 839 novel lncRNAs of high quality after stringent QC. Gene Set Enrichment Analysis (GSEA) of these novel lncRNAs reveal that many pathways relevant to neuron development are significantly down-regulated in schizophrenia patients vs controls. These findings may shed light on unraveling the molecular mechanism of schizophrenia.

1508T

Investigation of gene dosage sensitivity in 22q11.2 deletion syndrome.
A. Salaka, H. Moss, M. Niarchou, S. Chawner, M. Van Den Bree, M.J. Owen, N.M. Williams. MRC Centre for Neuropsychiatric Genetics& Genomics, Cardiff University, Cardiff, United Kingdom.

Individuals with chromosome 22q11.2 deletion syndrome (22q11.2DS) have an increased risk of developing a diverse spectrum of behavioural and psychiatric outcomes. However, there is a lack of understanding of the mechanisms explaining how such a degree of pleiotropy could arise from a relatively homogenous microdeletion on chromosome 22q11.2. As 22q11.2DS patients carry the same deletion it is most likely that haploinsufficiency of dosage sensitive genes spanning the deleted region influence the pathogenesis of the resulting psychiatric phenotypes. It is expected that deletion of 22q11.2 will have an impact on the level of gene expression of deleted genes which could be reduced by 50%. Consequently, reduction in the expression of these genes would change their biological function. We set out to identify dosage sensitive genes by differential gene expression analysis of the 22q11.2 deletion region and genome-wide in 22q11.2 deletion group and a normal control group. As not all genes are dosage sensitive the identification of genes that are differentially expressed between 22q11.2 deletion carriers and non-deleted controls will highlight those whose expression are influenced by the deletion. The expression of genes away from the CNV boundaries can be directly influenced the deletion. This phenomenon is called a 'position effect' which was investigated by performing genome-wide gene expression analysis to assess the expression of genes located outside 22q11.2. For gene expression analysis RNA samples of 22q11.2DS cases (n=33) and controls (n=36) were analysed by Illumina HumanHT-12 v4 Expression BeadChip which targets 47,231 probes. Samples and probes were subjected to quality control assessments. Differential gene expression analysis was applied by fitting a linear model and defining a contrast comparing cases to controls using Limma package on R. The result of differential gene expression analysis shows ~35% of genes located within the 3MB 22q11.2 region are significantly differentially expressed between 22q11.2 deletion cases and controls (FDR<0.05). There is a minor *in CIS* position effect of 22q11.2 deletion on nearby genes. However, the deletion has a negligible effect genome-wide, with only 2.2% of genes being significantly differentially expressed with FDR<0.05. As these studies have used RNA samples which were extracted from peripheral blood, future studies will be required to establish how they reflect gene expression in the human brain.

1509F

Childhood maltreatment and major depressive disorder: Indirect mediating effect of gene expression. *L. Spindola[1,2,3,4], P. Pan[2,3,4], P. Moretti[1,2,3,4], V. Ota[1,2,3,4], M. Santoro[1,2,3,4], H. Cogo-Moreira[2,3,4], A. Gadelha[2,3,4], G. Salum[3,8], G. Manfro[3,8], J. Mari[2,3,4], H. Brentani[3,5], R. Grassi-Oliveira[3,6,7], E. Brietzke[2,3,4], E. Miguel[3,5], L. Rohde[3,8], J. Sato[3,9], R. Bressan[2,3,4], S. Belangero[1,2,3,4].* 1) Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil; 2) LiNC - Interdisciplinary Laboratory of Clinical Neurosciences, UNIFESP, São Paulo, Brazil; 3) National Institute of Developmental Psychiatry for Children and Adolescents (INPD), Brazil; 4) Department of Psychiatry, UNIFESP, São Paulo, Brazil; 5) Department & Institute of Psychiatry, Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, Brazil; 6) Post-Graduation Program in Psychology, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil; 7) Developmental Cognitive Neuroscience Lab (DCNL), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil; 8) Department of Psychiatry, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre (UFRGS), Brazil; 9) Center of Mathematics, Computation and Cognition, Universidade Federal do ABC, Santo André, Brazil.

Investigating major depressive disorder (MDD) in childhood and adolescence can help reveal the relative contributions of genetic and environmental factors to MDD, since early stages of disease have less influence of illness exposure. Thus, we investigated the mRNA expression of 12 genes related to the hypothalamic–pituitary–adrenal (HPA) axis, inflammation, neurodevelopment and neurotransmission in the blood of children and adolescents with MDD and tested whether a history of childhood maltreatment (CM) affects MDD through gene expression. Whole-blood mRNA levels of 12 genes were compared among 20 children and adolescents with MDD diagnosis (MDD group), 49 participants without MDD diagnosis but with high levels of depressive symptoms (DS group), and 61 healthy controls (HC group). The differentially expressed genes were inserted in a mediation model in which CM, MDD, and gene expression were, respectively, the independent variable, outcome, and intermediary variable. *NR3C1*, *TNF*, *TNFR1* and *IL1B* were expressed at significantly lower levels in the MDD group than in the other groups. CM history did not exert a significant direct effect on MDD. However, an indirect effect of the aggregate expression of the 4 genes mediated the relationship between CM and MDD. In the largest study investigating gene expression in children with MDD, we demonstrated that *NR3C1*, *TNF*, *TNFR1* and *IL1B* expression levels are related to MDD and conjunctly mediate the effect of CM history on the risk of developing MDD. This supports a role of glucocorticoids and inflammation as potential effectors of environmental stress in MDD.

1510W

Genetic modifiers in patients with autism and Williams-Beuren syndrome. *M. Codina-Solà[1,2,3], D. Pérez-García[1,2,3], R. Flores[1,2,3], G. Palacios-Verdú[1,2,3], L. Pérez-Jurado[1,2,3], I. Cuscó[1,2,3].* 1) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain; 2) Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; 3) Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain.

The hallmark of the neurobehavioral phenotype of Williams-Beuren syndrome (WBS) is increased sociability and relatively preserved language skills, often described as the opposite to Autism Spectrum Disorders (ASD). However, the prevalence of ASD among WBS individuals is 12%, approximately 10 fold higher than in the general population. To look for phenotypic modifying factors, we have performed an unbiased study of genetic variants in individuals with WBS and co-occurring ASD. The study was conducted on eight individuals with WBS (4 males and 4 females) and a confirmed diagnosis of ASD by the Autism Diagnostic Interview-Revised (ADI-R). We first performed a detailed characterization of the deletion at the molecular level in order to find cis-acting modifiers. In addition we looked for genome-wide variants, both rare and common, using exome sequencing. The detailed study of the 7q11.23 region showed that patients did not differ in deletion breakpoints, but in seven of eight the deletion had originated in the paternal allele. In addition, we found an overrepresented rare SNP (rs12539160) in the remaining hemizygous allele in two individuals, which was previously associated to ASD by Genome-Wide Association studies. Our genome-wide analysis found large Copy Number Variants (CNVs), but identified several loss of function (LoF) mutations in ASD candidate genes and brain-expressed genes with a high probability of LoF intolerance. Some of these loci, such as *AGAP1* or *EPHB1*, are transcriptionally dysregulated in WBS. In addition, females carried a higher burden of rare deleterious variants in candidate genes compared to males (p=0.03). The increased frequency of paternal deletions in our cohort suggests a role of imprinting mechanisms that along with trans-acting factors in the remaining allele, such as ASD susceptibility loci, could influence the neurobehavioral profile. Our results also point towards a role for deleterious variants in functionally constrained brain expressed genes, that could act as second-hit modifying factors in an epistatic manner. Finally, the higher load of deleterious variants in ASD candidate genes in females reinforces the female protective effect for a more severe phenotype in a disorder of full penetrance.

1511T

Meta-analysis of 4,500 ASD exomes reveals novel genes and connections to developmental disorders. *L. He, J.C. Barrett, The DDD Study.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Authors: Liu He, Jeffrey C. Barrett, the DDD Study Recent research based on whole-exome sequencing has established a prominent role of *de novo* mutations in the etiology of sporadic autism spectrum disorders (ASD). However the rarity of *de novo* mutations, especially those expected to cause loss-of-function (LoF) through protein truncation, as well as the number of possible affected genes, require very large sample sizes to reliably identify risk genes. Here, using aggregated data from ~4500 published families affected by ASD, we re-examine the pathogenic contribution of *de novo* LoF mutations. By applying a stringent exome-wide significance cutoff, 15 genes were classified as ASD genes, among which 11 have been previously reported with high confidence and four more —*ANKRD11, ASXL3, MED13L, FOXP1* — though previously implicated in ASD by other means, have now also been confirmed based on a statistically significant enrichment of *de novo* LoF mutations in cases. Through comparison with recent published research from the DDD project focusing on *de novo* mutations in ~4300 families affected by severe developmental disorders (DD), 12 of these 15 genes were also associated with DD, which points to the converging genetics of *de novo* damaging variants and their substantial contribution to neurodevelopmental risk. In general, highly damaging *de novo* variants confer risk for a broad spectrum of neurodevelopmental and neuropsychiatric disorders, including ASD, SCZ, and DD. Intriguingly, in addition to the general overlap, we further find a significant (p= 0.001804) ~2-fold enrichment of *de novo* LoF mutations in ASD genes in DD probands also diagnosed with ASD, compared to the rest of DD probands. These findings support the repeatedly observed convergence of the biology and etiology between neurodevelopmental disorders and also illustrates the existence of potentially distinctive genetic architecture specific to different disorders, namely ASD & DD. .

1512F

Targeted deep sequencing reveals significant enrichment of putatively damaging rare variants in reading and language genes. *A.K. Adams[1], D.T. Truong[2], M.M.C. DeMille[2], B.F. Pennington[3], S.D. Smith[4], R.K. Olson[5], J.R. Gruen[1,2].* 1) Department of Genetics, Yale University, New Haven, CT; 2) Department of Pediatrics, Yale University, New Haven, CT; 3) Department of Psychology, University of Denver, Denver, CO; 4) Department of Pediatrics, University of Nebraska at Omaha, Omaha, NE; 5) Department of Psychology, University of Colorado at Boulder, Boulder, CO.

Reading disability (RD; dyslexia) is a common disorder with prevalence estimates ranging from 5% to 17%. Individuals with RD fall behind their peers by middle school in terms of reading performance, and remain impaired. Without intervention, affected individuals have lower SES and poorer educational outcomes over the course of their lives. Twin and family studies have demonstrated moderate to high heritability values of 0.4-0.8 for RD. Given the strong genetic component, a genetic screen could be used to identify children at risk, allowing them to receive intervention early, when it is most effective. Understanding the full genetic landscape of RD has proved troublesome, with known associations not accounting for a large portion of the heritability. This is at least partly due to a lack of large population or case-control cohorts limiting the ability to identify common variants through GWAS, and a paucity of resequencing reports of genes previously associated with reading and language to identify rare variants of high putative effect. Utilizing a targeted sequencing approach, 11 curated loci for reading and language performance were selected for deep sequencing from 96 RD cases. Cases were selected from the Colorado Learning Disability Research Center collections based on poor performance on a well-studied composite metric derived from the Peabody Individual Achievement Test. Loci of interest were isolated and sequenced. Resulting sequence was aligned to human reference genome hg19, and observed variants were scored for putative effect using SIFT and Polyphen-2. Two genes, *CCDC136* and *FLNC*, both encoded on 7q32, had 17 missense variants in 192 case chromosomes, showing a high disproportional enrichment of rare (MAF <0.05), and putatively damaging (Polyphen-2 score > 0.8) coding variants relative to the 23 other genes that were sequenced. Comparing these results to 503 individuals of European ancestry from the 1000 Genomes Project unselected for reading performance, revealed a nearly 2-fold enrichment of identified variants (2-tailed Fisher's Exact P= 0.0392) in these two genes. Other regions demonstrating disproportional enrichment of rare variants of high effect included the *DCDC2-KIAA0319* locus, within which the gene *FAM65B* (12 variants) was prominent. Taken together, RD follows the same pattern of other complex disorders, with enrichment of rare variants of high putative effect in reading and language genes.

1513W

Consanguinity negatively affects scholastic success. *N. Ghiasvand[1,2], H. Gholamhoseini[3], F. Rashidi[4], E. Samarbafzadeh[4], A. Solgi[1], R. Ghiasvand[5], F. Yaghoubi[1], M. Zabihi[1], S. BaniArdalan[1], M. Motejali[1], Z. Jahani[1], M. Sinclair[2], K. Szostak[6], N. Maghsoudi[4], A. Zali[1].* 1) Functional Neurosurgery Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) Biology Department, Grand Valley State University, Allendale, Michigan, USA; 3) Radiology Department, Modares Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 4) Neuroscience Research Institute, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 5) Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo, Norway; 6) Biomedical Sciences Department, Grand Valley State University, Allendale, Michigan, USA.

BACKGROUND: The association between consanguinity and simple autosomal recessive traits has long been established. However, some recent studies indicate a possible association between a number of complex traits and consanguinity. To investigate this controversial issue with regard to education educability we studied the relationship between consanguinity and scholastic success. METHOD: In Iran, higher education of children is of high priority for families, and annually more than a million competitors compete in the nation-wide university entrance exam for less than 100,000 seats. For this study the average coefficients of inbreeding (α) for four student groups with normal IQ, and also for their parents were calculated and compared. The four student groups included: 1) Struggling students who appeared to be incapable of obtaining a high school diploma. 2) Students who could successfully finish high school. 3) Students enrolled in a 4-year B.Sc. program in a top public university in Tehran. 4) Medical and engineering students who ranked among the top 400 in the nation-wide competition. For this study students were recruited after giving informed consent and obtaining internal review board approval. RESULTS: The coefficients of inbreeding for these groups were significantly different from each other; smaller values for those with greater scholastic success. Compared to their parents, the coefficients of inbreeding for groups 1 and 4 were significantly larger and smaller, respectively. The coefficients of inbreeding for groups 2 and 3 were not significantly different from those for their parents. CONCLUSION: Apparently, consanguinity results in inbreeding depression for scholastic success, suggesting that the genes involved with this complex trait act in a directional dominance - not in an additive model that has long been postulated for such traits. Based on this finding, countries with high prevalence of consanguinity may increase their national IQ and also their people's scholastic achievement by educating the public about the undesirable consequences of consanguineous marriages. .

1514T

Compound effects of copy number variants del 1q21.2 and dup 17q12.
A. Goldberg[1], C.K. Stein[2], R. Auerbach[3], L. Pellegrino[1], M. Byler[1], R.R. Lebel[1].
1) Center for Development, Behavior and Genetics, Dept Pediatrics, SUNY Upstate Medical University, Syracuse, NY; 2) Cytogenetics Laboratory, Dept Pathology, SUNY Upstate Medical University, Syracuse, NY; 3) Binghamton, NY.

A 4 year old female presented with the constellation of poor growth, developmental delay, dysgraphia, coordination disorder, language disorder, chronic cough, hoarse voice, myopia, and ADHD. She had been the product of a non-consanguineous union, born at 39 weeks of gestation by vaginal route to a 32 year old primigravid Caucasian woman whose pregnancy was uncomplicated. Paternal age was 35 at that time. The family history was notable for learning disabilities in both parents, and the mother is short of stature. Physical examination for the patient was notable for short stature, thin habitus, microcephaly, crowded dentition, hoarse voice, fifth finger clinodactyly, joint laxity, low pedal arches, and hypopigmented rings of spots on the irides (reminiscent of Brushfield spots). Microarray analysis revealed two clinically significant copy number variants (CNV), a 2.4 MB deletion of 1q21.2 and a 1.9 MB duplication of 17q12. Microarray analysis for the parents was declined under their insurance plan, but fluorescent in situ hybridization (FISH) analysis showed the CNVs to be of maternal (1q21.2) and paternal (17q12) origin. To our knowledge, there are no previous reports of this combination of anomalies, and the constellation of features is inconsistent with any existing identified syndrome. Girirajan et al (2012) found that patients with more than one copy-number variant often have more dramatic presentations than what would be expected with either of the copy-number variants alone. This appears to be the case for our patient as she manifests a more adverse phenotype than is found in either of her parents.

1515F

Complex form of mental retardation identified by next generation sequencing. *D. Panahi Zadeh.* Social Insurance, Ahvaz, Iran.

We found a nonsense change at c.486C>G, p.Tyr162* of the FAM126A gene that is to date unreported and regarding prediction programs is likely-pathogenic. This gene is also reported to the cause Leukodystrophy, hypomyelinating, 5 characterized by weakness and wasting of the lower limbs, scoliosis, progressive, developmental delay (apparent after the first year of life), delayed motor development, mental retardation, mild to moderate, cerebellar signs, abnormal folding of the myelin sheath, peripheral neuropathy with variable severity. Affected son is homozygous for detected change. His parents, and his healthy brother and sister are heterozygous. .

1516W

6q26 and 16p12.2 heterozygous deletions support oligogenic model for autism spectrum disorder. *C.C.V. da Silva-Camargo[1], J. Rosenfeld[2], S. Raskin[1,3], V.S. Sotomaior[1].* 1) Group for Advanced Molecular Investigation (NIMA), Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil; 2) Signature Genomics, PerkinElmer, Inc., Spokane, Washington, USA; 3) GENETIKA – Centro de Aconselhamento e Laboratorio de Genetica, Curitiba, Paraná, Brazil.

The development of autism spectrum disorder (ASD) is related to multiple factors in particular the genetics ones. With the highest heritability observed for commons neuropsychiatric disorder, ASD is caused frequently by copy number variations (CNVs). Studies of the genetic basis of this disorder have identified a large number of candidate mutations and genes, supporting an oligogenic model for the development of this disease. Through microarray-based comparative genomic hybridization (aCGH) we detected two microdeletions (in 6q26 and 16p12.2) in a 10-year-old autistic female with developmental and speech delay, seizures, difficulty with coordination and literacy, repetitive thoughts and movements and tendency to social isolation. The 6q26 deletion is entirely within the *PARK2* gene, which encodes the protein parkin that has important roles in mitochondrial homeostasis and degradation of damaged proteins and has been associated with development of ASD as well as other neurodegenerative diseases. The recurrent 16p12.2 deletion encompassing *OTOA*, *METTL9* and *IGSF6* genes, microdeletions in this region has been previously reported in individuals with autistic features, and was proposed to contribute to the development of ASD when other factors are present. We propose that these two deletions are contributing synergistically to the patient's phenotype, consistent with the proposed oligogenic model of ASD where multiple genetic factors interact to lead to autistic features.

1517T

Mouse models of mitochondrial dysfunction display autism endophenotypes. *T. Yardeni, M. Patel, M. McManus, D.G. Murdock, D.C. Wallace.* Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia, Philadelphia, PA.

Autism spectrum disorder (ASD) is characterized by impaired social interaction and communication, repetitive behavior, and restricted interests. Numerous studies have reported mild mitochondrial defects in ASD patients, including reduced mitochondrial enzyme levels and respiration in muscle, brain autopsies, and lymphoblastoid cell lines. It has been concluded, based on a meta-analysis, that children with ASD have a spectrum of differing severities of mitochondrial dysfunction. In addition, ASD has a strong male bias (4:1) which is inconsistent with Mendelian expectations but is characteristic of milder mtDNA diseases such as Leber Hereditary Optic Neuropathy (LHON). Based on the above facts, we have hypothesized that partial mitochondrial defects in mice due to nuclear DNA (nDNA), mitochondrial DNA (mtDNA), or environmental insults could result in autistic behavioral phenotypes. To test this hypothesis, we exposed our unique mouse models of mitochondrial dysfunction to a battery of behavioral tests that can reveal ASD endophenotypes. Our results on three month old mice harboring the mtDNA *ND6* nt 13997G>A P25L mutation with or without a null mutation in the nDNA nicotinamide nucleotide transhydrogenase gene *(Nnt)* revealed autism endophenotypes. These included repetitive activity (displayed as increased marble burying) and decrease in social interaction monitored by three-chamber social approach test. These results show for the first time that mutations that cause partial mitochondrial dysfunction are sufficient in themselves to cause autism-associated phenotypes. Since mitochondrial function requires one to two thousand nDNA genes plus hundreds of copies of the mtDNA, inactivation of one allele of a nDNA bioenergetic gene, functionally different mtDNA lineages, mildly deleterious mtDNA mutations, and/or mitochondrial toxins could all contribute to the complex etiology of autism.

1518F

Lack of association between *COL1A1* **polymorphisms and otosclerosis in Mexican patients.** *M.L. Arenas-Sordo[1], L. Casas-Avila[1], M. Valdes-Flores[1], G. Cordero-Olmos[2], C.E. Urquijo-Torres[1], L. Hernandez-Hernandez[3].* 1) Genetics, Instituto Nacional de Rehabilitacion, Mexico City, Mexico City, Mexico; 2) Othorinolaringology; 3) Audiology.

   Introduction. Otosclerosis is an illness that affects the optic capsule, caused by changes in bone metabolism. It is one of the most common causes of conductive hearing loss in adults and tends to be progressive in nature. It is the cause of 5–9% of the cases with hearing loss and of 18–22% of conductive hearing loss. From a genetic point of view, otosclerosis is usually considered as a multifactorial disease. Despite many studies over the time, a lot of etiological factors and theories have been suggested but the process of development of the otosclerosis remains unclear. *COL1A1* was the first gene to be associated with otosclerosis. This gene shared clinical and histopathological findings between otosclerosis and osteogenesis imperfect and is also related with osteoporosis. Some studies have demonstrated there are association between some polymorphisms and otosclerosis: rs2857396, rs16970089, rs113237935, rs9898186, rs 1800012, rs 2586498, rs2586488, rs1061237 (Americans), rs86498 y rs1061237 (Spanish), rs16970089, rs11327935, rs9898186, rs1800012, rs2586498, rs2857396, rs2586488, rs1061237 (Germans), rs16970089, rs11327935, rs9898186, rs1800012, rs2586498, rs2075555, rs2141279, rs2857396, rs2075559 (Belgium), and many others. The aim of this study was to determine if there is association between otosclerosis and the polymorphisms rs1107946, rs2586498, rs2141279 and rs1800012 of *COL1A1*. **Methods**.A case-control study was done with 90 cases and 135 controls. All patients and controls have complete medical records and audiological studies: otoscopy, audiometry, and signed the informant consent. The molecular study was done with RT-PCR **Results**: The 56% of the patients have family history of otosclerosis. From the 90 patients 70% were female and 30% male and in the controls 76% and 24% respectively. The mean age was 47.63 ± 9.8 years (23-65) in cases and 41.27 ± 12.05 years (24-75) in controls. The disease was bilateral in 82 cases (91%) and unilateral in 8 (9%), more frequent left side involved. The disease started before 40 years of age in 85% of the cases and the evolution mean time of evolution was 14.12 ± 9.7 years. None of the polymorphisms showed association with otosclerosis in Mexican patients. **Conclusions.** We did not found any association. This could be because we are different from other populations and/or the sample size was not big enough. Also one polymorphism (rs214179) was not in Hardy Weinberg equilibrium.

1519W

The first GWAS for tinnitus indicates a very low heritability for this trait. *E. Fransen[1,2], A. Gilles[3], P. Van de Heyning[3], G. Van Camp[1].* 1) Center for Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) StatUa Center for Statistics, University of Antwerp, Antwerp, Belgium; 3) ENT department, Antwerp University Hospital, Edegem, Belgium.

   Tinnitus, the perception of an auditory phantom sound in the form of ringing, buzzing, roaring or hissing in the absence of an external sound source, is perceived by approximately 15% of the population of which 2.5% experiences a severely bothersome tinnitus. The cause of tinnitus may be otologic (hearing loss, noise trauma, Ménière's disease, acoustic neurinoma, ototoxic medications or substances), neurologic (multiple sclerosis, head injury), metabolic (thyroid disorder, hyperlipidemia, vitamin B12 deficiency) or psychogenic (depression, anxiety, fibromyalgia). However, despite thorough clinical examination, tinnitus etiology often remains unclear. We performed a genome-wide association study on 916 independent individuals, of whom 18% reported episodes of tinnitus lasting for longer than 5 minutes. After imputation and quality checking, we tested the association between the tinnitus phenotype and 4,000,000 SNPs using PLINK (v1.07) assuming an additive model. None of the SNPs reached the threshold for genome-wide significance ($p < 5E-8$). The most significant SNPs, which are situated outside coding genes, reach a p-value of 3.4E-7. Using the Genetic Analysis of Complex Traits (GACT) software, we estimated the percentage of the variance explained by all SNPs in the GWAS. This showed a very low heritability of tinnitus (3.2%).

1520T

Linkage analysis of hearing loss severity identifies a potential modulator of *TMTC2*, a novel dominant mutation for familial sensorineural hearing loss. *J. Kent[1], C. Erbe[2], W.-M. Kwok[3], H. Guillen Ahlers[1], C.L. Runge[2], M. Olivier[1].* 1) Dept Genetics, Tx Biomed Res Inst, San Antonio, TX; 2) Dept Otolaryngology & Communication Sciences, Medical College of Wisconsin, Milwaukee, WI; 3) Dept Anesthesiology, Medical College of Wisconsin, Milwaukee, WI.

Familial sensorineural hearing loss (SNHL) is a complex of inherited pathologies of cochlear structure or the auditory nerve. While many SNHL genes have been identified, the full spectrum of genetic risk is not known. In a six-generation family of Northern European descent (N=46), we have recently identified a SNP, rs35725509, in *TMTC2* (chr12q21) as a novel autosomal dominant mutation sufficient to cause early-onset, moderate to severe SNHL in this family, and have confirmed association of this variant with SNHL in the general population. However, two branches of the family (first and second cousins in generation IV and their offspring) exhibit variation in severity and rate of progression of SNHL, suggesting the possibility of genetic modifier(s) segregating in the family. Using a subset of family members (N=36) with audiographic data and/or confirmed normal hearing, heritability of affection status was 1.0 (p=1e-07) using a probit model in SOLAR. We performed two-point genome-wide linkage scans in SOLAR using a panel of 79,163 Affymetrix SNP genotypes. We replicated our original primary linkage (LOD=2.9) of SNHL affection status to chr12q21, and this QTL accounted for all of the heritability for SNHL status. However, for a severity measure, audiographic threshold pure tone average (PTA, dB), heritability was 0.67+/-0.38 (p=0.03) and the peak linkage was at 5q13.2-q13.3 (LOD=9.3). The significance of the chr5q13 QTL as a potential modifier of SNHL severity needs further study. This locus has been reported as an autosomal recessive QTL for SNHL in a Qatari family. Genes in our linkage region of potential relevance include *MAP1B*, which encodes a microtubule-associated protein required for neurogenesis, and *S100Z*, which encodes a calcium binding protein. S100 family proteins are known to modulate calcium and potassium ion channels. TMTC2 is associated with the endoplasmic reticulum calcium uptake pump SERCA2B (ATP2A2), also on chr12. Other genes at chr5q13 may be relevant to the function of the TMTC2/SERCA2B complex, and normal auditory function.

1521F

Whole exome sequencing (WES) analysis in a Caucasian family with keratoconus. *Y. Bykhovskaya[1], E.L. Crowgey[2], Y. Liu[3], X. Li[4], Y.S. Rabinowitz[1].* 1) Cedars-Sinai Medical Center, Los Angeles, CA; 2) Nemours Alfred I. Dupont Hospital for Children, Wilmington, DE; 3) The Medical College of Georgia, Georgia Regents University, Augusta, GA; 4) Institute for Translational Genomics and Population Sciences, Biomedical Research Institute, Harbor-UCLA Medical Center, Los Angeles CA.

Genome-wide linkage analysis in a large four generation Caucasian pedigree with non-syndromic keratoconus mapped the responsible gene to a novel 8.2 MB genomic region located at 5q14.3-q21.1 with LOD of 3.53 which was further narrowed down to 5MB (96 Mb-100 Mb) using high density SNP panel with LOD of 2.49. We also identified SNPs associated with the trait under dominant model with p values ranging from 0.05 to 0.004. To identify keratoconus susceptibility variant (s), we performed whole exome sequencing in nine family members representing three family trios. After appropriate level of quality of the sequenced reads was confirmed by fastqc, a platform independent NGS quality tool, sequences were aligned to the human reference genome hg19 using bwa-mem (version 0.7.4). Following GATK best practices, Bam files were processed using Picard Tools (version 1.67) and GATK Haplotyper (version 3.4-0) for SNP and small indel detection under the default parameters. Variant call file (VCF version 4.1) was then annotated with SnpEff (Version 3.3a). In total, more than 500K variants were identified. No variants so far have been identified in top candidate gene in the region *CAST* which was implicated in genetic susceptibility to keratoconus in independent family and case-control panels. However, we have identified a combination of one common and one rare variant in the adjacent functionally related genes also located in the linkage region as a potential cause of keratoconus in this family. We are also performing comprehensive analysis of all sequencing data using multiple bioinformatic tools including variant prioritization using biological evidence and family segregation analysis using Phevor and CNV analysis using Conifer (version 0.2.2). We are also performing linkage analysis of all WES variants to identify additional evidence of linkage on chromosome 5 and/or other genomic regions. Identification of the variant(s) causing non-syndromic keratoconus in this family will provide significant new information on the genetic susceptibility to this ocular disease with complex genetic inheritance.

1522W

Identification of loci associated with response to intraocular pressure treatments in the Genetic Epidemiology Research on Adult Health and Aging cohort. *H. Choquet[1], K.K. Thai[1], R. Melles[2], E. Jorgenson[1]. 1) Kaiser Permanente, Division of Research, Oakland, CA; 2) Kaiser Permanente, Department of Ophthalmology, Redwood City, CA.*

Purpose: Elevated intraocular pressure (IOP) is the most important modifiable risk factor for glaucoma, a common eye condition which can lead to vision loss. Topical IOP lowering medications are used to control elevated IOP levels, however, there is variability among patients in response to these treatments, suggesting the influence of modifying factors. Here, we examine whether genetic factors may explain some of the variability in treatment response. **Methods:** We conducted a genome-wide association study (GWAS) of response to IOP lowering medications in the Kaiser Permanente Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort. As part of standard care, subjects had IOP measurements and were treated with IOP lowering medications for elevated IOP. We identified 4,037 patients of European ancestry with prescriptions for alpha-adrenergic agonists, beta-blockers, carbonic anhydrase inhibitors, Cosopt, and prostaglandin analogs. We selected IOP measurements in subjects' right eyes pre-treatment (up to 60 days prior to first prescription) and post-treatment (up to 120 days after). Response to IOP treatment was defined as the difference in IOP between the two measurements and as the percentage change in IOP measurement. Over 6.6 million genetic markers were imputed with $r^2>0.8$ and minor allele frequency > 0.05 using the 1000 Genomes reference panel. We first examined association between genetic markers and response to treatment, in the five treatment groups separately, using linear regression. A meta-analysis was then conducted by combining all the treatment groups. **Results:** We identified 12 loci that reached a suggestive level of association ($p<10^{-6}$) in individual treatment group analyses. No SNP associations reached genome-wide significance. By combining the five treatments in a meta-analysis, we observed a suggestive association ($p=4.12*10^{-7}$) between rs4876582 on chromosome 8 and response to IOP treatment. **Conclusions:** While we did not identify a genome-wide significant association with response to IOP treatment, we would expect 5 suggestive loci and we observed 12. Our results are consistent with a role for genetic factors underlying some of the variability among patients with elevated IOP. Despite this being the largest study to date of the genetics of IOP treatment response, additional analyses are needed to further refine these findings. Nonetheless, this is an important first step toward the pharmacogenetics of IOP treatment response.

1523T

Age-related macular degeneration patients carrying rare variants in the *CFH* and *CFI* genes have a reduced ability to degrade C3b. *M.J. Geerlings[1], M. Kremlitzka[2], B. Bakker[1], S.C. Nilsson[2], N.T. Saksens[1], Y.T. Lechanteur[1], M. Pauper[1], J. Corominas[1], S. Fauser[3], C.B. Hoyng[1], A.M. Blom[2], E.K. de Jong[1], A.I. den Hollander[1,4]. 1) Department of Ophthalmology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; 2) Division of Medical protein chemistry, Department of Translational Medicine, Lund University, Malmö, Sweden; 3) Department of Ophthalmology, University Hospital of Cologne, Cologne, Germany; 4) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands.*

BACKGROUND Tight regulation of the complement system is needed to protect the cells of our body from tissue damage. In age-related macular degeneration (AMD) the complement system is highly burdened by genetic variation. Rare variants in complement factor H (*CFH*), complement factor I (*CFI*), component 9 (*C9*) and component 3 (*C3*) have been described to play an important role in AMD pathogenesis. It has been postulated that patients carrying these variations may have higher complement activation levels than AMD patients in general, but this has not yet been experimentally proven. Therefore, we aimed to determine the effect of rare genetic variants on levels of complement components in serum. In addition, we compared the activity of the complement system in rare variant carriers compared to non-carriers to evaluate if carriers are more likely to benefit from complement inhibiting therapy. **METHODS** Whole exome sequencing data of 22 severely affected AMD families was filtered for rare variants in complement genes *CFH*, *CFI*, *C9*, and *C3*. In addition, 3200 individuals from the European Genetic Database (EUGENDA) were genotyped using KASP Genotyping System (LGC). Serum samples were obtained from carriers of identified rare variants (n=177) and a non-carrier control set (n=157). The concentrations of Factor H, Factor I, C9 and C3 in serum samples were measured by ELISA. Additionally, serum samples were incubated with C3b-AFG47 and separated on 10% SDS-PAGE. The ability to degrade C3b was calculated as the ratio between the α'chain and the 43-kDa product. **RESULTS** Novel and previously described rare variants were predominantly found in cases and contributed to an increased risk to develop AMD (OR 2.1; p<0.001). *CFI* carriers (Gly119Arg; Leu131Arg) showed decreased factor I serum levels. Carriers of the *C9* Pro167Ser variant had increased C9 serum levels compared to non-carriers. Factor H and C3 serum levels did not differ between carriers and non-carriers. Carriers of rare variants in *CFH* (Ser193Leu; Arg175Gln) and *CFI* (Gly119Arg; Pro553Ser; Leu131Arg) genes had a reduced ability to degrade C3b compared to non-carriers. Additionally, we detected a correlation between both FH and FI levels with the ability to degrade C3b in non-carriers individuals (R^2 0.35 and R^2 0.31, respectively; p<0.001). **CONCLUSION** Carriers of rare variants in *CFH* and *CFI* have higher levels of complement activation and may benefit more from complement inhibiting therapy than AMD patients in general.

1524F

A cumulative genetic risk score is associated with younger age of diagnosis in primary open angle glaucoma. *J. Wiggs[1], R. Guo[1], J.N. Cooke Bailey[2], R.P. Igo, Jr.[2], J.L. Haines[2], L.R. Pasquale[1], NEIGHBORHOOD consortium.* 1) Dept Ophthalmology, Harvard Med Sch, MEEI, Boston, MA; 2) Dept of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio.

   Purpose: To determine if dosage of 12 primary open-angle glaucoma (POAG) genetic risk variants are associated with glaucoma features related to disease severity: age of diagnosis, disease severity based on visual field mean defect (MD) , and need for glaucoma filtration surgery. **Design:** Retrospective observational study with case-control and case-only components **Participants:** 976 POAG cases and 1140 controls from the Glaucoma Genes and Environment Study (GLAUGEN) and 1971 cases and 2347 controls from the National Eye Institute Glaucoma Human Genetics Collaboration (NEIGHBOR). **Methods:** We aligned the variant associated with increasing risk of POAG for 12 SNPs known to be associated with POAG in our case control sets. In case only analyses we formulated a weighted genetic risk score (GRS) from this panel and assessed the association with glaucoma features of interest. Multivariable results from cohort specific analyses were combined with meta-analytical techniques. **Main outcome measures:** Age of disease onset, Mean deviation (MD) on visual field testing (Humphrey) and need for incisional glaucoma surgery. **Results:** In case control analyses 10 of 12 SNPs had at least nominal association with POAG (p<0.05) and the GRS was strongly associated with POAG (P=5.1E-71). In case only analysis, the GRS ranged from 7 to 21 (maximum score was 24). In multivariable analysis, each unit higher in GRS was associated with a 0.26 year earlier age at diagnosis (95% confidence interval: -043, -0.08; p=0.005). These results were most significantly driven by GAS7 rs9897123 (beta = -0.72; 95% CI: -1.27, -0.17; p=0.01), FOXC1 rs2745572 (beta = -0.67; 95% CI: -1.29, -0.04; p=0.036) and CDKN2B-AS1 rs4977756 (beta = -0.69; 95% CI: -1.28, -0.10; p=0.021). Individuals with >19 risk alleles had an average age of diagnosis 4 years earlier than those with <9 risk alleles. The GRS was not related to maximum MD from the eye with more functional loss (beta = -0.019; 95% CI: -0.15, 0.11; p=0.77); nor was it associated with the need for filtration surgery in GLAUGEN where this information was available (OR=1.02; 95% CI: 0.93, 1.11; p=0.71). **Conclusion**: A higher dose of POAG risk alleles was associated with an earlier age of glaucoma diagnosis. On average, POAG patients with the highest GRS had 4.0 year earlier diagnosis of disease. Discovery of more POAG loci is needed to fully predict glaucoma features from genetic data. Funded by NIH/NEI R01EY022305.

1525W

Family-based rare variant association study of familial myopia in Amish and Ashkenazi Jewish families. *D. Lewis[1], C.L. Simpson[1,2], A.M. Musolf[1], L. Portas[1], F. Murgia[1], D. Stambolian[3], J.E. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore Maryland, United States; 2) University of Tennessee Health Sciences Center, Memphis, Tennessee, United States; 3) Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States.

   Myopia is the largest cause of uncorrected visual impairments globally and constitutes a major public health problem that is dramatically increasing in many populations. Its complex multifactorial component is governed by both genetic and environmental factors that promote its development. Genome-wide association studies (GWAS) and linkage studies have identified multiple loci influencing the risk of developing myopia but, few causal variants have been identified. Compared to common variants, rare variants are more likely to be mutations of recent origin and therefore more likely to be population specific. Therefore, because of population history and differences in linkage disequilibrium (LD) patterns, it can potentially lead to spurious associations. The case-parent trio design has an advantage of dealing with confounding due to population structure by employing a family-based association test such as the transmission disequilibrium test (TDT). We have performed family based association analysis examining common and rare variants using the family-based association test (FBAT) software package on 37 Amish and 63 Ashkenazi Jewish families with strong family history of myopia from the Penn Family Study using genotypes from the Illumina HumanExome v1.1 array plus 24,263 custom SNPs. Individuals in the families were defined as myopic if their average refractive error was <= -1 Diopter (D) and were considered unaffected if their average refractive error was > 0.0 D; others were coded as having an unknown phenotype. More stringent rules were used to code children as unaffected since children's eyes become more myopic fairly rapidly during childhood and adolescence. Noteworthy are two variants (rs135 and rs136) for Amish families and one variant (rs6972578) for Ashkenazi Jewish families, all in the same gene (OSBPL3) that were suggestively associated (*p* < 1.4 x10[-4]) with common myopia under a significant linkage peak previously detected in a set of African American families from the Penn Family Study. To follow-up this observation of association of different variants in the same gene in two different samples, we are using rare-variant (RV) association analysis in the TDT framework to perform gene-based tests with the rare variants in these exome chip data. The RV-TDT framework can control for both admixture and substructure and thus avoid spurious associations. This method has the potential to improve our power to detect causal genetic variants.

1526T

***RRAGC* is associated with age-related macular degeneration (AMD) progression rate from intermediate to advanced AMD, implicating mTORC1 signaling in AMD pathogenesis.** *P.J. Persad[1], R.J. Sardell[1], S.S. Pan[1], P. Whitehead[1], L.D. Adams[1], R.A. Laux[2], J.A. Fortun[3], M.A. Brantley Jr.[4], J.L. Kovach[3], S.G. Schwartz[3], A. Agarwal[4], J.L. Haines[2], M.A. Pericak-Vance[1], W.K. Scott[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami , FL; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 4) Ophthalmology and Visual Sciences, Vanderbilt University Medical Center, Nashville, TN.

Age-related macular degeneration (AMD) is a complex disease for which 52 risk variants across 34 loci have been identified. However, only some of these loci (*CFH, ARMS2, C2/CFB, C3, LIPC, ABCA1, COL8A1,* and *RAD51B*) have been associated with progression rate to advanced AMD. Identifying genes influencing progression could provide novel targets for therapeutics aimed at slowing disease pathogenesis and preserving visual acuity. Genome-wide association studies to identify novel loci influencing progression from intermediate to advanced AMD are lacking. Therefore, to identify progression loci, we examined a dataset of 397 European-descended subjects (606 eyes with intermediate AMD at baseline of which 178 eyes progressed) with at least two examinations and a mean follow-up time of 2.5 years (range 0.05-13). The mean age at baseline was 76.5 years, and 58% of subjects were female. Genome-wide genotyping was performed using the Illumina HumanCoreExome array and the dataset was imputed to 6.3 million variants (with imputation quality (R^2)>0.3 and minor allele frequency$\geq$0.05) using IMPUTE2 and the 1000 Genomes Project Phase I reference panel. Cox proportional hazards regression models were used to assess association between time to progression in each eye from intermediate AMD (extensive intermediate-size drusen or any large drusen) to advanced AMD (geographic atrophy and/or choroidal neovascularization) and imputed SNP dosage (with non-progressors censored at the time of last examination). Models included the top two principal components from population stratification analyses, sex, and age at baseline, and accounted for correlation between an individual's eyes. Thirteen SNPs in the *RRAGC* gene had genome-wide significant main effects [minimum p-value (rs188028731)=1.15×10^{-9}, hazard ratio (HR)=3.12, 95% CI=2.16-4.51]. These SNPs were associated with a shorter time to progression. Two other intergenic regions on chromosomes 3 and 12 produced genome-wide significant results. The 52 known AMD risk variants were not nominally statistically significant (p<0.01). *RRAGC* regulates mTORC1 signaling. This pathway influences neovascularization in AMD and is under consideration as a target for drug development. Thus, identification of *RRAGC* as a new progression locus may facilitate development of strategies for slowing progression rate from intermediate AMD to advanced AMD, slowing vision loss, and improving management of the disease in patients.

1527F

Genetic screening of the homozygous p.V37I variant in *GJB2* may facilitate early detection of postnatal childhood hearing loss. *T. Yang, Y. Chen, H. Wu.* Dept of Otolaryngology - Head And Neck Surgery, Xinhua Hospital, Shanghai, Shanghai, China.

Purpose: Widespread genetic screening of the homozygous p.V37I variant in *GJB2* may provide a new strategy for early detection of the postnatal childhood hearing loss. The effectiveness of this strategy, however, has not been evaluated in practice. Our study aimed to identify children with postnatal childhood hearing loss due to the homozygous p.V37I variant in a large Chinese Han cohort. **Methods:** Children with homozygous p.V37I were identified from 28378 Chinese Han children born in 2006 by high resolution melting analysis. Their hearing status was evaluated and compared with that of 200 randomly selected wild-type children at the age of eight. Postnatal childhood hearing loss was diagnosed with reviewing the newborn hearing screening result. **Results:** The homozygous p.V37I variant was detected in 126 (0.44%) of the 28378 children. Among them, 18 (14%) had sensorineural mild-to-moderate hearing loss at the age of eight including 8 (6.3%) with postnatal childhood hearing loss who passed their newborn hearing screening at birth. The homozygous p.V37I variant was associated with an increased risk of postnatal childhood hearing loss (RR, 27.1; 95% CI, 1.6-465.2; *P*=0.023). Most (7/8) children with postnatal hearing loss did not have any risk-indicating features listed in the Position Statement 2007 from the Joint Committee on Infant Hearing. **Conclusion:** Widespread newborn screening for the homozygous p.V37I variant in *GJB2* may reveal the genetic predisposition to congenital and postnatal childhood hearing loss in 0.44% of Chinese Han general population. With continued monitoring of the children's hearing, early detection of the postnatal childhood hearing loss can be achieved in approximately 6.3% of this specific susceptible group.

1528W

Genome-wide association study of Fuchs endothelial corneal dystrophy in a Japanese population. *M. Nakano[1], H. Adachi[1], Y. Tokuda[1], H. Nakagawa[2], Y. Ikeda[2], M. Ueno[2], C. Sotozono[2], S. Kinoshita[2,3], K. Tashiro[1].* 1) Dept Genomic Med Sci, Kyoto Prefectural Univ Med, Kyoto, Japan; 2) Dept Ophthal, Kyoto Prefectural Univ Med, Kyoto, Japan; 3) Dept Front Med Sci & Tech Ophthal, Kyoto Prefectural Univ Med, Kyoto, Japan.

Fuchs endothelial corneal dystrophy (FECD, MIM: 136800) is an inherited, progressive and degenerative disease affecting the corneal endothelium. In United States, FECD is regarded as a common disease - ~5% of the population over 40 years old suffer from the disease, whereas the prevalence of FECD is considered to be rarer in Japanese. In Caucasian, an intronic variant (rs613872) in the transcription factor 4 (*TCF4*) has been identified by a genome-wide association study (GWAS) as a significant risk factor of late-onset FECD. Furthermore, the expansion of CTG trinucleotide repeat in the third intron of *TCF4* was found to be associated with FECD patients in Caucasian. We also revealed that the CTG repeat in *TCF4* was significantly expanded in Japanese FECD patients compared to the controls, although the frequency of patients possessing expanded repeats was less than the Caucasian patients. These results suggested that additional undiscovered variants should be involved in the pathogenesis of FECD in Japanese. In this study, we therefore performed a GWAS in an attempt to identify novel susceptible variants in Japanese FECD patients. We earnestly recruited the patients together with the controls with informed consent from Sep. 2010 through Aug. 2015 at the University Hospital of Kyoto Prefectural University of Medicine and the Baptist Eye Clinic in Kyoto. Genomic DNA from the subjects was genotyped on illumina Infinium HumanCoreExome BeadChip with 538,448 variants according to the manufacturer's protocol. After removing samples and variants failed to pass the quality controls, the genotype data derived from 55 FECD patients and 445 controls were applied to a GWAS analyzing 278,032 autosomal variants. The association analyses were performed by PLINK and R software. As the results, we were not able to obtain any significant variants within 1 Mb of the *TCF4* locus. Especially, rs613872 in *TCF4* turned out to be monomorphic in our population, which was supported by the Japanese data based on NCBI dbSNP. By contrast, we succeeded in obtaining some suggestive loci other than *TCF4* locus, which seemed to be associated with the Japanese FECD patients. In particular, a genome-wide significant variant ($P < 5.0 \times 10^{-8}$) was identified at 6q15, and we are currently performing the replication analysis to confirm the result. Overall, these findings would provide the foundation for future studies to reveal the mechanism determining the FECD pathogenesis in different populations.

1529T

Genome-wide association study using the Japonica array for cold medicine-related Stevens-Johnson syndrome with severe ocular complications. *M. Ueta[1,2], H. Sawai[2], R. Shingaki[3], Y. Kawai[4], C. Sotozono[5], K. Kogima[4], M. Nagasaki[4], S. Kinoshita[1], K. Tokunaga[2].* 1) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 3) Life Science Business Department, Healthcare Medical Business Promotion Division, Toshiba Corporation Healthcare Company, Tokyo, Japan; 4) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 5) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

A genome-wide association study (GWAS) for cold medicine-related Stevens-Johnson syndrome (CM-SJS) with severe ocular complications (SOC) was performed in a Japanese population. A recently developed ethnicity-specific array with genome-wide imputation that was based on the whole-genome sequences of 1,070 unrelated Japanese individuals was used. For GWAS, we used samples from 117 Japanese individuals with CM-SJS/TEN with SOC and 691 healthy Japanese volunteers and for validation, we included additional 21 Japanese CM-SJS/TEN with SOC patients and additional 192 healthy Japanese volunteers . Validation analysis with additional samples identified two new susceptibility loci on chromosomes 15 and 16. Moreover, interactive effects between the previously identified risk allele, *HLA-A*02:06*, and the newly detected risk variants were observed. This study confirmed the usefulness of GWAS using the ethnicity-specific array and genome-wide imputation based on large-scale whole-genome sequences. Our findings contribute to the understanding of genetic predisposition to CM-SJS with SOC.

1530F

Exploring the heritability of age-related macular degeneration endophe-notypes in the Ohio Amish. *N.A. Restrepo[1], E.S. Yeunjoo[1], R.J. Sardell[2], M.G. Nittala[3], L.D. Adams[2], R. Laux[1], D. Fuzzell[3], L. Reinhart-Mercer[2], L. Caywood[2], D. Dana[3], W.K. Scott[2], S.R. Sadda[2], D. Stambolian[4], M.A. Peri-cak-Vance[2], J.L. Haines[1].* 1) Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Ophthalmology, Doheny Eye Institute, Los Angeles, CA; 4) Departments of Ophthalmology and Genetics, University of Pennsylvania, Philadelphia, PA.

Studies of the impact of vision impairment and blindness are rapidly accruing, resulting in a clearer picture of how vision loss will burden future economies and quality of life for global citizens. Genetics plays an integral role in the risk and onset of vision impairment and disease. Determining the genetic component to disease risk is the first step to studies of the genetic architecture of vision health. Heritability studies not only define the extent to which genetics influences interindividual variation of a phenotype, but also provide insight into environmental factors driving disease. Currently, 60% of the heritability of age-related macular degeneration (AMD) can be attributed to known common and rare genetic variants. The remaining unexplained heritability may partly reflect challenges defining the complex phenotype; the current classification system uses broad-scale grades that may not adequately represent the underlying biological basis of the disease. We hypothesized that parsing the AMD phenotype into heritable sub-phenotypes, each with a distinct genetic basis, will improve prediction of risk and increase our understanding of the genetic architecture of AMD. We sampled 228 related individuals from Amish families with early/intermediate AMD cases in Ohio. Individuals underwent detailed ophthalmologic exams including Ocular Coherence Tomography imaging to quantify retinal features, fundus imaging, and intraocular pressure. The Amish provide an excellent opportunity to analyze the heritability of complex traits given large nuclear families and relatively homogeneous environmental and genetic backgrounds. Utilizing a modified Twin study design under the ACE model, we calculated the additive genetic heritability for IOP as $h^2=0.88$ (0.84-0.93) and the unique environment at 11.7%. Estimates of CDR put additive heritability at $h^2=0.76$ (0.68-0.85) and unique environment at 14.2%. This study is the first reported estimate of the heritability of IOP and CDR in an Amish population, highlighting heritable, quantitative ocular traits that are both measurable in all individuals. Heritability estimates for IOP and CDR appear inflated compared to published estimates in other populations (0.29-0.74 and 0.56-0.78, respectively). This may be due to a highly shared environment which can inflate estimates. Additionally, heritability estimates are population specific and our results may simply reflect a population with low environmental burden for disease risk.

1531W

Pharmacogenetics of age-related macular degeneration: From a candidate gene to a genome-wide approach. *L. Lorés-Motta[1], F. Van Asten[1], M. Riaz[2], M. Grunin[3], J. Corominas[4], M. Pauper[4], C.B. Hoyng[1], E.K. de Jong[1], P. Mitchell[5], P. Baird[2], I. Chowers[3], R.K. Koenekoop[6], S. Fauser[7], A.I. den Holland-er[1,4], International AMD Genomics Consortium.* 1) Department of Ophthalmology, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, the Netherlands; 2) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Ophthalmology, Department of Surgery, University of Melbourne; 3) Department of Ophthalmology, Hadassah Hebrew University Medical Center, Jerusalem, Israel; 4) Department of Human Genetics, Radboud university medical center, Nijmegen, the Netherlands; 5) Centre for Vision Research, Department of Ophthalmology and Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia; 6) Departments of Pediatric Surgery, Human Genetics and Ophthalmology, McGill University Health Centre, Montreal, QC, Canada; 7) Department of Ophthalmology, University Hospital of Cologne, Cologne, Germany.

The highly variable response to anti-vascular endothelial growth factor (VEGF) drugs in neovascular age-related macular degeneration (nvAMD) patients is, in part, due to genetic predisposition. Several studies have implicated variants in genes associated with AMD and VEGF signaling, however, the results have been inconsistent. As a consequence, the nature of this genetic component needs to be further investigated. We sought to identify new variants influencing response to anti-VEGF therapy in nvAMD using candidate gene and genome-wide approaches. Response was defined as the change in best corrected visual acuity (VA) after the loading dose of three monthly ranibizumab or bevacizumab injections. In our candidate gene approach, we interrogated the role of four single nucleotide polymorphisms (SNPs) located in neuropilin-1 (*NRP1*), a co-receptor for VEGFA. The study cohort consisted of 377 nvAMD patients treated with ranibizumab. Patients carrying the GA or AA genotypes of SNP rs2070296 performed significantly worse than individuals carrying the GG genotype. A cumulative effect of rs2070296 in the *NRP1* gene and rs4576072 located in the *KDR* gene, previously associated with treatment response, was observed. Patients carrying two or more risk alleles performed significantly worse than patients carrying zero or one risk allele. In our hypothesis-free approach, 678 patients of European descent treated with ranibizumab or bevacizumab at five different clinics were genotyped using a custom-modified HumanCoreExome array (Illumina). After imputation and quality control procedures, 6,089,769 variants with a minor allele frequency ≥5% were assessed for association with treatment response per cohort. Linear regression analysis was conducted adjusting for baseline VA, age at first injection and the first two principal components, and the results were combined in a meta-analysis. Five suggestive association signals with a p-value <10-5 were identified. Replication of the top SNPs in a cohort of 900 European descent nvAMD patients is in progress and will be presented. A combined analysis of the discovery and replication cohorts may reveal novel genes influencing nvAMD treatment response. The cumulative effect of the variants identified in this project may be used in prediction models for treatment response to anti-VEGF injections, which in the future may help tailor medical care to individual needs.

1532T

Transcriptome profiling of human keratoconus corneas through RNA sequencing identifies collagen synthesis disruption and downregulation of core elements of TGF-β, Hippo, and Wnt pathways. *J.A. Karolak[1,2], M. Kabza[1,3], M. Rydzanicz[4], M.W. Szczesniak[3], D.M. Nowak[1,2], B. Ginter-Matuszewska[1,2], P. Polakowski[5], R. Ploski[4], J.P. Szaflik[5], M. Gajecka[1,2].* 1) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland; 2) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 3) Department of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University in Poznan, Poznan, Poland; 4) Department of Medical Genetics, Medical University of Warsaw, Warsaw, Poland; 5) Department of Ophthalmology, Medical University of Warsaw, Warsaw, Poland.

Keratoconus (KTCN, OMIM 148300) is a degenerative eye disorder characterized by progressive stromal thinning, resulting in a conical shape of the cornea. Structural changes in the corneal layers cause optical aberrations, leading to a loss of visual function. To understand better the factors contributing to KTCN we performed comprehensive transcriptome profiling of human KTCN corneas for the first time using an RNA-Seq approach. 25 KTCN and 25 non-KTCN corneas were enrolled in this study. After RNA extraction, total RNA libraries were prepared and sequenced. The discovery RNA-Seq analysis (in eight KTCN and eight non-KTCN corneas) was conducted first, after which the replication RNA-Seq experiment was performed on a second set of samples (17 KTCN and 17 non-KTCN corneas). Over 82% of the genes and almost 75% of the transcripts detected as differentially expressed in KTCN and non-KTCN corneas were confirmed in the replication study using another set of samples. We used these differentially expressed genes to generate a network of KTCN-deregulated genes. We found an extensive disruption of collagen synthesis and maturation pathways, as well as downregulation of the core elements of the TGF-β, Hippo, and Wnt signaling pathways influencing corneal organization. This first comprehensive transcriptome profiling of human KTCN corneas points further to a complex etiology of KTCN. Supported by National Science Centre in Poland, Grant 2012/05/E/NZ5/02127.

1533F

Genetic analyses of Chinese patients with congenital cataract. *C. Hao, X. Chen.* Beijing Children's Hospital, Capital Medical University, Beijing, China.

Purpose: Cataract is the leading cause of poor vision and blindness in children. Children with congenital cataract failed to be early diagnosed and treated would led to permanent amblyopia. Pre-symptomatic gene diagnosis is critical to congenital cataracts treatment. Methods: The patients' ophthalmic performance and clinical data were recorded in detail. Genomic DNAs were extracted from the blood samples of 10 clinically diagnosed congenital cataract patients and 100 unaffected subjects. Genomic DNA of the probands were investigated using a Next-Generation Sequencing panel which included 40 cataract related genes. All candidate mutations were verified using Sanger sequencing. To exclude the previously unidentified alleles from polymorphisms, samples from 100 unaffected controls were sequenced for the same regions of variations. Possible functional impact of the mutations were predicted by the online software PolyPhen-2 and SIFT. The secondary structures and hydrophobic properties of the mutant amino acid were analyzed by Protscale software. Results: In all 10 patients, 6 unrelated pedigrees involved with autosomal dominant congenital cataracts was identified. Four cases had no apparent pathologic mutations in any of the genes studied. Three missense mutations in GJA8(Connexin50) have been found, including two novel mutation c.A188G(p.N63S) and c.C761T(p.S254F), and a known mutation c.C565T(p.P189S) which was previously found in a Danish family but firstly identified in Chinese pedigree. Other two novel missense mutations were also found in CRYGD c.C107T(p.A36V) and MIP c.T701C(p.L234P); one nonsense mutation was found in CRYGC c.G393A(p.W131Stop). Conclusions: This information will be useful for gene diagnosis and genetic counseling of congenital cataract in China.

1534W

Association between primary open-angle glaucoma and genetic variants associated with intraocular pressure. *F. Mabuchi[1], N. Mabuchi[1], Y. Sakurada[1], S. Yoneyama[1], K. Kashiwagi[1], Z. Yamagata[2], H. Iijima[1].* 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, Univ Yamanashi, Chuo, Yamanashi, Japan.

Purpose: To assess the association between primary open-angle glaucoma (POAG), including normal tension glaucoma (NTG) and high tension glaucoma (HTG), and the genetic variants associated with intraocular pressure (IOP). **Methods:** Two hundred and twenty five Japanese patients with NTG, 214 patients with HTG, and 181 control subjects without glaucoma were analyzed for 6 genetic variants associated with IOP; rs1052990 (near gene: *CAV2*), rs11656696 (*GAS7*), rs7081455 (*PLXDC2*), rs59072263 (*GLCCI1/ICA1*), rs2472493 (*ABCA1*), and rs58073046 (*ARHGEF12*). The total number of risk alleles for IOP elevation of these genetic variants was calculated for each participant and compared between the control subjects and patients with NTG or HTG. The risk (odds ratio) of POAG for each genetic variant was calculated using logistic regression model and the product of the odds ratios of these genetic variants was compared between the control subjects and patients with NTG or HTG. **Results:** The total number of risk alleles in patients with NTG (6.42 ± 1.46, mean ± standard deviation) and HTG (6.44 ± 1.62) were significantly more (P = 0.021, analysis of variance; P = 0.016 and P = 0.014 respectively, Bonferroni post hoc test) than that (6.06 ± 1.47) observed in the control subjects. Similarly, the product of the odds ratios in patients with NTG (2.51 ± 1.01) and HTG (2.53 ± 1.10) were also significantly larger (P = 0.0020, analysis of variance; P = 0.0022 and P = 0.0017 respectively, Bonferroni post hoc test) than that (2.21 ± 0.81) observed in the control subjects. **Conclusion:** These data support that POAG is a polygenic disorder by additive effects of multiple genetic variants associated with IOP.

1535T

Shared genetics between age-related macular degeneration (AMD) and complex diseases/traits suggest common pathways of AMD with cardiovascular and auto-immune diseases. *B. Weber[1], F. Grassmann[1], C. Kiel[1], M. Gorski[2], K. Stark[2], I. Heid[2], International AMD Genomics Consortium.* 1) Inst Human Gen, Univ Regensburg, Regensburg, Germany; 2) Department of Genetic Epidemiology, Univ Regensburg, Regensburg, Germany.

Purpose: Age-related macular degeneration (AMD) is the leading cause of vision loss with strong environmental and genetic components influencing disease development. With regard to the latter, the International AMD Genomics consortium (IAMDGC) recently reported the association of 52 genetic variants at 34 loci with AMD risk. To evaluate an overlap with the genetic risk of other complex diseases/traits, we calculated their genetic risk scores and assessed their association with AMD. Methods: First, we catalogued previously published, genome-wide significant variations for 62 complex diseases/traits and extracted a single variation per independent locus. Next, we calculated a genetic risk score for each trait/disease by calculating the weighted sum of risk alleles for all late stage AMD cases and controls. We then compared the genetic risk scores between late stage AMD cases and controls to identify significant overlaps between AMD risk and the risk for other diseases/traits. Results: The association of the genetic risk scores was evaluated with AMD risk and significant results were found for 17 of the complex diseases/traits (FDR < 0.01). By restricting the analysis to 1,824 variants initially used to compute the genetic risk scores, we identified 28 novel AMD risk variants representing 20 novel AMD-associated loci so far not recognized by AMD genome-wide association studies. Genes in the novel loci reinforce previous findings which strongly implicate the complement system in AMD pathogenesis. Conclusion: Our findings highlight the fact that unrelated pathologies share correlated genetic risk profiles with AMD and thus may point to common disease pathways.

1536F

Loss of function variants in a new gene increase risk for primary congenital glaucoma with variable expressivity. *T.L. Young[1], S. Tompson[1], B. Thompson[2,3], D. Berner[4], O. Siggs[5], V. Limviphuvadh[6], K. Whisenhunt[1], J. Wiggs[7], J. Craig[5], S. Quaggin[2,3], F. Pasutto[4], International Primary Congenital Glaucoma Genetics Consortium.* 1) Ophthalmology and Visual Sciences, University of Wisconsin, Madison, WI; 2) .Feinberg Cardiovascular Research Institute, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 3) Division of Nephrology/Hypertension, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 4) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; 5) Department of Ophthalmology, Flinders University, Adelaide, Australia; 6) Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore; 7) Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA.

   Purpose: Primary congenital glaucoma (PCG) accounts for 5% of childhood blindness, and results from defects in the aqueous outflow system of the eye leading to high intraocular pressure (IOP), globe expansion and optic neuropathy. *CYP1B1* mutations are responsible for ~30-50% of cases, with some reports suggesting involvement of *LTBP2*, *FOXC1* and *MYOC* gene variants. We aimed to identify novel candidate genes for PCG using exome sequencing in unsolved families. **Methods**: An international multi-center cohort of 189 PCG families was assembled, all of which were Sanger sequenced and mutation negative for the 4 known genes. Exome sequencing was performed on 105 families using various platforms. Common and unrelated variants present in the ExAC public database and 119 internal control exomes, respectively, were filtered out. Sanger sequencing was used to validate candidate disease gene variants and identify further variants in 84 additional PCG families and 787 families with later-onset glaucoma (JOAG/POAG). Using a mouse line containing a doxycycline-inducible knock-out allele, animals lacking a functional allele were generated to model the human genotype. **Results**: Heterozygosity for 1 rare missense and 2 novel nonsense variants in the Angiopoietin-1 gene, *ANGPT1*, was identified in 3 of 189 unrelated PCG families. Heterozygosity for an additional novel nonsense and 2 novel missense variants was also identified in 3 of 787 families with JOAG/POAG. The 3 missense variants were located at evolutionarily well-conserved residues, but were predicted to be relatively tolerated by *in silico* tools. However, the 3 nonsense changes likely represented true loss-of-function (LoF) alleles. Remarkably, the ExAC database contains only a single heterozygous LoF variant in *ANGPT1* from more than 121,000 'normal' alleles, indicating LoF variation is not tolerated for this gene. Although mice hemizygous for *Angpt1* from embryonic day 13.5 did not shown any overt phenotype, conditionally homozygous knock-out mice did show a significantly raised IOP. **Conclusions**: For the first time, we have implicated mutations in *ANGPT1* as a risk factor for PCG and later-onset forms of glaucoma. Our data suggest the requirement for vigilant glaucoma screening in relatives of patients that carry an *ANGPT1* mutation, as they may develop glaucoma in later years.

1537W

Common medical comorbidities have no impact on cognitive or behavioral outcome in children with 22q11.2 deletion syndrome: The deletion trumps all! *D.M. McDonald-McGinn[1,2], T.B. Crowley[1], M. Unolt[1,3], K. Grand[1], D.E. McGinn[1], E. Moss[1], L. Katz[1,2], E. Goldmuntz[1,2], R. Sharkus[1], J.W. Gaynor[1,2], B.S. Emanuel[1,2], E.H. Zackai[1,2].* 1) Dept Clinical Genetics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Perelman School of Medicine of the University of Pennsylvania; 3) La Sapienza University Policlinico Umberto I Rome.

   Background: Congenital heart disease (CHD) is one of the most common features (75%) associated with 22q11.2 deletion syndrome (22q). Neurodevelopmental dysfunction is the most frequent adverse outcome of CHD. Hypocalcemia too is common in 22q (50%) reportedly affecting long-term developmental outcome. Mean Full Scale IQ (MFSIQ) scores in children with 22q fall in the borderline range (70-79) and several behavioral differences are observed including ADHD, anxiety and ASD. Knowledge of medical associations with potential to alter cognitive outcomes/behavioral phenotypes is useful in providing anticipatory care and counseling for both families and caregivers. **Methods:** We retrospectively reviewed our records on 1321 patients with 22q for the presence and severity of CHD, hypocalcemia, a behavioral diagnosis of ADHD/anxiety/ASD and IQ scores. **Results:** 1102/1321 patients (83% of the total cohort) had a complete cardiology evaluation, of whom 366 (33%) had an age appropriate Wechsler neuropsychological evaluation yielding a FSIQ score. Likewise, 265/852 (31%) patients had both calcium data and a FSIQ score on record. Medical comorbidities and presence or absence of ADHD/anxiety/ASD were then compared with FSIQ scores in these groups. MFSIQ for patients with CHD (75.86; n=207) was not statistically different from those without CHD (76.86; n=159). We then stratified our patients with 22q and CHD two dimensionally (anatomically, physiologically) and found no significant difference in FSIQ in those patients with a simple, moderate or complex CHD as compared to those with no CHD or a CHD that was considered insignificant, e.g. VSD closing spontaneously. Similarly, FSIQ scores for those with a history of hypocalcemia (77.09; n=146) was not significantly different from those with no hypocalcemia (77.27; n=119). Furthermore, FSIQ scores for those patients with CHD and hypocalcemia (77.94; n=95) was no different than for those with no CHD nor hypocalcemia (77.88; n=59). Finally, neither presence nor absence of CHD/hypocalcemia affected behavioral outcomes nor did behavioral differences impact FSIQ. **Conclusion:** Common medical comorbidities have no impact on cognitive or behavioral outcomes in children with 22q including those with complex CHDs requiring surgical intervention in early infancy. This data may reflect the benefit of early diagnosis and optimal remediation or the intractable effect of haploinsufficiency of 22q11.2 trumping all other effects on neurocognition.

1538T

Genetics of microtia: Discovery of new genes using exome sequencing. *D.V. Luquetti[1,2], A.E. Timms[2], P.M. Hurtado[3], G.L. Porras[4], P. Ayala[5], E.E. Turner[1,2], J. Rosin[2], H.M. Pachajoa[6], T.C. Cox[1,2], I.M. Zarante[5].* 1) Craniofacial Division, Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Seattle Children's Research Institute, Seattle, WA, USA; 3) Pontificia Universidad Javeriana, Cali, Colombia; 4) Clinica Comfamiliar Risaralda, Pereira, Colombia; 5) Pontificia Universidad Javeriana, Bogota, Colombia; 6) Universidad ICESI, Cali; Colombia.

Microtia is a congenital malformation of the external ear ranging from mild structural abnormalities to complete absence of the ear. As with most birth defects Microtia can be syndromic or non-syndromic. The prevalence of non-syndormic microtia varies widely, in the US it is estimated to be between 1 and 2 per 10,000 births, whereas in Colombia is between 4 and 5 per 10,000 births. We have enrolled 85 patients with non-syndromic microtia from four sites in Colombia following a standardized protocol including physical exam, questionnaire on maternal risk factors and family history, 2D photographs and biological sample collection. We submitted DNA from 25 of these patients and their parents for whole exome sequencing (WES). Here, we present data from the WES on these 25 patients; sequencing results from additional 11 trios are pending and will be added to the presentation at the meeting. Our patient population consists of; 76% males, 21% with bilateral microtia and 70% with unilateral right microtia, 85% presented ear canal atresia, 35% had preauricular tags and 8% had a cleft palate. Abdominal ultrasound was performed in 65% of patients with no significant abnormalities found. When analyzing sequence variation, we found a range of 0-3 *de novo*, 0-5 homozygous and 0-14 compound heterozygotes variants per individual. We did not find recurrence of the same variants or genes. For the 30 *de novo* variants found, 16 seem likely causative based on protein prediction of deleterious effect (Polyphen2>0.85), CADD Phred (>15), sequence conservation in evolution (GERP>3.0), population frequency (<0.0001). Of those, 10 are expressed in branchial arch 2 (BA II-embryonic origin of the external ear) during ear development in the mice. We are currently performing in situ hybridizations to confirm the expression of those genes in BAII and confirm spatial representation. None of the genes found in the *de novo* analysis have been previously related to microtia, the data though suggests that these newly discovered genes are involved in the genetic susceptibility to microtia. For example, EPHA4 and NDST2, two of the genes with pathogenic mutations, directly interact with FGFs and FGFRs; FGF3 and 10 and FGFR2 and 3 are already known to cause syndromic microtia. Our future studies include gene expression in the developing human ear and functional studies of the variants in mice models.

1539F

Microduplication at *GRHL3* gene in an individual with non syndromic cleft lip and palate. *L.A. Ribeiro-Bicudo, M.S. Maranhão, J.T. Arruda, N.A. Bergamo, F. Almas.* Genetics, Federal University of Goias, Goiania, Brazil.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is the most common orofacial birth defect with a wide range prevalence among different populations. Although the exact environmental and genetic risk factors associated with NSCL/P remains unclear, the understanding of the genetic mechanisms involved in this malformation are evolving. Syndromic CL/P is mainly caused by Mendelian disorders such as Van der Woude Syndrome (VWS). VWS, first been described in 1954, is an autosomal dominant developmental disorder in which CL/P is a representative hallmark .Previous linkage analysis identified 1q32-q41 and 1p34 as VWS loci. IRF6, located in VWS1 locus 1q32-q41, was identified as a causative gene that can explain ~70% of VWS cases. Given the similar phenotype between syndromic and nonsyndromic forms of CL/P, it is believed that causative genes identified in syndromic CL/P are promising candidate genes for NSCL/P as well. As such, identification of *IRF6* (Interferon Regulatory Factor 6) in VWS and NSCL/P is a typical example of the candidate gene approach. Recently, *GRHL3* (Grainyhead-like 3), located in VWS2 locus 1p34, was identified as a novel causative gene of VWS. As *IRF6*, *GRHL3* can be another example of candidate gene approach for NSCL/P. In the present work we search for microdeletions/duplications in *IRF6* and *GRHL3* genes, through MLPA analysis using the SALSA MLPA P304 kit, in 80 NSCL/P individuals. This study was approved by the institutional review board of the Institution, and written informed consent was obtained from the participants or their parents.MLPA results showed a duplication at exon 4 of *GRHL3* gene in a non syndromic CLP individual. MLPA results from parents did not show any variation. Mutations in *GRHL3* have been identified in VWS patients without *IRF6* mutations. Therefore, *GRHL3* was recognized as a novel causative gene of VWS and represents a promising candidate gene for NSCL/P as well. Although the association between *GRHL3* and NSCL/P remains unknown, recent report identified significant association between *GRHL*3 and NSCL/P susceptibility. This is the first report on a microduplication at *GRHL3* in a individual with NSCLP. Further analysis are necessary to clarify this finding. .

1540W

Gorlin-Goltz syndrome: Case report. *I. Briceño, J. Martinez, M. Sotelo, C. Montes, C. Martinez, A. Zarate, A. Saavedra, L. Arias.* medicine, Universidad de La Sabana, bogota, bogota, Colombia.

   Abstract The Gorlin-Goltz syndrome is a disease of embryonic development characterized by the appearance of multiple queratoquistics tumors, basal cell carcinomas, dimples palm and / or plantar, Vertebral, syndactyly, among others. Patients usually have high predisposition to cancer. The objective of this report is to present the case of a patient Operation Smile foundation with this peculiar condition. phenotypic characteristics found in relation to diagnostic characters for Gorlin-Goltz evaluated further evidence Comparative Genomic Hybridization in which no known mutations were detected, whereby genetic counseling performing molecular diagnostics he performed more robust techniques such as sequencing exome.

1541T

Clinical aspects associated with orofacial clefts in a Colombian population. *J. Martinez[1], I. Briceno[1], M. Sotelo[1], L. Arias[1], C. Montes[1], C. Martinez[1], A. Zarate[1], A. Saavedra[1], A. Montero[1], M. Zamora[1], A. Collins[2].* 1) Medicine, Universidad de La Sabana, Bogota, bogota, Colombia; 2) Genetics, University of Southampton, UK.

   Abstract Objectives: To present descriptive epidemiology of Orofacial Clefts and to determine the association of syndromic forms with antenatal high-risk conditions, preterm birth, and comorbidities among nested-series of cases. **Methods:** A study of nested-series of cases was conducted. Frequencies of cleft type, associated congenital anomalies, syndromic, non-syndromic and multiple malformation forms, and distribution of Orofacial Clefts according to sex and affected-side were determined. Odds ratios were calculated as measures of association between syndromic forms and antenatal high-risk conditions, preterm birth and comorbidities. A total of three hundred and eleven patients with Orofacial Clefts were assessed in a 12-month period. **Results:** The most frequent type of Orofacial Clefts was cleft lip and palate, this type of cleft was more frequent in males, whereas cleft palate occurred more often in females. The most common cases occurred as non-syndromic forms. Aarskog-Scott syndrome showed the highest frequency amongst syndromic forms. Hypertensive disorders in pregnancy, developmental dysplasia of the hip, central nervous diseases and respiratory failure showed significant statistical associations (p <0.05) with syndromic forms. **Conclusions:** These data provide an epidemiological reference of Orofacial Clefts in Colombia. Novel associations between syndromic forms and clinical variables are determined. In order to investigate causality relationships between these variables further studies must be carried out.

1542F

Causes of death associated with 22q11.2 deletion syndrome. *K. Coleman[1,5], L. Kobrynski[2], M. Oster[2,3], B. Pearce[4], J. Cubells[5,6,7], O. Ousley[6,7], C. Gunter[1,2,5].* 1) Marcus Autism Center, Children's Healthcare of Atlanta, 1920 Briarcliff Road NE, Atlanta, GA. 30329; 2) Emory University School of Medicine Department of Pediatrics, 2015 Uppergate Drive, Atlanta, Ga 30322; 3) Sibley Heart Center, 1405 Clifton Road NE, Atlanta, GA 30322; 4) Emory University Rollins School of Public Health, 1518 Clifton Road NE, Atlanta, Ga 30322; 5) Emory University School of Medicine Department of Human Genetics, 615 Michael Street, Suite 301, Atlanta, GA 30322; 6) Emory University School of Medicine, Department of Psychiatry, 12 Executive Park Dr, NE, 2nd Floor, Atlanta, GA 30329; 7) Emory Autism Center, 1551 Shoup Court, Atlanta, Georgia 30322.

Purpose: The purpose of this study is to describe the causes of death reported in patients with the 22q11.2 deletion syndrome (22qDS). This is a retrospective study of 715 patients from 1995 – March, 2016 maintained securely in the Southeast Regional Phenotypic 22q11DS (SERPh22) database by a consortium of researchers in Atlanta, Georgia. We cross-checked causes of death with two other 22qDS databases: the Sibley Heart Center and Emory Pediatric Immunology.Results: There were 53 deaths in the SERPh22 database; however, after cross-checking the other two databases, we could identify causes of death for an additional 5 patients. The total number of known deaths is 58 (8.1%) out of the 715 patients. These data are consistent with published reports on phenotype and known causes of morbidity/mortality in 22qDS patients. Of the deaths that occurred in patients younger than 18 years, 36% are due to cardiopulmonary arrest; 12.7% are due to sepsis or the flu; 3.6% are due to central apnea; 3.6% are due to intraventricular hemorrhages; and 9% are due to other causes (sudden death at home, oncology process with proliferation of cells in pulmonary artery, post-op complications from pacemaker insertion, withdrawal of treatment, and hospice treatment). As the majority of these data are collected from a pediatric healthcare system there are only 3 deaths known to have occurred in patients older than 18 years of age. Two patients had sudden deaths at home and 1 was an accidental death resulting from impulsive behavior.Conclusions: In this study, the main causes of death in 22qDS patients under 18 years is secondary to cardiopulmonary arrest with or without cardiac surgery, sepsis or infections, central apnea, and intraventricular hemorrhages. The low number of patients over 18 years is due to the population served by a pediatric healthcare system. Because this population includes deaths that occur outside of any healthcare facility, the 8.1% of deaths in this cohort is most likely an under-estimate. Future work will include cross-checking all 715 patients with the Georgia Vital Statistics/Death Registry.

1543W

Novel copy number variant (CNV) and neurodevelopmental profile (NDP) in boy with neurofibromatosis type 1 (NF1 [MIM 162200]). *C. Cappello[1], C. Keen[1], C. Samango-Sprouse[1,2,3].* 1) The Focus Foundation, Crofton, MD; 2) Department of Pediatrics, George Washington University, Washington, DC; 3) Department of Human and Molecular Genetics, Florida International University, Miami, FL.

Introduction: NF1 is an autosomal dominant neurogenetic disorder with a prevalence of 1:4650. It is either inherited or caused by a spontaneous mutation of chromosome 17q11.2 (MIM 613113). Physical manifestations include skinfold freckling, café-au-lait spots and neurofibromas. Cognitive dysfunction is common in NF1, specifically in executive function, attention, language, visual perception and learning. We present a 6-year-old male diagnosed with NF1 who also has a 58kb microdeletion on 3q24, affecting one copy of the *PLOD2* gene (MIM 609220). His NDP is complicated by global apraxia (GA), childhood apraxia of speech (CAS), autism spectrum disorder and sensory processing disorder. No previous literature reports NF1 patients with this microdeletion of 3q24 with CAS and GA. This case expands upon the genotypic and phenotypic profile of NF1 and CNV.**Statement of Purpose:** We investigated the NDP of a child with NF1 complicated by a novel CNV. We explored the impact and association of the CNV on this NDP compared to the common phenotypic profile of NF1.**Methods:** NDP was assessed using the Preschool Language Scale-5 (PLS-5), Wechsler Scales (WPPSI), Child Behavior Checklist (CBCL) and receptive vocabulary (ROWPVT-4), which probe cognitive, motor, language and behavioral domains.**Results:** Receptive vocabulary was within the low-average range (standard score=83). He had deficits in auditory and expressive comprehension (1st percentile) on the PLS-5, supporting expressive and receptive language disorder. Inability to imitate nonverbal movements, sounds, words or large motor movements demonstrated GA and CAS, especially since fluid reasoning on WPPSI-V (composite score=100) was an area of preservation. Sensory processing dysfunction impacted neurobehavioral outcomes, as the CBCL demonstrated deficits in executive dysfunction and social development common in NF1, CAS and GA.**Discussion:** This case presents a unique NDP for NF1. The GA and CAS are novel pieces in this phenotypic profile, which have not been described before in NF1. Further studies regarding the 3q24 microdeletion-NF1 interaction must be conducted to ascertain their link to GA, CAS and other neurodevelopmental sequelae. This report and other published findings suggest that atypical phenotypic presentation of NF1 may warrant further investigation of CNV to understand the complexity of the genotypic-phenotypic presentation.

1544T

Loss-of-function rare variants in *ARHGAP29* **contribute to the etiology of non-syndromic cleft lip and palate.** *A.C. Faria[1], C.P. Savastano[1], L.A. Brito[1], N. Seto-Salvia[2,3], E. Peskett[2], C.M. Musso[1], L. Alvizi[1], S.A.M. Ezquina[1], C. James[2], P. Beales[2], M. Lees[4], G.E. Moore[2], P. Stanier[2], M.R. Passos-Bueno[1], GOSgene.* 1) Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; 2) Genetics and Genomic Medicine, UCL Institute of Child Health, London, UK; 3) Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; 4) Department of Clinical Genetics, Great Ormond Street Children's Hospital, London, UK.

   Background: Non-syndromic cleft lip with or without cleft palate (NSCL/P) represents one of the most common congenital human malformations, affecting about 1 in 700 liveborn children worldwide. Genome-Wide Association Studies (GWAS) have successfully detected common susceptibility alleles associated with NSCL/P, including SNVs at the 1p22 region, in which *ARHGAP29* has emerged as the best candidate gene. *ARHGAP29* re-sequencing studies in NSCL/P patients have identified rare variants, mostly missense, with uncertain involvement in the defect etiology. In this study, screening of *ARHGAP29* was conducted in 188 familial NSCL/P cases (173 families from Brazil and 15 families from the United Kingdom) and 1210 controls (609 from Brazil and 601 from United Kingdom). **Method:** Sequencing data from *ARHGAP29* was obtained from whole-exome sequencing in Illumina HiSeq2000, targeted gene sequencing in Illumina MiSeq and or Sanger sequencing in ABI3730 DNA Analyzer. Overall burden of *ARHGAP29* rare variants in patients and controls was tested using gene-based Sequence Kernel Association Test (SKAT). **Results:** We identified 10 rare variants in *ARHGAP29:* five missense, one inframe deletion, and four loss-of-function (LoF) variants. All missense changes already were described in public databases. Only one missense variant was predicted to be possibly pathogenic by *in silico* programs, but discarded after segregation analysis. The in-frame deletion was predicted to be non-pathogenic by *in silico* tools, so it was also not considered further. The four LoF variants, including three splice site and one stopgain variant, were all predicted to disrupt the protein based on *in silico* analysis. Segregation analysis of the LoF variants performed in two out of the four families were in accordance with an autosomal dominant pattern with incomplete penetrance (K=59%). Significant mutational burden of LoF (SKAT, P=0.0005), but not missense, variants were observed in cases compared to controls, suggesting that only LoF variants contribute to the etiology of NSCL/P. **Conclusion**: Our data demonstrate that rare LoF variants in *ARHGAP29* represent an important mechanism in the etiology of NSCL/P, occurring in 2.1% of familial cases. Genetic testing for this gene might be taken into consideration in genetic counseling of familial cases. Support: Fapesp/CEPID/CNPq/NHS Foundation/UCL/Illumina.

1545F

Novel association between rs1063588, an eQTL for *MRPL53* **(2p13)***,* **and nonsyndromic cleft lip with or without cleft palate.** *C. Masotti[1,2], L.A. Brito[1], A.C. Nica[3], D.Y. Sunaga[1], K.U. Ludwig[4], C. Malcher[1], .S.G. Ferreira[1], G.S Kobayashi[1], C.P. Savastano[1], M. Aguena[1], D. Meyer[1], T. Hünemeier[1], D.F. Bueno[2], N. Alonso[5], D. Franco[6], E. Mangold[4], E. Dermitzakis[3], M.R. Passos-Bueno[1].* 1) Department of Genetics, Institute of Biosciences, University of São Paulo, Sao Paulo, Brazil; 2) Hospital Sírio-Libanês, São Paulo, Brazil; 3) Department of Human Genetics and Development, University of Geneva, Switzerland; 4) Institute of Human Genetics, University of Bonn, Sigmund-Freud-Str.25,D-53127 Bonn, Germany; 5) Divison of Plastic Surgery, Medical School, University of São Paulo, São Paulo, Brazil; 6) Departamento de Cirurgia Plástica, Hospital Clementino Braga Filho, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Brazil.

   Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent complex disorder. Many susceptibility loci have been uncovered by genome-wide association studies (GWAS), most of which are low-effect variants falling within noncoding regions or gene deserts. In this regard, GWAS hits may be implicating regulatory regions, which can be active during craniofacial development. In this study, we investigated the association of expression quantitative trait loci (eQTLs) with NSCL/P, in order to directly examine regulatory variants. We firstly mapped eQTLs using mesenchymal stem cell culture samples derived from *orbicularis oris* muscle (OOMMSC) obtained from NSCL/P individuals. *Orbicular oris* muscle is a compromised tissue in NSCL/P patients, and easily collected during reconstructive surgery. By correlating genome-wide expression levels and genotypes, we mapped 119 *cis*-eQTLs related to 18 genes (P<0.0001; FDR=14%), and 31 were selected for further association analysis with 624 NSCL/P patients and 668 controls from Brazilian population. European, African and Amerincian ancestries were estimated for each individual using ancestry informative markers, in order to account for stratification bias. We found a novel association for the *MRPL53* gene best eQTL (rs1063588, P=0.0008). To investigate if a particular ethnic subgroup is driving this association, we stratified the analysis by individual ancestry, and found a stronger association signal for the subgroup enriched with Amerindian ancestry (264 patients x 165 controls with >20% of Amerindian genome: *P=0.008*, $OR_{het\_rs1063588}$ = 1.99 [1.09-3.64 95%CI], $OR_{hom\_rs1063588}$ = 2.77 [1.49-5.15]), compared to the non-Amerindian subgroup (*P*= 0.21; 356 patients x 510 controls with <20% of Amerindian genome), indicating that this locus may be a susceptibility factor in the Amerindian population. *MRPL53* (2p13) encodes a 39S protein subunit of mitochondrial ribosomes and interacts with cMYC, evidences suggesting that ribosome dysfunction may have an etiological impact on NSCL/P. FAPESP, CNPq.

1546W

Trans-ethnic genome-wide meta-analysis of over 9,000 individuals reveals multiple associations with earlobe attachment. *E. Orlova[1], M.K. Lee[2], K. Adhikari[3], J. Li[4], C. Gallo[5], G. Poletti[5], L. Schuler-Faccini[6], M. Bortolini[6], S. Canizales-Quinteros[7], F. Rothhammer[8], G. Bedoya[9], R. González-José[10], D.J. Balding[11], E.J. Leslie[2], E. Feingold[1], J.T. Hecht[12], G.L. Wehby[13], L.M. Moreno[14], A. Ding[4], J. Tan[15], Y. Yang[15], L. Jin[4], M.L. Marazita[2], T. Cox[16], S. Wang[4], A. Ruiz-Linares[3], J.R. Shaffer[1], S.M. Weinberg[2].* 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Oral Biology, University of Pittsburgh Pittsburgh, PA; 3) Department of Genetics, Evolution and Environment, University College London, London, United Kingdom; 4) Chinese Academy of Sciences Key Laboratory of Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P.R. China; 5) Laboratorios de Investigacion y Desarrollo, Facultad de Ciencias y Filosofía, Lima, Peru Universidad Peruana Cayetano Heredia; 6) Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 7) Facultad de Medicina, UNAM, Ciudad de México, Mexico; 8) Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile; 9) Laboratorio de Genetica Molecular (GENMOL), Universidad de Antioquia, Medellin, Colombia; 10) Centro Nacional Patagónico (CENPAT), CONICET, Chubut, Argentina; 11) Centre for Systems Genomics, University of Melbourne, Parkville, Victoria, Australia; 12) Department of Pediatrics, McGovern Medical School University of Texas, Houston, Texas; 13) Department of Health Management and Policy, University of Iowa, Iowa City, Iowa; 14) Department of Orthodontics, University of Iowa, Iowa City, Iowa; 15) State Key Laboratory of Genetic Engineering, Fudan University, Shanghai, China; 16) Center for Developmental Biology & Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA.

The first genome-wide significant associations with the trait were shown in 2015 involving SNPs near the genes *EDAR* (2q13) and *SP5* (2q31), along with a suggestive signal at the 6q24 locus. We performed a meta-analysis of GWAS results from three ethnically distinct cohorts totaling 9155 individuals in an effort to uncover additional genetic signals contributing to variation in earlobe attachment and to investigate whether these signals are population-specific. Earlobe attachment data were available on US individuals of European ancestry (n=1791), a sample of admixed individuals from Latin America (n=4715) and a sample of Han Chinese individuals (n=2649). In each cohort, an ordinal tripartite phenotype definition was used to classify earlobes as free, partially attached, or attached. Results from a common set of 5,366,603 imputed SNPs were meta-analyzed using the inverse variance weighted method in METAL. We observed six genome-wide significant associations: rs3827760 (2q13;p=6.65E-10), rs6756973 (2q31;p=1.40E-28), rs12695694 (3q23;p=4.82E-13), rs58122955 (6q24;p=4.43E-14), rs7096127 (10p12.2;p=2.65E-08), and rs1950357 (14q13;p=4.86E-09). The previously associated signals near *EDAR* and *SP5* were each independently replicated in at least one additional cohort. The previously suggestive signal at 6q24 now reached genome-wide significance. Although there was statistical evidence of effect size heterogeneity at the 3q23 locus (driven primarily by the Chinese cohort), the direction of effect for each of the associated alleles was consistent among cohorts. The associated loci contained several genes involved in morphogenesis. To further investigate these regions, expression of genes in syntenic intervals around the GWAS peaks was investigated in second branchial arch tissue from wildtype E10.5 mouse embryos and two mutants with distinct pinna phenotypes. Of note, *Ranbp2* (2q13), *Vta1* (6q24), *Pax9* (14q13), and *Mipol1* (14q13) were highly expressed in wildtype tissue. *Sp5* and *Gad1*, both located at the 2q31 locus, were expressed at relatively low levels in wildtype embryos, but showed evidence of differential expression in the mutants. These findings shed light on the complex genetics underlying variation in ear morphology and highlight the role of several genes in craniofacial morphogenesis.

1547T

Polymorphisms G894T and A-922G of *NOS3* **gene and periconceptional exposure to vascular disruption agents as a risk factor for non-syndromic cleft lip/palate.** *C. Peña Padilla[1], E.L. Mellin Sanchez[1], L. Bobadilla Morales[2,3], A. Corona Rivera[2,3], H.J. Pimentel Gutierrez[2], C. Ortega de la Torre[2], F. Sanchez Zubieta[2], J.R Corona Rivera[1,3].* 1) Genética, Hospital Civil de Guadalajara Dr. Juan I. Menchaca, Guadalajara, Mexico; 2) Unidad de Citogenética, Onco-Hematología, Pediatría, Hospital Civil de Guadalajara Dr. Juan I. Menchaca, Guadalajara, Mexico; 3) CUCS, Universidad de Guadalajara, Guadalajara, Mexico.

Introduction. Non-syndromic cleft lip/palate (NSCLP) is a multifactorial congenital anomaly with an incidence of 1-3/1000 NB. A previous publication propose the association between NSCLP and polymorphysms G894T and A-922G of *NOS3* gene, related to angiogenesis and vascular disruption, although only in exposure of tobacco and lack of folate intake in the first trimester of pregnancy. **Objectives.** Determine if exist an association between the polymorphysm G894T and A-922G of *NOS3* gene and the exposure to vascular disruptor agents (VDA) as a risk factor for NSCLP in newborns in a school hospital from Guadalajara, Mexico. **Methods.** We studied 161 newborns (47 cases with NSCLP and 114 healthy controls) born in our school hospital during the period 2009 - 2014. DNA extraction form peripheral blood and genotyping by Sanger sequencing were performed previous informed consent. Exposure to VDA was determinate by standardized questionnarie to the mother in both groups. Multivariate analysis with binary logistic regression was performed to determinate the association. **Results.** Genotype frequencies for cases and controls respectively, were: a) *NOS3* A-922G: AA (wild-type) 30 (63.8%) vs 68 (59.6%); AG (heterozygous): 14 (29.8%) vs 41 (35.9%); GG (homozygous) 3 (6.4%) vs 5 (4.4%) and b) *NOS3* G894T: GG (wild-type): 31 (65.9%) vs 73 (64%); GT (heterozygous) 14 (29.8%) vs 35 (30%); TT (homozygous) 2 (4.2%) vs 6 (5.3%). These frequencies were in Hardy - Weinberg equilibrium in the control group, there were no differences stastistically significant, and not association. The multivariate analysis of exposure to VDA showed slightly association only with Coffee (OR= 2.01, CI 95%: 1.00 - 4.04). **Conclusion.** We do not found association between NSCLP and presence of polymorphisms studied, neither for the VDA previously reported. The association found between NSCLP and Coffee, suggest a relation Gene - Environment, but additional investigation is required.

1548F

Replication study of GWAS candidate genes for nonsyndromic cleft lip and/or cleft palate in Vietnamese population. *S. Suzuki[1,2], M. Ono[1], T. Niimi[1], M. Yamamoto[6], H. Imura[1], C. Nguyen[8], E.J Leslie[3,4], M.E Cooper[5], N. Natsume[1], M. Nguyen[8], M.L. Marazita[3,5], J.C. Murray[2,7].* 1) Div.res and Tx for Oral and Maxillofac. Congenital anomalies., Aichi-Gakuin University, Nagoya City, Aichi, Japan; 2) Department of Pediatrics, University of Iowa; 3) Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; 5) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 6) Faculty of Psychological and Physical Science, Aichi-Gakuin University, Japan; 7) University of Iowa, Departments of Pediatric Dentistry, Epidemiology and Biology, Iowa City, IA; 8) Odonto Maxillo Facial Hospital, 263-265 Tran Hung Dao Street, District 1, Ho Chi Minh City, Vietnam.

Background: Nonsyndromic cleft lip with or without cleft palate (CL/P) and cleft palate only (CPO) are among the most common human birth defects caused by both genetic and environmental factors and/or their interactions. Genome wide association studies (GWAS) and meta-analyses of CL/P have been identified multiple genetic risk loci including *MAFB, VAX1, PAX7, FOXE1 and ABCA4*. Previous studies of the *DMD* gene also showed significant association with oral clefts that include cleft palate. **Methods:** We used 404 independent case-parent trios from three case groups: nonsyndromic cleft lip only (CLO; 104 case-parent trios), cleft lip and palate (CLP;135 case-parent trios) and cleft palate only (CPO;165 case-parent trios). Participants were genotyped by either TaqMan assays ® or Fluidigm ®. We genotyped 48 SNP markers, which were utilized and found statistically significnat in previous GWAS studies. Transmission disequilibrium tests were performed by the Family based Association Test and PLINK. This study was approved by institutional review boards and informed consents were obtained from the patients. **Results:** Statistically significant associations were observed with SNPs in genes *IRF6, FOXE1 and MAFB*. Regarding CL/P, TDT analysis showed associations with *IRF6* (rs2235371, *p*= 6.80E-06 and rs2013162 *p*= 0.03486) and *MAFB* (rs17820943, *p*= 0.002266, rs13041247 and rs11696257, *p*= 0.001862). Also *FOXE1* (rs894673 and rs3758249, *p*= 0.02) was slightly associated with CL/P. We could not confirm statistical significance between *VAX1* markers and the Vietnamese population that have previously shown an association in the Mongolian population. However, after Bonferroni correction of the above results, only *IRF6* showed significant association. **Conclusions:** Our study replicated previous GWAS findings in *48* markers from 22 genes for the Vietnamese population and identified that *IRF6* may contribute to a part of the cleft lip and/or cleft palate etiology. Our findings of this study also suggest that *MAFB* might be associated with CL/P, similar to our previous study of the Mongolian population. This work was supported by JSPS KAKENHI Grant Number JP24249092, JP26861757.

1549W

Genome-wide analyses reveal subtype-specific genetic effects in orofacial cleft birth defects. *E.J. Leslie[1], J.C. Carlson[2], J.R. Shaffer[2], A. Butali[3], C.J. Buxo[4], K. Christensen[5], F. Deleyiannis[6], J.T. Hecht[7], L. Moreno[3], I.M. Orioli[8], A.R. Vieira[2], G.L. Wehby[3], E. Feingold[2], S.M. Weinberg[1], J.C. Murray[3], T.H. Beaty[9], M.L. Marazita[1].* 1) Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA; 2) University of Pittsburgh, Pittsburgh, PA; 3) University of Iowa, Iowa City, IA; 4) University of Puerto Rico, San Juan, Puerto Rico; 5) University of Southern Denmark, Odense, Denmark; 6) University of Colorado School of Medicine, Denver, CO; 7) University of Texas Health Science Center at Houston, Houston, TX; 8) ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (National Institute of Population Medical Genetics), Rio de Janeiro, Brazil; 9) Johns Hopkins University, Baltimore, MD.

Complex diseases may include multiple subtypes manifesting similar physiological/anatomical features. While such subtypes are not often readily apparent from available clinical information for many complex diseases, orofacial clefts (OFCs) exhibit easily documented phenotypic heterogeneity, and thus are excellent models for studying phenotypic and genetic heterogeneity. OFCs are caused by incomplete fusion of the upper lip and/or palate, resulting in three major subtypes: cleft lip (CL), cleft palate (CP), and CL plus CP (CLP). Combined they represent the most common human craniofacial birth defects. Historically, CL and CLP were hypothesized to share an etiology, while CP was considered a distinct disorder. However, increasing evidence points to a mixture of independent and shared genetic risk factors across all OFC subtypes. We performed a genome-wide meta-analysis in 901 OFC cases, 1700 controls, and 3427 OFC trios from global populations. We considered 28 loci showing evidence of association with CL, CLP, and/or CP (p<1E-05), and compared estimated effect sizes and confidence intervals. Non-overlapping confidence intervals were considered evidence of differential effects between subtypes. We found eight loci where the effects could not be distinguished between CL, CLP, and CP, indicating these loci may be common to all orofacial clefts. Two loci showed evidence of association only with CLP; two were associated only with CP; while none appeared to be specific to CL. Sixteen loci had overlapping effects between CL and CLP, but not with CP. Among these, we found two loci contained several SNPs with differential effects between CL and CLP. At the *IRF6* locus, multiple SNPs showed larger estimated effects for CL than CLP. The top CL SNPs and the top CLP SNPs are not in linkage disequilibrium, suggesting these were independent signals. We confirmed this with reciprocal conditional analyses on the top hits within each subtype. However, each signal showed some evidence of association within the other subtype, suggesting these visible OFC subtypes are not genetically homogeneous. In conclusion, although multiple shared risk factors exist across subtypes, these do not follow traditional divisions of CL+CLP vs. CP. By identifying OFC subtype-specific signals, this study has facilitated delineation of genetically determined OFC subtypes which should allow us to examine more homogenous groups in future studies and to better understand mechanisms leading to OFCs.

1550T

Genome-wide interaction study implicates *VGLL2* **and alcohol exposure in orofacial clefting.** *J.C. Carlson[1], E.J. Leslie[1], J.R. Shaffer[1], K. Christensen[2], F. Deleyiannis[3], J.T. Hecht[4], L.M. Moreno Uribe[5], G.L. Wehby[5], E. Feingold[1], S.M. Weinberg[1], M.L. Marazita[1].* 1) University of Pittsburgh, Pittsburgh, PA; 2) University of Southern Denmark, Odense, Denmark; 3) University of Colorado School of Medicine, Denver, CO; 4) University of Texas Health Science Center at Houston, Houston, TX; 5) University of Iowa, Iowa City, IA.

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is the most common craniofacial birth defect in humans, affecting approximately 1 in 700 births worldwide. NSCL/P has complex etiology including several known genes and environmental factors. However, known genetic risk variants only account for a small fraction of the heritability of NSCL/P. It is commonly suggested that gene-by-environment interactions (GxE) may help explain some of the missing heritability. We separately examined three prenatal exposures – alcohol, smoking, and vitamin use during the three months prior to conception through the end of the first trimester of pregnancy – in 242 NSCL/P cases and 351 controls of European descent from Denmark, Hungary, and the US. Joint tests of genetic and GxE effects were tested genome-wide (5,935,570 SNPs with MAF>0.10) using logistic regression assuming an additive genetic effect and adjusting for five principal components of ancestry. We further interrogated loci with suggestive results from the joint test (p<1E-05) by examining the GxE effects from the same model. Suggestive results (p<1E-05) for the joint test were observed in the 2q24.2 (alcohol), 4q28.3 (alcohol), 6q22.1 (alcohol), 7p14.1 (alcohol and smoking), 8p23.2 (vitamins use), 8q24.21 (alcohol and smoking), 12p13.31 (alcohol and smoking), and 20q13.2 (smoking) loci. From these regions, a GxE effect was seen for rs117083 in the 6q22.1 locus in the alcohol scan (p=9.54E-07, OR=3.86[2.25, 6.62]). This locus contains *VGLL2*, a gene required for craniofacial development in zebrafish. Interestingly, previous studies indicate that loss of vgll2 and alcohol exposure independently result in death of cranial neural crest cells supporting our observation that genetic variants at this locus and alcohol exposure interact to increase risk of orofacial clefting. R01DE016148, X01HG007485.

1551F

Genetic analysis of vascular malformations reveals both germline and somatic mutations. *W.L. Wooderchak-Donahue[1,2], J.H. Tu[3,4], R. Margraf[1], J. Le[1], P. Bayrak-Toydemir[1,2], J.M. Teng[3].* 1) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT; 3) Department of Dermatology, Stanford University, Stanford, CA; 4) Columbia University College of Physicians and Surgeons, New York, NY.

Vascular malformations are localized abnormal clusters of vessels that can affect capillaries, arteries, veins, lymphatics, or any combination of these. Somatic mutations in *PIK3CA* which codes for the catalytic subunit of phosphatidylinositol 3-kinase (PI3K) were recently identified as a major cause of lymphatic and venous malformations. We hypothesized that genetic factors in addition to mutations in *PIK3CA* account for the diverse clinical phenotypes and treatment responses observed. Here, we determined the prevalence of *PIK3CA* mutations and sought to identify additional germline and somatic variants in a cohort of pediatric patients with lymphatic, venous, or mixed venolymphatic malformations. DNA was extracted from formalin fixed paraffin embedded (FFPE) or fresh/frozen tissue and matched blood samples. A custom SureSelect next generation sequencing panel was used to target the coding regions of 451 genes implicated in vascular malformations and the PI3K/AKT/mTOR pathway. Samples were sequenced on a HiSeq2500 instrument using 2x100 paired-end reads. Sequences were aligned to the human genome reference sequence, and PCR duplicates were removed. Local realignment and variant calling were performed using GaTK, and FreeBayes was used to identify somatic variation down to 1%. Of the tissues evaluated, 7 of 11 (63.6%) had a somatic *PIK3CA* mutation ranging from 1-9% variant. Deleterious germline mutations were identified in two cases that did not have a *PIK3CA* mutation. One case with congenital systemic lymphangiectasia had two heterozygous *CCBE1* mutations causing Hennekam Syndrome, and another case with a venolymphatic malformation had two heterozygous *PIEZO1* variants causing a recently recognized lymphatic dysplasia syndrome. Additional novel germline and somatic variants from several genes whose products are upstream or downstream of PI3K in the mTOR pathway (*AKT1, TSC2, RPTOR, PIK3C2G,* and *MTOR*) were identified in cases that had a somatic *PIK3CA* mutation. This is the first report of the coexistence of germline mutations in the mTOR pathway with somatic *PIK3CA* mutations in patients with lymphatic and venolymphatic malformations. These results may provide new mechanistic insights as to how vascular anomalies develop and lead to better treatments. Our results highlight the genetic complexity of vascular anomalies and underscore the importance of genetic testing in these patients, as their mutation profile may influence differences in response to treatment.

1552W

Tetraploid-diploid mosaicism in a patient with pigmentary anomalies of hair and skin: A new dermatologic feature. *J. Schacht[1], E. Farnworth[1], J. Hogue[3], L. Rohena[1,2,4].* 1) Department of Pediatrics, San Antonio Military Medical Center, San Antonio, TX; 2) Department of Pediatrics, Division of Genetics, San Antonio Military Medical Center, San Antonio, TX; 3) Department of Pediatrics, Division of Medical Genetics, Madigan Army Medical Center, JBLM, Fort Lewis, WA; 4) Department of Pediatrics, University of Texas Health Science Center at San Antonio, San Antonio TX.

Tetraploid-diploid mosaicism in humans is exceedingly rare. Phenotypes are variable, but common features include dysmorphic facial features, growth retardation, developmental delays, and intellectual disability. Few cases report dermatologic findings with pigmentary changes. We present a 10 year old male with tetraploid-diploid mosaicism and coexistent hypopigmentation of his hair with hypo and hyperpigmentation of his skin. This case expands the current literature as we are not aware of previous documentation of this unique combination of pigmentary anomalies. Other significant clinical features affecting several organ systems in our patient include a horseshoe kidney with renal cysts on both moieties, bilateral cryptorchidism, pectus excavatum, 13 rib pairs, bilateral pes cavus, scoliosis, delayed gastric emptying, poor growth requiring g-tube placement, Dandy-Walker malformation, tethered cord, epilepsy, intellectual disability, sensorineural hearing loss, high myopia, lagopthalmos, persistent asthma, and hirsutism.

1553T

Psychosocial, behavioural, cognitive and molecular-cytogenetic aspects of Prader-Willi syndrome (PWS): An Egyptian study. *M. Shehab[1], H. Afifi[2], H. El-Bassyouni[3], E. Salah[4].* 1) Human Cytogenetics, National Research Centre, Cairo, Giza, Egypt; 2) Clinical Genetics; 3) Clinical Genetics; 4) Child Health.

Prader-Willi syndrome is a multifaceted complex genetic disorder that is not widely known in Egypt, and there is very evident lack of differentiation between children with PWS and those with other causes of IDandobesity. This study aimed to examine psychological and behavioral profile of a sample of children with different modes of inheritance. We screened all patients5-14yrs old who presented to CGDclinic, NRC, during5years for at least 3 of the following signs: obesity, hypotonia, delayed milestones and hypogonadism.Starting by105patients;we ended by45patients. They were23males and22femaleswith mean age of 8years.All patients were subjected to clinical examination , Pedigree analysis, Nutritional and Bone Mineral Density evaluation, Children WISC , Autism Rating Scale ,Behavior Checklist for ages4-18,Conners' Parent Rating Scale-Revised,Cytogenetic and FISH analysis. All patients presented with hypotonia and had8-12positive criteria. Positive family history was66.7%and positive consanguinity86.7%.Hypogonadism in82.3% (males showed micropenis, hypospadias and undescended testis; females had absent labia minora and clitoris). The majority of PWS patients53.3% had mild to moderate ID, borderline mentality 6.7%, mild ID20%, and severe ID 20%, mild autistic features6.6% and significant autism2.2%.FISH confirmed deletion in30 patients66.6%.No correlation between FISH results and the IQ. All patients presented with hyperphagia, Cognitive rigidity 86.6% and Inattention76.9%.No association between the prevalence of oppositional defiant,cognitive rigidity or psychosocial disorders and the degree of ID or hyperphagia.Oppositional personality trait was found to be positively related with FISH–ve patients. Conner's results confirmed the positive relation between hyperactivity, Anxiety disorder, Social problems and impulsivity subscale results and FISH–ve patients. ADHD severity index showed a significant +ve inclination towards the FISH–ve patients.DSM-IV Inattention score didn't show any stat. significant difference, DSM-IV Hyperactivity- Impulsivity score was related to the FISH –ve patients.The total score was suggesting a strong correlation of ADHD symptoms with the non deletion types of PWS. This study is a cross sectional study that focuses on a specific age range in PWSpatients, follow up studies assessing those patients till adulthood could result in different conclusions as the patients' psychological and behavioral profiles can vary markedly with age.

1554F

Neotenic Complex Syndrome: A rare condition characterized by neoteny and multiple congenital anomalies. *B.A. Peters[1,3], R.F. Walker[2], S. Ciotlos[1], Q. Mao[1], R. Chin[1], N. Barua[1], S. Drmanac[1], M.R. Agarwal[1], R.Y. Zhang[1], Z. Li[3], M. Wu[1], J.S. Liu[1], P. Carnevali[1], R. Drmanac[1].* 1) Department of Research, Complete Genomics, Inc., 2071 Stierlin Court, Mountain View, CA 94043; 2) Retired; 3) 3BGI-Shenzhen, Shenzhen 51803, China.

We describe a novel syndrome in six patients with extreme developmental delay and other congenital anomalies. The defining characteristic of this syndrome is neoteny, such that the patients appear to be much younger than there biological age. In each case there is no history of such phenotypes in the family suggesting either a *de novo* genetic or environmental cause. All are female with failure to thrive as demonstrated by height and weight well below the 5th percentile. The patients range in age from 3-23 yet none have developed an ability to communicate. Whole genome sequence (WGS) analysis of the patient, their parents, all unaffected siblings, and in some cases extended family members was performed on Complete Genomics' nanoarray platform for massively parallel sequencing to an average coverage of ~50X. The genomes of each patient and sibling were also analyzed to ~100X coverage using Complete Genomics' Long Fragment Read (LFR) technology to provide complete variant phasing information and additional mutation verification. The patients in this study all had normal karyotypes and were free of large structural amplifications and deletions as determined by aCGH and WGS read coverage analysis. We identified 45-135 *de novo* mutations (DNMs) in the genomes of the patients and 58-115 in their unaffected siblings. On average, we discovered ~1.2 coding DNMs per patient and ~1.4 per sibling, this difference was not found to be significant and suggests there is nothing unusual about the number of coding DNMs found in the patients. Two DNMs were found in genes previously found with DNMs in the patients of other severe developmental syndromes. In the remaining patients with DNMs in genes not associated with other syndromes, 80% were found in genes intolerant to variation compared with only 36% of the genes with DNMs in the siblings, lending support to some of these DNMs being potentially causative. Interestingly, the genes found to harbor DNMs in the patients were found to be involved in transcription regulation and chromatin modification; the functions of genes containing DNMs in the siblings were more broadly dispersed. Analysis of inherited variants in the patients and their unaffected siblings did not uncover any other potentially causative variants. In addition, there are no obvious environmental factors that could contribute to these patients' syndrome.

1555W

A rare missense mutation in *CHFR* is associated with IMAGe syndrome. *N. Paria[1], H. Hu[2], F. Payne[3], C. Logan[4], L. Bicknell[5], H. O'conner[6], R. Gray[6], R. Gupta[1], J. Huibregtse[6], R. Semple[7], A. Jackson[4], C. Huff[2], P. White[8], M. Zacharin[9], M. Hutchison[10], J. Rios[1,8,11,12].* 1) Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX; 2) Department of Epidemiology, MD Anderson Cancer Center, Houston, TX; 3) The Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 4) Institute for Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 5) Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; 6) Department of Molecular Biosciences, University of Texas at Austin, Austin, TX; 7) The University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Cambridge, UK; 8) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; 9) Division of Medicine, Royal Children's Hospital, Melbourne, Victoria, Australia; 10) Department of Pediatrics, Medical University of South Carolina, Charleston, South Carolina; 11) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX; 12) Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX.

IMAGe syndrome (MIM# 614732) is an extremely rare growth disorder presenting with severe dwarfism, adrenal insufficiency, skeletal dysplasia and genital anomalies. Previous studies identified mutations in the proliferating cell nuclear antigen (PCNA)-binding domain of CDKN1C, a negative regulator of cell proliferation. We performed whole-genome sequencing in four IMAGe patients from three families in whom no *CDKN1C* mutations were identified. pVAAST analysis identified a genome-wide significant association at a single locus harboring the *CHFR* (p=1x10[-7]) and *GOLGA3* (p=3x10[-7]) genes. Rare variants identified in each gene cosegregated on a 1.71 Mb haplotype shared among patients in all three families; however, genome-wide ERSA analysis confirmed no significant relatedness between samples to at least 9 generations. Both variants were screened in three additional unrelated individuals with IMAGe; two carried both variants while the third carried only the *CHFR* variant (CHFR[Q93R]), suggesting *CHFR* is associated with IMAGe syndrome. The CHFR[Q93R] variant is located at a highly conserved residue within the forkhead-associated (FHA) domain of the CHFR protein, an E3 ligase that functions as a cell-cycle checkpoint and regulator of cell proliferation. CHFR and the related RNF8 proteins are the only E3 ligases known to have both FHA and RING finger domains, and both have been shown to regulate poly- and monoubiquitination of PCNA, respectively. Cultured dermal fibroblasts from IMAGe patients displayed proliferation defects, including binucleated cells with disorganized actin, and failed to respond to growth factors in culture, consistent with altered cell cycle regulation. Transciptome profiling identified preferential expression of the mutant allele in patients compared to carrier parents. Furthermore, we screened for the CHFR[Q93R] variant among patients undergoing clinical whole-exome sequencing for whom no genetic diagnosis was identified. Among five patients carrying the CHFR[Q93R] allele, two adolescent/adult males and one infant female were identified with multiple features resembling IMAGe syndrome and a fourth infant female is described with multiple non-specific features (failure to thrive, global developmental delay, some growth abnormalities). Our results support *CHFR* as a new gene for IMAGe syndrome. We hypothesize that the CHFR[Q93R] allele is inherited with an as-yet unidentified trans-acting factor that may affect allele-specific expression. .

1556T

Sequencing the *GRHL3* coding region reveals the first common susceptibility variant and rare truncating mutations for nonsyndromic cleft palate. *A.C. Böhmer[1,2], E. Mangold[1], N. Ishorst[1,2], A. Hoebel[1,2], P. Gültepe[1,2], H. Schuenke[1,2], J. Klamt[1,2], A. Hofmann[1,2], L. Gölz[3], R. Raff[1], P. Tessmann[1,2], S. Nowak[1], H. Reutter[1,2,4], J. Stavusis[5], M. Ishida[6], R. Seselgyte[6], G.E. Moore[6], M.M. Nöthen[1,2], G. Borck[7], K.A. Aldhorae[8], B. Lace[5], P. Stanier[6], M. Knapp[9], K.U. Ludwig[1,2].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; 3) Department of Orthodontics, University of Bonn, Bonn, Germany; 4) Department of Neonatology and Pediatric Intensive Care, Children's Hospital, University of Bonn, Bonn, Germany; 5) Medical Genetics and Mitochondrial Research Group, Latvian Biomedical Research and study Centre, Riga, Latvia; 6) Institute of Child Health, University College London, London, United Kingdom; 7) Institute of Human Genetics, Ulm University, Ulm, Germany; 8) Orthodontic Department, College of Dentistry, Thamar University, Thamar, Yemen; 9) Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany.

Orofacial clefts are common and occur either in isolation or as part of a syndrome. The two frequent forms are nonsyndromic cleft lip with or without cleft palate (nsCL/P) and nonsyndromic cleft palate only (nsCPO). Both are considered multifactorial with a strong genetic background, but with very limited overlap regarding the genetic etiology. The most common syndromic form of orofacial clefting is Van der Woude syndrome (VWS). Recently, ~5% of VWS patients were identified with mutations in the grainy head-like 3 (*GRHL3*) gene. VWS patients have either CL/P or CPO often accompanied with lower lip pits. As lip pits have an incomplete penetrance, VWS can sometimes mimic nonsyndromic clefting. To investigate *GRHL3* in nonsyndromic clefting, we sequenced its coding region in 576 Europeans with nsCL/P and 96 with nsCPO. Most strikingly, nsCPO-affected individuals had a higher minor allele frequency for rs41268753 (0.099) than control subjects (0.049; $P = 1.24 \times 10^{-2}$). This association was replicated in nsCPO/control cohorts from Latvia, Yemen, and the UK ($P_{combined} = 2.63 \times 10^{-5}$; $OR_{allelic} = 2.46$ [95% CI 1.6-3.7]) and reached genome-wide significance in combination with imputed data from a GWAS in nsCPO triads ($P = 2.73 \times 10^{-9}$). Notably, rs41268753 is not associated with nsCL/P ($P = 0.45$). Rs41268753 encodes the highly conserved missense variant p.Thr454Met (c.1361C>T) (GERP = 5.3), which numerous *in silico* prediction programs denote as deleterious. It has a CADD score of 29.6 and increases protein binding capacity *in silico*. Sequencing also revealed four novel truncating *GRHL3* mutations including two that were *de novo* in four families, where all nine individuals harboring mutations had nsCPO. This is important for genetic counseling: given that VWS is rare compared to nsCPO, our data suggest that dominant *GRHL3* mutations are more likely to cause nonsyndromic than syndromic CPO. Thus, with rare dominant mutations and a common risk variant in the coding region, we have identified an important contribution for *GRHL3* in nsCPO.

1557F

Strategies for selecting candidate genes for studying epistasis based on prior knowledge. *S. Ukraintseva, I. Zhbannikov, M. Duan, L. Arbeeva, K. Arbeev, A. Kulminski, I. Akushevich, A. Yashin.* Center for Population Health Aging, Duke University, Durham, NC.

Biological justification of candidate genes used in analyses of gene x gene interaction effects on aging phenotypes and longevity is critically important for obtaining meaningful results. We suggested several approaches to selecting candidate genes for analysis of epistasis, based on current biological knowledge about their functions and involvement in particular pathways, processes and health disorders. Then we selected SNPs in these genes and estimated the effect of pairwise epistasis on longevity and physiological aging-related phenotypes in four longitudinal human datasets (CHS, MESA, FHS, ARIC). Then we compared results of respective analyses, to see which approach lead to a more biologically plausible interpretation. Specifically, we tested the following sets of candidate genes: 1) Hub genes from aging/longevity related pathways that are known to be interacting; 2) Key regulatory genes involved in cell/tissue responses to damage, such as apoptosis, senescence, proliferation, autophagy, ECM clearing, and inflammation, regardless of their involvement in particular pathway; 3) Genes that are highly pleiotropic and influence risks of multiple diseases, according to information placed in NHGRI GWAS catalog; 4) Genes corresponding to pairs of SNPs that showed increase in LD with age in our data. Our analysis shows that candidate genes from groups (1) and (2) are more prone to significant epistatic interactions that influence aging and longevity than members of two other groups. These results support selection of candidate gene directly involved in aging related processes in cell and tissue for analyses of gene x gene interaction effects on phenotypes of physiological aging and longevity.

1558W

Largest ever GWAS of human lifespan. *P.K. Joshi[1], K. Fischer[2], T. Esko[2,3,4], J.F. Wilson[1,5], LifeGen.* 1) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG UK; 2) Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge Center 7, Cambridge, 02242, MA, USA; 4) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Cambridge, 02141, MA; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU UK.

We shall present the results of a GWAS of ~1.4million parent lifespans: 1 million from UK Biobank and 400,000 from the other LifeGen Cohorts.   Lifespan is a trait of interest to everyone. Until recently GWAS have focussed on long-livedness as a binary trait, with *APOE* and perhaps *FOX03* being the only loci that have been robustly replicated (GWAS of longevity in CHARGE consortium confirms *APOE* and *FOXO3* candidacy, Broer *et al.* 2015 *J Gerontol A Biol Sci Med Sci.)*. Focussing on long-livedness has strengths and limitations: biologically it may focus on ageing rather than common disease, whilst statistically study of trait extremes should increase power, for a given sample size. On the other hand long-livedness may miss early life effects and subject recruitment may be difficult and, very importantly, requires specific recruitment of the aged, rather than permitting general population cohorts to take part. In a pilot study(with around one sixth of the present sample count), we recently developed a GWAS approach to lifespan that overcomes these difficulties and discovered two genome wide significant variants, both of which replicated – using parental lifespan information, coupled with Cox models (Joshi PK *et al.* Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nature Commun* 7, 11174, March 2016*).*   We shall present fresh results of the largest ever GWAMA of human lifespan and discuss the methodological extensions to cox modelling that made this tractable in large GWAS and appropriate to meta-analyse. .

1559T

Whole genome sequencing of a healthy aging cohort. *A. Torkamani[1], G.A. Erikson[1], D.L. Bodian[2], M. Rueda[1], B. Molparia[1], E.R. Scott[1], A.A. Scott-Van Zeeland[1], S.E. Topol[1], N.E. Wineinger[1], J.E. Niederhuber[2], E.J. Topol[1].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Inova Translational Medicine Institute, Falls Church, Virginia.

Studies of long-lived individuals have revealed few genetic mechanisms for protection against age-associated disease. This may be due to the fact that, in part, medical technology can promote longevity as a function of medical interventions per se. Yet, heritability estimates have demonstrated that longevity is in fact partially inherited, with a negligible genetic contribution to survival up to approximately 60 years of age, after which an increasing genetic component to survival is observed. Therefore, we pursued whole genome sequencing (WGS) of a related phenotype – healthy aging – to understand the genetics of disease-free aging without medical intervention. WGS was performed with the Complete Genomics platform on 511 healthy aging individuals were compared to 686 matched population controls.   In contrast with studies of exceptional longevity, we found that healthy aging is not associated with known longevity variants. We observe a significant depletion of *APOE*-ε4 alleles, which is associated with longevity but considered to be 'frailty' allele for late onset disease rather than a true longevity allele. Similarly, we observe a reduced overall genetic susceptibility to Alzheimer and coronary artery disease in healthy aging, as determined by genetic risk scores, but no reduced risk for type 2 diabetes, stroke, and common cancers. Additionally, we observe no association with a decreased rate of rare pathogenic variants for common Mendelian disorders (ACMG secondary findings genes, inherited cancer, nor familial dementia), potentially indicating the presence of disease-resistance factors. In keeping with this possibility, by performing a whole genome sequence based common variant genome-wide association study as well as a rare variant burden analysis; we identify suggestive common and rare variant genetic associations implying that protection against cognitive decline is a genetic component of healthy aging. Interestingly, the gene most enriched for rare variants in the health aging cohort is *COL25A1* – a neuron specific collagen that plays a critical role in regulating the buildup of Aβ fibrils.   These findings, based on a relatively small cohort, require independent replication. Overall, our results suggest healthy aging demonstrates characteristics similar to other complex polygenic phenotypes, and that healthy aging is a genetically overlapping but distinct phenotype from exceptional longevity that may be enriched with disease-protective genetic factors.

1560F

GWAS of frailty in older Caucasians and African Americans: The health and retirement study. *C. Wu, M.C. Odden, S.A. Ramsey, H. Bae.* Oregon State University, Corvallis, OR.

Background. Frailty is a clinical syndrome of decreased resilience to stressors, resulting from age-related declines in multiple organ systems. Frailty is common in older adults and is associated with death, disability, fractures, and falls. Heritability estimates of frailty ranged from 19% to 43%, suggesting genetic variation may play an important role in frailty. However, only a limited number of genes have been examined in candidate gene studies, with no consistent findings. No genome-wide association study (GWAS) has been published on frailty. **Methods.** We performed a GWAS to identify genetic variants of frailty in 4898 Caucasians and 650 African Americans (AAs) aged ≥65 years from the Health and Retirement Study (2006 and 2008 waves). Frailty was measured on a 0-5 scale counting the number of characteristics: low grip strength, slow gait speed, excessive weight loss (%), and low physical activity, defined based on the lowest quintile, and self-reported exhaustion. GWAS was conducted on ~1.65 million genotyped single nucleotide polymorphisms (SNPs) with minor allele frequency >5%. Race-stratified linear regression was conducted, adjusting for age, sex, and the top six eigenvectors derived from principal components analysis. The GWAS results were post-analyzed using the gene expression data from the Genotype-Tissue Expression project. **Results.** We found no genome-wide significant SNPs ($p < 5 \times 10^{-8}$), but 30 suggestive associations ($p < 5 \times 10^{-5}$) were present in Caucasians. Three SNPs, in strong linkage disequilibrium (LD), are located at *NTF3* gene, which is involved in the nervous system. One SNP was associated with the expression of *HDHD3* gene in 17 tissues, such as artery and muscle. Among AA, 2 SNPs, in strong LD, reached genome-wide significance, and 16 SNPs were suggestively related to frailty. The two genome-wide significant SNPs are located at *DYNC1I2* gene whose related pathways are involved in the immune system. One SNP that was suggestively associated with frailty in Caucasians was replicated in AAs ($p < 0.05$; consistent direction of estimate). However, strength and direction of many associations were different between Caucasians and AAs, suggesting there may be different genetic variants contributing to frailty by race. **Conclusions.** Many genetic variants may be associated with frailty. Caucasians and AAs may have different genetic architecture underlying frailty. Future GWAS using independent racially-diverse cohorts is needed to replicate these findings.

1561W

Somatic genetic variation in the brain during aging and in Alzheimer disease. *H.T. Helgadottir[1], P. Lundin[2], E. Wallén Arzt[1], L. Lilius[3,4], A-K. Lindström[3,4], C. Graff[3,4], M. Eriksson[1].* 1) Karolinska Institutet, Department of Biosciences and Nutrition, Center for Innovative Medicine, Huddinge, Sweden; 2) Science for Life Laboratory (SciLifeLab), Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; 3) Karolinska Institutet, Department of NVS, Center for Alzheimer Research, Division for Neurogeriatrics, Huddinge, Sweden; 4) Deptartment of Geriatric Medicine, Karolinska University Hospital, Huddinge, Sweden.

Mutation accumulation in cells creates genetic mosaicism in a tissue. This accumulation may contribute to aging, and to neurodegenerative disease such as Alzheimer disease (AD; MIM 104300). Alzheimer disease is a neurological disorder that causes cognitive dysfunction and mainly affects elderly people. The neuropathological findings seen in the brain of AD patients can also be seen in older unaffected individuals, suggesting neurodegeneration has a more common effect on aging than previously thought. In this study, 2.85 Mb of the genome was sequenced with the average depth of 660x. Individuals with early-onset AD (n=4), late-onset AD (n=4), and age- and gender-matched non-dementia individuals (n=8) were included in the study. Blood and brain (temporal cortex) samples from each individual were sequenced with the aim of identifying low-degree somatic mutations that might contribute to aging. Variants were called according to the GATK Best Practices workflow, and the somatic callers, Varscan2 and MuTect1. Variants with minimum 100X depth in both tissues and minimum 5% alternate allele frequency were included in the study. To determine brain- or blood-specificity of each variant we applied the Fisher Exact test ($p < 0.05$) on observed reference and alternate reads from each tissue. When comparing the variant calls from the somatic callers, little overlap was observed, highlighting the complexity of detecting rare variants. Overall, fewer brain-specific variants were observed in late-onset AD patients compared to the early-onset AD. In non-dementia individuals, more variants were observed in older brains compared to the middle-aged brains. The amount of cells in the brain is expected to decrease during normal aging, and even more so in AD patients where cell loss is one of the hallmarks of the disease. Therefore, fewer mutations were expected to be seen in the brain of AD patients, as evidenced in our study. Several variants are recurrently mutated in different individuals, and clusters of mutations are seen within the same loci. Our results indicate different genetic variation in the brain and blood, and that AD has an effect on the genetic variation in the brain. .

1562T

Genome-wide association and pathway analysis of smoking cessation. *S. Kim[1], H. Kim[1], Y. Yun[1], Y. Chang[2], S. Ryu[2], H. Shin[3], H. Kim[1].* 1) Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, South Korea; 2) Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea; 3) Department of Family Medicine and Health Screening Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea.

The aim of this study was to investigate genetic variants and biological pathway associated with smoking cessation. The discovery samples included 679 subjects recruited from the Kangbuk Samsung Cohort Study which evaluated a wide range of health traits and diseases among Korean adults. The replication samples consisted of 3228 subjects selected from rural Ansung and urban Asan cohorts which are two community-based cohorts. SNP genotyping was conducted using Illumina or Affymetrix microarray chips. Quality control (QC) procedures were conducted to remove SNPs with missing genotyping rate > 5%, MAF > 0.05, HWE<10^{-6}. After QC, Genotyped data were imputed by IMPUTE2 using 1000 genome as a reference data. SNPs with an imputation quality score (R^2) greater than 0.8 and MAF>0.05 were used for the association analysis. Genome-wide association analysis (GWA) was performed using the logistic regression in Plink Ver. 1.90 to test the association of SNPs with smoking status as categorical variables. To identify biological pathways, GSEA with MAGENTA algorithm were conducted. GSEA p-value with threshold of 75 percentile cutoff. Pathway information was obtained from KEGG and REACTOME of the molecular signature database (MsigDB v4.0). In the discovery phase, the smallest P-valued was identified for a marker of rs6531222 located in chromosome 2, but it was not replicated. Although no SNPs were observed as associated with smoking cessation at genome-wide significance level, suggestive evidence of association was found in rs59476775 located near DUSP26 gene. In pathway analysis, Regulation of Insulin Secretion was top-ranked pathway in the discovery sample. Regulation of Insulin Secretion and MAPK Signaling Pathway were also found in the replication study. These findings contribute to our understanding of the genetic effects on smoking cessation and provide critical clues to biological mechanisms that influence them. [This research was supported by Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare (HI14C0072) and the National Research Foundation of Korea (NRF) funded by ICT & Future Planning (NRF-2014R1A2A2A04006291) and the Ministry of Education (NRF-2013R1A1A2062702).].

1563F

Genome-wide association study of male sexual orientation. *A.R. Sanders[1,2], G.W. Beecham[3], S. Guo[3], K. Dawood[4], G. Rieger[5], J.A. Badner[6], E.S. Gershon[2], R.S. Krishnappa[7], A.B. Kolundzija[8], J. Duan[1,2], P.V. Gejman[1,2], J.M. Bailey[9], E.R. Martin[3].* 1) Res Inst, NorthShore Univ HealthSystem, Evanston, IL; 2) Dept of Psychiatry and Behavioral Neuroscience, Univ of Chicago, Chicago, IL; 3) Dept of Human Genetics, Univ of Miami, Miami, FL; 4) Dept of Psychology, Pennsylvania State Univ, University Park, PA; 5) Dept of Psychology, Univ of Essex, Colchester, England, UK; 6) Dept of Psychiatry, Rush Univ Medical Center, Chicago, IL; 7) Dept of Psychiatry, Icahn School of Medicine at Mount Sinai, Elmhurst, NY; 8) Dept of Sociomedical Sciences, Mailman School of Public Health, Columbia Univ, New York, NY; 9) Dept of Psychology, Northwestern Univ, Evanston, IL.

Male sexual orientation is a scientifically and socially important trait shown by family and twin studies to be influenced by environmental and complex genetic factors. Genetic linkage studies have identified pericentromeric chromosome 8 and chromosome Xq28 as containing genetic variants contributing to its development, however, the linkage peaks are large and specific trait genes have not been identified. We conducted a genome-wide association study (GWAS) on a sample (N=2,381) including family data (i.e., GWAF) using Affymetrix single nucleotide polymorphism (SNP) arrays. Our analyzed European ancestry sample consisted of 1,273 sporadic (without family history data) samples (88 homosexual men and 1,185 heterosexual men) and 1,108 samples from families with homosexual brothers (1,022 homosexual men, 40 bisexual men, and 46 heterosexual men). After quality control, imputation, and analysis, we identified several SNPs with $p<10^{-5}$, including regions of multiple supporting SNPs on chromosomes 13, 14, and X. The genes nearest to these peaks have functions plausibly relevant to the development of sexual orientation. On chromosome 13, *SLITRK6* is a neurodevelopmental gene mostly expressed in the diencephalon, which contains a region previously reported as differing in size in men by sexual orientation. On chromosome 14, *TSHR* genetic variants in intron 1 could conceivably help explain past findings relating familial atypical thyroid function and male homosexuality. Furthermore, skewed X chromosome inactivation has been found in the thyroid condition, Graves' disease, as well as in mothers of homosexual men in other studies. On chromosome Xq25, *GRIA3* alterations in a mouse model resulted in unusual social behaviors in male-male interactions. On pericentromeric chromosome 8 within our previously reported linkage peak, we found nominal support for a SNP association reported in a previous scientific meeting (ASHG 2012). Although increasing evidence highlights genetic contributions to male sexual orientation, our current understanding of contributory loci is still limited, consistent with the complexity of the trait. Further increasing genetic knowledge about male sexual orientation should help advance our understanding of the biology of this important trait.

1564W

Methylenetetrahydrofolatereductase C677T polymorphism and the risk of Acute Myeloid Leukemia in Indian population: A case-control association study. *J.C. Silas[1,2], R. Reddy[3], S. Vishnupriya[4], V. Sugunakar[4], S. Rajender[3], K. Thangaraj[2].* 1) Taibah University, Madina, Saudi Arabia; 2) Centre for Cellular and Molecular Biology, Hyderabad, India; 3) Central Drug Research Institute, Lucknow, India; 4) Osmania University, Hyderabad, India.

Background: Methylenetetrahydrofolate reductase (MTHFR) enzyme is involved in folate metabolism and is crucial for maintaining DNA integrity in the cells**.** MTHFR gene polymorphisms have been previously associated with acute myeloid leukemia in many populations. The aim of this study is to evaluate the association of MTHFR C677T polymorphism with acute myeloid leukemia in Indian population. **Methods:** A total of 170 cases with acute myeloid leukemia and 243 age matched controls were tested for the C677T polymorphism using polymerase chain reaction followed by Sanger sequencing of the amplified product. **Result:** No significant association was found between MTHFR C677T Polymorphism and acute myeloid leukemia **Conclusion:** MTHFR C677Tpolymorphism is not strongly associated with acute myeloid leukemia in Indian population.

1565T

Haplotypes of vitamin D receptor gene: Both risky and protective for fibromiyalgia (FMS). *M. Akyol[1], A. Baklarlı[2], E. Tepeli[3], H. Balkarlı[1], S. Temel[4], V. Cobankara[4].* 1) Dept. of Medical Biology and Genetics, Akdeniz Universitiy, Antalya,Turkey; 2) Dept. Rhomotology-internal medicine, Antalya Training and research Hospital, Antalya,Turkey; 3) Dept. of Medical Genetics, Pamukkale Universitiy, Denizli,Turkey; 4) Dept. Rhomotology-internal medicine, Pamukkale Universitiy, Denizli,Turkey.

Background:Clinical observations suggest that vitaminD may influence fibromyalgia etiology. FMS and vitamin D deficiency share a similar symptom profile;the relationship between these conditions is unclear.VitaminD receptor is a crucial mediator of the symptoms of vitaminD deficiency.Thus, vitaminD receptor gene polymorphisms or haplotypes may contribute to vitaminD resistance.Objectives: This study evaluated the clinical relationship between FMS and vitaminD gene among Turkish FMS patients.Methods:73 patients with FMS were diagnosed according to the ACR 2010 criteria; the control group included 61 healthy unrelated volunteers.The rs2228570(Fok I), rs1544410(Bsm I), rs7975232(Apa I) and (rs731236.Taq I polymorphisms were examined using PCR-RFLPResults:In individual SNP analyses, none of the snps have been found to be associated with FMS(p values are 0.491, 0.477, 0.207, and 0.736 respectively). However, it has been detected that there is a strong association between haplotypes of VDR gene polymorphisms and FMS. As a result of this study, the effect of haplotypes is shown to be both risky and protective. P value of risk haplotypes ATC(rs2228570,rs1544410,rs7975232) is 0,015 (26,5%) and TTT is 0,0068 (12,8%).The p value of collected risk haplotypes (ATC, TTT) is 0,006 (39,3%) and (ATC, TGT, and TTT)0,000009 (58,7%).Frequent protective haplotype TTC is 0,0385 (6,8%).The p value of collected protective haplotypes (ATT, TCG, and TTC) is 0,008 (38%).Conclusions:The present study is the first study evaluating VDR gene in patients with FMS.Our results suggest that haplotypes in VDR are strongly associated with fibromyalgia expressing both risk and protective effect.Our findings may help guide future research needed to define the role of vitaminD in FMS.

1566F

Improving the resolution of GWAS in populations of African descent. *S. Chavan[1], M. Daya[1], N.M. Rafaels[1], H.R. Johnston[2], M. Boorgula[1], I. Ruczinski[3], T.H. Beaty[4], M. Taub[3], R.A. Mathias[4,5], K.C. Barnes[1], CAAPA consortium.* 1) Department of Medicine, University of Colorado, Denver, CO; 2) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 3) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD; 4) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD; 5) Department of Medicine, Johns Hopkins University, Baltimore, MD.

The first genome-wide association study (GWAS) of asthma focused exclusively on African ancestry populations (464 + 471 African American cases and controls, and 299 + 355 African Caribbean cases and controls) identified an association between dipeptidyl peptidase 10 (*DPP10*) and clinical asthma (Mathias et al 2009). This gene was first identified as a candidate gene for asthma by positional cloning in 2003, with some evidence for association, but this evidence for association was not replicated across independent populations. We hypothesized that inadequate tagging of African ancestry genes by the commercial genotyping array used in the 2009 study (Illumina 650Y), combined with subtle differences in ancestry between populations, contributed to a failure to replicate associations for individual SNPs, even though gene-based evidence was seen. In addition, none of the HapMap imputed SNPs around the top SNP, rs1435879, were supportive, possibly due to inadequate representation of African ancestry in the imputation reference panel. As part of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), where whole genome sequence (WGS) data (30X coverage) was generated on 883 unrelated individuals of self-reported African ancestry from 15 North, Central, and South American and Caribbean populations (plus Yoruba-speaking individuals from Ibadan, Nigeria) the 'African Diaspora Power Chip' (ADPC) was developed. The ADPC is a genotyping array consisting of tagging SNPs useful in comprehensively identifying African specific genetic variation when combined with a commercially available GWAS chip. We genotyped a subset of subjects from the original Mathias et al. GWAS on the ADPC (678 cases and 723 controls), and imputed the combined GWAS and ADPC data sets against the CAAPA genomes, which provided a much better representation of African haplotypes compared to the current 1000 genomes reference panel. In this new imputed data set, rs1435879 was still the most significant association in the DPP10 region, but several intronic SNPs surrounding rs1435879 now supported the association signal. This finding highlights the importance of adequately tagging African ancestry variants, and using sufficient African ancestry representation for imputation reference panels, when doing genetic association studies in African ancestry populations.

1567W

Phenome-wide association study of *SULT1A1* copy number variation and correlation with estrogen metabolism. *J. Liu[1], R. Zhao[1], Z. Ye[2], A. Frey[3], N. Snyder[3], S. Hebbring[1].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 2) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI; 3) A.J. Drexel Autism Institute, Drexel University, Philadelphia, PA.

Human cytosolic sulfotransferase 1A1 (SULT1A1) is believed to be one of the most important SULT isoforms. This theory is driven by observations that SULT1A1 is widely expressed in multiple human tissues and enzyme experiments demonstrate SULT1A1 has a wide range of phenolic substrates. As such, SULT1A1 has been hypothesized to play a significant role in drug, hormone, and neurotransmitter metabolism, while also being involved in detoxification and carcinogenesis. Interestingly, *SULT1A1* is subject to common copy number variation (CNV) where the entire gene can be deleted and duplicated. To better understand the biological role of SULT1A1 in human health, we genotyped the CNV in 14,275 Marshfield Clinic patients linked to an extensive electronic health record (EHR) system. The majority of patients (64%) had 2 copies of *SULT1A1* whereas 36% had >2 copies (max 6 copies) and 11 patients (0.08%) were homozygous deleted. This is the first time individuals with 0 copies of *SULT1A1* have been identified. To determine if individuals with 0 copies is detrimental, we manually examined the clinical data from these patients and found no obvious shared health conditions. To further assess the importance of *SULT1A1* CNV, we conducted comprehensive phenome-wide association studies (PheWASs). PheWAS results using ICD9 coding and clinical text data provided evidence that *SULT1A1* CNV may play an important role in response to common allergens. The strongest association included the clinical documentation of "Nasacort" (p=6.2E-7), which is a medication commonly used to treat allergic rhinitis. This association passed a conservative Bonferroni correction. To further understand the potential biological involvement of SULT1A1 activity on hormone metabolism, most notably estrogen metabolism, we quantified estrogen metabolites in plasma from 100 patients with *SULT1A1* CNV genotype data. The patients were selected for extreme CNV variation and included 50 males between the ages 40-50 and 50 premenopausal women between ages 30-40. Results showed that *SULT1A1* CNV was positively correlated with estrone to estrone sulfate conjugate ratio (E1/E1S; p=0.03, r=0.32). With evidence demonstrating for the first time that *SULT1A1* CNV is associated with common allergies and may be involved in hormone metabolism according to ex vivo experiments, this study highlights the potential broad importance of SULT1A1 in a variety of biological processes that are important in human health.

1568T

SCN10A **gene variant in a case of sudden infant death syndrome.** *K. Leavitt[1], T. Goldwaser[1], S. Klugman[1], R. Marion[2], Y. Tang[3], T.V. McDonald[2].* 1) Division of Reproductive and Medical Genetics, Department of Obstetrics, Gynecology and Women's Health, Montefiore Medical Center, Bronx, NY; 2) Division of Genetic Medicine, Department of Pediatrics, Montefiore Medical Center, Bronx, NY; 3) Molecular Genetics Laboratory, Office of Chief Medical Examiner, New York, NY.

Background: Sudden infant death syndrome (SIDS), has an incidence of 38.7 deaths per 100,000 live births. The majority of cases appear to be multi-factorial, a combination of genetic susceptibility and environmental risk factors. Several single gene disorders have been recognized as causes of SIDS, such as mutations in one or more cardiac channel genes and fatty acid oxidation disorders. Case: 34 y/o G8P4032 who was referred to genetics because of a history of infant deaths at 2 months of age. The first one was a male from a different partner who died at 2 months of age in El Salvador in 1996. The second infant death was in October 2015, when her daughter of 2 months of age was found unresponsive on her crib. The patient has 2 healthy living children. The cause of death was preliminarily established as SIDS. Clinical course and diagnostic work-up: An autopsy was obtained through the Office of the Chief Medical Examiner of New York City, which included molecular testing of 95 cardiac arrhythmogenic genes, revealing a variant of uncertain significance, NP_006505.2: p.Leu867Phe in one allele in the SCN10A gene in the deceased child. The SCN10A gene encodes for the protein Nav1.8 of the voltage-gated sodium channel alpha subunit. In 2015, mutations in this gene have been reported in Brugada syndrome and atrial fibrillation, and are inherited in autosomal dominant pattern. However, other genetic and environmental factors influence disease expression and severity. Maternal testing was obtained and the results show that the mother does not carry that variant on the SCN10A gene, while testing on the father of both the deceased child and the current pregnancy is still pending. The patient declined amniocentesis. Plan: Given the possible association of an SCN10A variant, the cause of death of her previous child, and the unknown status of the father, the following were recommended: delivery at a tertiary hospital, continuous neonatal monitoring and electrocardiogram, presume the risk for Brugada syndrome and proceed accordingly. The patient was also cautioned to avoid high risk drugs, use crib monitoring for breathing and movement and consider umbilical cord blood sampling for targeted DNA sequencing for the SCN10A variant. Conclusion: The use of a molecular panel for cardiac arrhythmia genes has been included in the diagnostic work-up for SIDS in some states. This case is ongoing to determine the role of this SCN10A variant in this patient's multiple infant deaths.

1569F

Systematic evaluation of risk loci from candidate gene and genome-wide association studies of fatigue. *E.C. Corfield[1], S.M. Marshall-Gradisnik[2], N.G. Martin[3], D.R. Nyholt[1].* 1) Faculty of Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia; 2) National Centre for Neuroimmunology and Emerging Diseases, Menzies Health Institute Queensland, School of Medical Science, Griffith University, Southport, Australia; 3) QIMR Berghofer Medical Research Institute, Brisbane, Australia.

Chronic fatigue syndrome (CFS) is a complex, neurological disorder of unknown pathophysiology. However, the immune, endocrine, nervous, cardio-vascular, and digestive systems have been implicated. Additionally, classical twin studies have been utilised to estimate the heritability of CFS in females. The estimated contribution of additive genetic, common environmental, and unique environmental factors has been estimated at 51%, 12%, and 36%, respectively (Buchwald, et al., 2001). Limited studies have been conducted to investigate the underlying genetics of CFS, the majority of which involved very small sample sizes. The genetic contribution has been implicated through gene expression analyses, bioinformatic data mining approaches, candidate gene studies, and most recently genome-wide association (GWA) studies. However, replication studies are required to confirm these exploratory findings. Therefore, the present study aims to evaluate the findings from candidate gene and GWA studies. Additionally, the selected genes will be evaluated in a more general fatigue phenotype. Seventeen genes with nominal evidence for association were selected from published candidate gene studies while 323 genes plus 188 inter-genic SNPs were selected from published CFS GWA studies. SNPs within the study genes and 15 kb regions upstream and downstream (Pickerell, et al., 2010) were investigated in GWA data of 105 CFS cases and 71 controls. In addition, the selected genes were investigated in 307 fatigue cases and 745 controls. A kinship matrix was utilised to account for relatedness within the fatigue cohort, which was derived from a twin cohort. A Bonferroni-corrected significance threshold was calculated to assess the association of SNPs with CFS and fatigue. Although both GWA analyses contained relatively small sample sizes the CFS dataset contained over 2.5 times the number of cases compared to previously published CFS GWA studies. In addition to confirming association between previously implicated risk loci, results will be presented concerning the genetic similarity of CFS and more general fatigue.

1570W

Exome chip studies in allergic disease – Utilising multinomial logistic regression to uncover genetic associates of sensitisation disease phenotypes. *J.A. Curtin[1], A. Custovic[2], D.C. Belgrave[2], A. Simpson[1].* 1) University of Manchester, Manchester, United KingdomCentre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester, Manchester, Manchester, United Kingdom; 2) Department of Paediatrics, Imperial College London, London, UK.

   Background: Much of the heritability of complex traits (e.g. atopy, asthma) remains unexplained by genome wide association studies (GWAS), possibly due to poor phenotype definition. Conventionally, atopy is defined as the presence of a positive skin test or specific IgE to any allergen. We have proposed that this diagnostic label encompasses a number of different phenoypes. We applied machine learning approach to all available skin test and sIgE data collected from birth to age 8 years in the Manchester Asthma and Allergy Study (MAAS) to cluster children into different atopy classes (Simpson *et al.* AJRCCM 2010;181:1200-6); the majority of atopic children clustered into 4 distinct classes. We hypothesized that genetic associates of different phenotypes will be characteristic. **Methods:** We genotyped children from MAAS using the "Human Exome-12 v1.1 Beadchip", which resulted in 39714 SNPs after quality control and removing uninformative SNPs. We carried out a multinomial logistic regression likelihood test using the newml option in SNPT-EST (v2.5.2) on the sensitisation clusters where the null hypothesis is that are no differences between the different clusters. The four sensitisation classes were: "Multiple early" (122 cases), "Mite" (68 cases), "Grass" (34 cases); and "Persistent Sensitisation to a wide variety of allergens" (67 cases). The "No sensitisation" class (n=541) was the reference. We compared these results to those obtained from "Conventional atopy" (272 cases; 559 controls). **Results:** Using the additive model, we observed a trend association from the likelihood ratio test for 4 SNPs and differences between the 4 classes - 2 amino acid changing SNPs in *VWDE* (rs2192828 and rs848016 on 7p21; p=2.5E-06 and p=3.8E-05 respectively), 1 amino acid changing SNP in *OR52N4* (rs7394584 on 11p15; p=9E-05) and an intronic SNP in a gene with homology to mouse *Mir1192* (rs12446956 on 16q22 p=6E-05). In contrast we observed one intronic SNP in *WDR36* (rs10051830, 5q22; p=5.5E-05) was associated with conventional atopy. **Conclusions:** Atopy comprises multiple latent atopic vulnerabilities with different genetic risk factors. To date identifying the genetic associates of latent classes has been hampered by the multinomial nature of these phenotypes. Using a newly implemented multinomial logistic regression, we observed trend associations between different genetic regions precisely defined phenotypes of sensitisation.

1571T

Genome-wide association study identifies sex-specific genetic determinants of emphysema. *A. El Boueiz[1,2], M. Hardin[1,2], M.H. Cho[1,2], G.R. Washko[2], T.H. Beaty[3], J.D. Crapo[4], E.K. Silverman[1,2], B.J. Make[4], D.L. DeMeo[1,2] for the COPDGene Investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston; 2) Pulmonary and Critical Care Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 4) Division of Pulmonary Medicine, Department of Medicine, National Jewish Health, Denver, CO.

   Introduction: Chronic obstructive pulmonary disease (COPD) is the leading cause of chronic respiratory morbidity and mortality in the US. Despite the growing recognition that differences in COPD susceptibility and severity exist between male and female smokers, sex-specific determinants of COPD remain unknown. Emphysema, destruction of lung parenchyma, is commonly seen in smokers with COPD, and males and females with COPD demonstrate different percentages of radiographic emphysema at each stage of lung obstruction. The aim of this study was to identify sexually dimorphic genetic risk factors for emphysema. **Methods:** From the COPDGene cohort, we analyzed 4983 males and 4321 females with complete genotype and CT densitometry data. Genotyping was performed on the Illumina Omni Express platform with additional markers imputed using the 1000 Genomes Project. We conducted sex-stratified genetic association tests for radiographic emphysema using an additive model adjusting for age, body mass index, pack-years of smoking, current smoking, scanner type and genetic ancestry. Inverse normal transformation was applied to the phenotype to reduce the impact of deviation from normality. Separate analyses in non-Hispanic whites and African Americans were combined into a pooled analysis. **Results**: Stratification on the basis of sex reveals genome-wide significant genetic effects for markers in the 4q31 locus near *HHIP* in females but not in males (Effect size: 0.07±0.02, P: 3E-5 *(M)*; Effect size: 0.11±0.02, P: 1E-8 *(F)*). We did not detect any other genome-wide significant common SNP differences between men and women. However, three interesting top sub-threshold differential associations were found on chromosomes, 8p22 near *DLC1* (Effect size: -0.08±0.02, P: 8E-7 *(M)*; Effect size: -0.02±0.02, P: 0.23 *(F)*), 6q25 near *ESR1* (Effect size: -0.002±0.024, P:0.92 *(M)*; Effect size: -0.14±0.03, P: 4E-7 *(F)*) and 2p24 near *KCNK3* (Effect size: -0.02±0.02, P: 0.26 *(M);* Effect size: -0.12±0.02, P: 8E-7 *(F)*). **Conclusion:** There is mounting evidence of an important role for sex-differentiated effects in the architecture of complex traits such as COPD. This GWAS of quantitative emphysema in smokers identified potential risk loci with differential associations in males and females. These findings may help elucidate sexually-dimorphic mechanisms of the disease and inform sex-specific therapeutics and preventive strategies

1572F

Functional mutations in GRHL3 gene are associated with cleft palate in sub-Saharan Africa. *M.A. Eshete[1], W.L Adeyemo[2], L.J.J Gowans[3], P.A Mossey[4], T. Busch[5], O. James[2], E.A Augustine-Akpan[8], M.L Marazita[6], A.A Adeyemo[7], J.C Murray[5], A. Butali[8].* 1) Addis Ababa University, Addis Ababa, Select a Country; 2) University of Lagos, Nigeria; 3) Kwame Nkrumah University of Science and Technology, Ghana; 4) University of Dundee, UK; 5) Department of Pediatrics University of Iowa, USA; 6) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, U.S.A; 7) National Human Genomic Research Institute , Bethesda MD; 8) Department of Oral Pathology, Radiology and Medicine, University of Iowa, USA,.

Background/Objectives: Although progress has been made towards understanding the genetic etiology of cleft lip with or without cleft palate (CL/P), few significant findings have been reported for isolated cleft palate (CPO). The first gene with clear evidence of causality in CPO, Fas-associated factor-1 (*FAF1*), was reported by Ghassibe et al. (2011) . Recently*,* a common coding variant of *GRHL3* was shown to be associated with risk for non-syndromic CPO in Europeans (Leslie et al ., 2016). Mutations in this gene have also been reported in families with Van der Woude syndrome. The current study sequenced the *GRHL3* gene in cases with CPO from Africa. **Materials and methods:** We recruited participants from Ghana, Ethiopia and Nigeria. This cohort included case-parent trios, cases and other family members, as well as controls. Seventeen coding regions (exons) of *GRHL3* were amplified using Polymerase Chain Reaction (PCR). We sequenced exons of this gene in DNA from 134 non-syndromic and also the parent samples to identify *de novo* variants. Novel variants observed in cases were sequenced in 270 controls. **Results:** Six novel mutations were discovered (2 Ghanaian, 1 Ethiopian and 3 Nigerians,). The mutations were: missense (p.Pro166His, p.Asp410Gly and p.Arg603Lys), splice site (p.Ser428Arg), frameshift mutation (p.Gly65AlafsTer55) and nonsense mutation (p.Tyr559X). None of these mutations were seen in 270 controls and in any known exome and whole genome databases including the 1000 genomes database that has data from Africa. The p.Pro166His, p.Asp410Gly and p.Gly65AlafsTer55 segregate in the family and were seen in unaffected mothers. Mutations p.Ser428Arg, p.Arg603lys and p.Tyr559X were not seen in mothers and we could not ascertain de novo since father samples were unavailable. The p.Gly65AlafsTer55 and p.Tyr559X will lead to premature truncation of the protein. The p.Asp410Gly mutation damaged a sumoylation site. Serine and Threonine phosphorylation is also affected by the p.Ser428Arg mutation which damaged the CK1 phosphorylation site. Based on the amino acid properties and conservation of the wild type at position p.Asp410Gly and p.Ser428Arg in several species, these mutations were predicted to be probably damaging to the protein by Polyphen and SIFT. **Conclusions:** Our study provides evidence that, as in Caucasian populations, mutations in *GRHL3* may contribute to the risk of non-syndromic CPO in the African population.

1573W

Genetics of bilirubin serum levels and chronic obstructive pulmonary disease. *N. Fishbane[1], M. Obeidat[1], N.N. Hansel[2], N. Rafaels[3], R. Mathias[4], I. Ruczinski[5], T.H. Beaty[6], K.C. Barnes[3], P.D. Paré[1], D.D. Sin[1].* 1) Centre for Heart Lung Innovation, University of British Columbia, Vancouver, BC, Canada; 2) Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA; 3) Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO, USA; 4) Division of Genetic Epidemiology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA; 5) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 6) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

Background: Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide. The disease is characterized by increased oxidative stress. Bilirubin, a product of heme catabolism, has been shown to have antioxidant and anti-inflammatory properties. Serum levels of bilirubin correlate positively with lung function and high levels are associated with reduced incidence of COPD. We hypothesized that serum bilirubin levels are under genetic control in COPD subjects. **Methods**: The Lung Health Study (LHS) is a longitudinal multicenter study of mild-to-moderate COPD subjects, who were followed over 11 years. In year 5, bilirubin serum levels were measured in 4,101 individuals who also had whole genome genotypes. A genome-wide association study (GWAS) for bilirubin levels was undertaken using 1000 genome imputed genotypes, assuming an additive genetic model, and adjusting for age, sex, BMI, and genetic principal components. The top bilirubin-associated single nucleotide polymorphisms (SNPs) were tested for association with lung function in LHS subjects and in a number of large publically available lung function GWAS data sets including the UKBiLEVE study (n=15,000) and the SpiroMeta consortium (n=38,000). **Results**: One previously known locus was strongly associated with bilirubin levels on chromosome 2 in the UDP glucuronosyltransferase family 1 member A complex locus (UGT1A, P = 6.77E-135). The UGT1A genes are involved in the conjugation of bilirubin. Given this was the only genome-wide significant locus, a second GWAS was performed conditioning on the UGT1A locus. Three loci were borderline significant in this analysis: the previously identified solute carrier organic anion transporter family member 1B1 (*SLCO1B1*) gene region on chr12 (P= 1.44E-7), and two novel loci including *KIAA1217* gene on chr10 (P=5.8E-8) and an intergenic region on chr21 (P= 8.05E-08). The most significant SNP from the UGT1A locus was nominally associated with lung function in heavy smokers in the UKBiLEVE study (P = 0.039). **Conclusion**: The bilirubin GWAS in COPD subjects replicated previously identified loci and suggested two novel regions. Integrating bilirubin's genetic determinants with lung function suggests that bilirubin may contribute to the pathogenesis of COPD.

1574T

A genome-wide association study of vertical cup-disc ratio in Latinos. *X. Gao[1], D. Nannini[1], M. Torres[2], Y.I. Chen[3], K.D. Taylor[3], J.I. Rotter[3], R. Varma[2]. 1) Dept Ophthalmology & Visual Sci, Univ Illinois, Chicago, Chicago, IL; 2) USC Roski Eye Institute, Department of Ophthalmology, University of Southern California, Los Angeles, CA; 3) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics and Medicine at Harbor-UCLA, Torrance, CA.*

Vertical cup-disc ratio (VCDR) is used as a clinical assessment measure to identify and monitor glaucomatous damage to the optic nerve. Previous genetic studies conducted in European and Asian populations have identified many loci associated with VCDR. The genetic factors in other ethnic populations, such as Latinos, influencing VCDR remain to be determined. Here, we describe the first genome-wide association study (GWAS) on VCDR in Latinos. We conducted this GWAS on VCDR using 4,537 Latinos who were genotyped using either the Illumina OmniExpress BeadChip (~730K markers) or the Illumina Hispanic/SOL BeadChip (~2.5 million markers). Study subjects were 40 years of age and older. Linear regression, adjusting for age, gender, and principal components of genetic ancestry, was conducted to assess the associations between single nucleotide polymorphisms (SNPs) and VCDR. We imputed SNPs from the 1000 Genomes Project to integrate additional SNPs not directly genotyped.We replicated two previously reported SNPs that reached GWAS significance, rs1900005 and rs7916697, in the *ATOH7-PBLD* region, as well as identified two suggestive associations in the *CDC7-TGFBR3* region on chromosome 1p22.1 and in the *ZNF770-DPH6* region on chromosome 15q14. We discovered a novel SNP, rs56238729 (P = 8.56E-11), in the *ATOH7-PBLD* region that is significantly associated with VCDR in Latinos. We replicated 10 previously reported regions, including *COL8A1, CDKN2BAS, BMP2,* and *CHEK2* (P < 2.17E-03). Our results discovered a novel SNP that is significantly associated with VCDR in Latinos and confirmed previously reported loci, providing further insight into the genetic architecture of VCDR.

1575F

Genetic and environmental factors influencing cardiovascular, anthropometric, and metabolic biomarker trait variation in sub-Saharan African populations. *M.E.B. Hansen[1], A. Ranciaro[1], S. Thompson[3], J. Hirbo[4], W. Beggs[1], M. Ibrahim[5], T. Nyambo[6], S. Omar[7], C. Wambebe[8], D. Meskel[9], G. Belay[9], M. Bilow[10], E. Eskin[10], C. Abnet[11], M. Yeager[11,12], S. Chanock[11], S.A. Tishkoff[1,2]. 1) Genetics Dept., University of Pennsylvania, Philadelphia, PA; 2) Biology Dept. University of Pennsylvania, Philadelphia, PA; 3) University College London, London, UK; 4) Vanderbilt University, Nashville, TN; 5) Dept. of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan; 6) Dept. of Biochemistry, Muhimbili University of Health and Allied Sciences, Dares, Salaam, Tanzania; 7) Kenya Medical Research Institute, Center for Biotechnology Research and Development, Nairobi, Kenya; 8) International Biomedical Research in Africa, Abuja, Nigeria; 9) Dept. of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 10) Depts. of Computer Science and Human Genetics, UCLA, Los Angeles, CA; 11) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD; 12) Cancer Genomics Research Laboratory, NCI-DCEG, SAIC-Frederick Inc., Frederick, MD.*

The African continent is home to a diverse range of indigenous peoples that have adapted to a wide range of ecological environments and subsistence lifestyles. Many complex traits are expected to display variation between populations due to demographic history and/or natural selection to these diverse environments. In an effort to survey phenotypic variation in Africa and begin to understand the genetic and environmental factors that contribute to this variation, we measured 22 traits in a total cohort of ~4,900 individuals, with numbers per trait ranging from ~900 to 4,900 individuals. The cohort is comprised of agricultural, pastoral, and hunter-gatherer populations across eastern and western sub-Saharan Africa (Cameroon, Nigeria, Sudan, Ethiopia, Kenya, and Tanzania). The traits include anthropometric measurements (e.g. height, weight, waist circumference), cardiovascular measurements (e.g. systolic and diastolic blood pressure), and a panel of metabolic blood biomarkers (e.g. LDL, HDL, triglycerides, glucose, NEFA, C-peptide). We present the observed variation in these traits between sexes, across populations, and across dietary categories. We find significant differences between populations in both the mean trait values and the tails of the trait distributions. A subset of ~1,700 individuals were genotyped on the Illumina 1M-Duo or the Illumina 5M Omni SNP array. We performed genome-wide association tests using a linear mixed model approach on this subset, and analyzed the reproducibility of associations for loci identified in non-African cohorts. Several traits show strong ethnicity specific associations, including height and blood pressure. We estimated the trait heritability values and gene-by-diet/environment interactions, and discuss the degree to which genetic ancestry and subsistence practice impacts trait variation. This work underscores the need to include ethnically diverse populations in gene association studies and to develop better methodological approaches for leveraging association studies from small cohorts with diverse genetic and long-term dietary/environmental conditions.

1576W

The role of novel and low-frequency variants contributing to body fat distribution. *A. Justice[1], T. Karaderi[2], H. Highland[1], K. Young[1], M. Graff[3], P. Auer[3], V. Turcot[4], K. Lu[5], C. Schurmann[6], N. Heard-Costa[7,8], E. Marouli[9,10,11], J.A. Pospisilik[12], I. Borecki[13], L.A. Cupples[7,14], R.J.F. Loos[5], K.E. North[1], C.M. Lindgren[15] on behalf of GIANT, GoT2D, CHARGE, BBMRI-NL, and deCODE.* 1) Epidemiology, University of North Carolina, Chapel Hill, NC, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) Department of Biostatistics, University of Wisconsin-Milwaukee, Milwaukee, WI, USA; 4) Montreal Heart Institute, University of Montreal, Canada; 5) Icahn School of Medicine at Mount Sinai, New York, NY, USA; 6) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 7) National Heart, Lung, and Blood Institute, the Framingham Heart Study, Framingham, MA, USA; 8) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 9) William Harvey Research Institute, London, UK; 10) Barts and The London School of Medicine and Dentistry, London, UK; 11) Queen Mary University of London, London, UK; 12) Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany; 13) Department of Genetics Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA; 14) Boston University School of Public Health, Boston, MA, USA; 15) The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, UK.

Body fat distribution, measured here as waist-to-hip ratio adjusted for BMI (WHR), is a heritable trait and a leading risk factor for cardiometabolic diseases. The majority of genetic studies focus on common, non-coding variants, but the genetic underpinnings of WHR may include rare and protein-coding variants with large effects. We aim to identify protein altering (coding/splicing), low-frequency variants (LFV) (MAF<5%), and genes influencing WHR using exome array data. First, we meta-analyzed study- specific single variant (SV) association results for ~240,000 variants from 344,369 individuals of European, African, Hispanic/Latino, South and East Asian ancestry. We conducted gene-based association tests using the sequence kernel association test (SKAT) method including up to 16,222 genes defined by LFVs predicted as damaging. Also, due to known gender differences in effects on WHR, we analyzed associations stratified by sex. We took the top 70 variants (P<2E-6) forward to validation and meta-analyses in an additional 132,177 European individuals. All tests assumed an additive model. For combined sexes, 47 (21 novel) coding/splicing variants reached array-wide significance (P<2E-7) in our meta-analysis, including six novel LFVs in *FGFR2*, *R3HDML*, *HIST1H1T*, *ACVR1C*, *RAPGEF3*, *PCNXL3*. Four additional variants were identified in the sex-specific analyses, including 2 novel LFV variants in *UGGT2* (men) and *ANGPTL4* (women). The largest effects were observed for LFV in *FGFR2* and *HIST1HIT* genes with 0.02 per allele change in WHR. Three genes identified in our SV analyses were also significantly associated in our gene-based results, including *RAPGEF3* for our total sample and *ACVR1C* and *ANGPTL4* in our women-only analysis (P<2.5E-6). Like known WHR genes, both *ANGPLT4* and *RAPGEF3* are involved in angiogenesis, a necessary pre-cursor for pre-adipocyte differentiation. Further, *ACVR1C* encodes ALK7, which is highly expressed in the brain, pancreas, and adipose tissues. Both ANGPLT4 and ALK7 may be involved in lipid homeostasis. We then performed a look up of our potential genes in adult *Drosophila* knockdowns to assess their role in the regulation of fat storage. We found significant increases in fat storage for two genes with common nonsynonymous variants associated with WHR, *PLXND1* and *DNAH10*. Our approach examining variants often missed by GWAS yielded specific novel genes that may impact adiposity, stressing the importance of targeting protein altering and LFV.

1577T

Genetic markers as risk or protection for mountain adaptation and disease. *S. Kohli[1,2], G. Mohammad[3], R. Kukreti[1,2], M.A.Q. Pasha[1,2].* 1) CSIR-Institute of Genomics and Integrative Biology, New Delhi, India; 2) Academy of Scientific and Innovative Research (AcSIR), New Delhi, India; 3) Department of Medicine, Sonam Norboo Memorial Hospital, Leh, Ladakh, Jammu, and Kashmir, India.

High-altitudes (HA) (1500–3500m), characterized by hypobaric hypoxic environment, are home to >140million people. The HA natives, over the generations, have undergone genetic remodelling to adapt in this extreme environment and hence exhibit distinct physiologic traits. However, in susceptible sojourners, hypobaric hypoxia stimulates oxidative stress triggering endothelial damage and vascular wall remodelling leading to high altitude pulmonary edema (HAPE). Hence, genetic analysis of HA natives and sojourners would offer a unique platform to understand the pattern of variations for mountain adaptation and diseases. We conducted a genome-wide study using Omni 1M beadchip, followed by replication using customized 1.1K chip, to identify differentiated variants, i.e., single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) for adaptation and maladaptation in three well defined study groups namely HA natives, HAPE–controls and HAPE–patients, recruited through Sonam Norboo Memorial hospital, Leh (3500m), Ladakh, Jammu & Kashmir, India. Extensive statistical analysis and population stratification with HAPMAP and Indian genome variation populations identified variants, which were further validated for functionality by appropriate downstream experiments. Genetic analysis of HA natives revealed SNPs in *EPAS1* and *EGLN1*, and CNV in *LRRIQ1* and *SLC6A15* as most significant signals (P<1.00E−10); suggesting the regulation of vascular homeostasis and cellular oxygen sensing to direct adaptive cellular responses in natives. In HAPE, SNPs in *CHI3L1* and *DLC1* emerged as most significant signals (P≤8.30E−4); and at mRNA and protein level, CHI3L1, a promoter of vasodilation, was down-regulated by 1.83 and 1.58 folds (P≤0.0001), respectively, and DLC1, a GTPase-activating protein involved in cytoskeletal changes, was up-regulated by 2.30 and 1.44 folds (P≤0.005), respectively, suggesting the importance of pulmonary vasoconstriction and alveolar leakage in HAPE pathogenesis. Similarly, in HAPE, the most significant signals for CNVs emerged in region spanning *IKZF3* and *MEX3D* (P≤1.30E−05), the genes that directly or via BCL-2, are known to up-regulate HIF1α expression, the master transcriptional regulator of the adaptive response to hypoxia. To conclude, the present study identifies genes related to cellular and hypoxia signalling thereby regulating cellular response at HA.

1578F

A genome-wide association scan identifies multiple common genetic variants on *MECOM* and *SALL3* for human fingerprint patterns. *J. Li[1,2], H. Zhang[2], Y. Yang[2,3], J. Tan[2], Q. Peng[1,2], A. Ding[1,2], S. Wu[1,2], Y. Liu[1,2], S. Zheng[1], L. Jin[1,2,3], S. Wang[1,2].* 1) Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, 200031 Sh; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, 200433 Shanghai, China; 3) Fudan-Taizhou Institute of Health Sciences, 1 Yaocheng Road, 225300 Taizhou, Jiangsu, China.

The fingerprint patterns are highly individualistic and have been utilized in methodological studies. They provide a permanent record of events during the first 4 months of foetal life. Here we report a genome-wide association study (GWAS) in 2,907 Han Chinese for patterns of fingerprint (arch, loop and whorl) on each digit. We find genome-wide significant association at two genomic regions (in 3q26 and 18q23) affecting patterns of fingerprint on multiple fingers. SNPs at 3q26 ($p=5.04\times10^{-12}$, $\beta=-0.097\pm0.014$) near *MECOM*, a complex locus of *MDS1* (myelodysplasia syndrome 1) and *EVI1* (Ecotropic virus integration site 1) are associated with patterns on six fingers (digit 2, 3 and 4 of both hands). The protein encoded by *MECOM* is a transcriptional regulator and oncoprotein that may be involved in hematopoiesis, apoptosis, cell differentiation and proliferation. *EVI1* also has the critical role in the development of forelimbs and fingers in humans. SNPs at 18q23 ($p=1.98\times10^{-12}$, $\beta=0.084\pm0.012$) near *SALL3* (Spalt-Like Transcription Factor 3), a key regulator of limb development at early stages, are associated with patterns on digit 5 of both hands. Previously reported variants in *ADAMTS9-AS2* at 3p14.1, associated with increased frequency of whorls in Caucasians, do not show any significance in our study (p=0.229). Our findings suggest that different genes may affect the formation of fingerprint patterns in different populations.

1579W

Genome-wide association study of 6,939 individuals identifies five novel loci and suggests that prenatal exposure to testosterone is not a major determinant of individual differences in 2D:4D finger ratio. *N.M. Warrington[1,2,3], G. Hemani[4,5], P.G. Hysi[6], M. Mangino[6], G. McMahon[4,5], M. Hickey[7], D. Wolke[8], G.W. Montgomery[9], M.J. Wright[9], D.A. Mackey[10], N.G. Martin[2], G. Davey Smith[4,5], C.E. Pennell[3], T.D. Spector[6], S.E. Medland[2], D.M. Evans[1,4,5].* 1) Diamantina Institute, University of Queensland, Brisbane, Queensland, Australia; 2) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 3) School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia; 4) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, England; 5) School of Social and Community Medicine, University of Bristol, Bristol, England; 6) Department of Twin Research and Genetic Epidemiology, King's College London, London, England; 7) Department of Obstetrics and Gynaecology, University of Melbourne, Royal Women's Hospital, Victoria, Australia; 8) Department of Psychology and Health Sciences, University of Warwick, Coventry, England; 9) Queensland Brain Institute and Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia; 10) Lions Eye Institute, Centre for Ophthalmology and Visual Science, The University of Western Australia, Perth, Western Australia, Australia.

It has long been hypothesized that during critical periods of development, including the prenatal period, sex steroids permanently alter the developing nervous system, which in turn influences behavior in later life. There is good evidence for the validity of this theory, referred to as the "Organizational Hypothesis", in experimental animals. However, it is difficult to test this hypothesis in humans as prenatal exposure to androgens is extremely difficult to measure. Based on several lines of indirect evidence, it has been hypothesized that the ratio of the length of the index finger to the length of the ring finger (2D:4D) is a marker of prenatal androgen exposure and could therefore be used as a retrospective non-invasive biomarker of testosterone exposure *in utero*. The 2D:4D ratio is sexually dimorphic; however, many studies of 2D:4D and sex related traits have involved small sample sizes and have been difficult to replicate, leading to substantial controversy over the utility and precise meaning of the measure. In the largest genome-wide association study of 2D:4D to date (N=6,939), we find little evidence to support the hypothesis that prenatal androgen exposure is a major driver of individual differences in 2D:4D. We show that the length of the CAG microsatellite in the androgen receptor gene, which influences sensitivity to testosterone, shows little evidence of association with 2D:4D. We also performed a genome-wide association study meta-analysis of 2D:4D and identified seven loci (five novel) that do not contain genes obviously related to testosterone. These seven loci explain 2.38% of the variance in the mean 2D:4D, which is equivalent to approximately half of the variance explained by sex. Furthermore, genetic variants associated with (adult) testosterone/sex hormone-binding globulin were not associated with 2D:4D. Together these findings suggest that 2D:4D is unlikely to represent a simple biomarker of prenatal androgen exposure and any effect of testosterone exposure on the measure may be small relative to the combined influence of other sources of variation. Accordingly, studies using 2D:4D as a biomarker for prenatal androgen exposure need to be cautious with respect to their broader interpretations relating to the Organizational Hypothesis.

1580T

Individuals at the tails of common phenotypes deviate from a polygenic mode of inheritance. *A.R. Wood, K.S. Ruth, R. Beaumont, J. Tyrrell, S.E. Jones, M.A. Tuke, H. Yaghootkar, R.M. Freathy, A. Murray, T.M. Frayling, M.N. Weedon. Genetics of Complex Traits, University of Exeter Medical School, Exeter, Devon, United Kingdom.*

Genome-wide association studies have shown that common genetic variants explain a large fraction of the heritability for many common diseases and quantitative traits. However, the relevance of common variants in individuals with extreme phenotypes is less understood. These individuals may carry a large number of common trait raising or lowering alleles, consistent with a polygenic model. Alternatively, they may deviate from this model due to monogenic or environmental factors. Using 120,000 British individuals from a single study, the UK Biobank, we aimed to determine whether participants at the extreme tails of several quantitative traits carried the number of common alleles expected under a polygenic model, and to identify individuals deviating from the polygenic model. Using published variants, we generated weighted allele scores (WAS) and tested the observed average WAS against the distribution of average WAS for a given percentile, derived from repeated simulations of the polygenic model. For all traits tested, the polygenic model was observed across the distribution until the extreme tails. For height, the shortest 1.5% of individuals carried fewer alleles associated with short stature than expected (P=5E-10), consistent with our previous work (Chan et al 2011). Inspection of ICD10 disease codes confirmed that some of these individuals had monogenic disorders of skeletal growth, including Turner's syndrome, spondyloepiphyseal dysplasia, and achondroplasia. For BMI, the thinnest 2% of individuals carried fewer BMI-lowering alleles than expected (P=6E-5). The most overweight 2% carried more BMI raising alleles than expected, although only in women (P=9E-5). This excess in the upper BMI tail remained after removing females self-reporting poor health, dieting, and regular exercise (P=4E-4). For menopause age and systolic blood pressure (SBP), we observed a significant excess and deficit of trait-raising alleles in the lower and upper tails, respectively. For menopause age, we observed deviation from a polygenic model in the youngest 2.5% (P=1E-6) and oldest 1.5% (P=5E-4). For SBP, we observed deviation in the lowest 1% (P=1E-3) and highest 0.5% (P=8E-3). We conclude that common genetic variants are associated with complex traits at the extremes as well as across the population, but additional factors consistent with either rare non-additive genetic or rare non-genetic factors become more prominent at the extremes.

1581F

Towards an innovative early prediction of pregnancy-related diseases with integrated analysis of daily life log records and multi-omics data. *T. Yamauchi[1], O. Daisuke[1], Y. Harada[2], R. Yamashita[2], T. Takai[2], S. Ogishima[2], N. Minegishi[2], O. Tanabe[2], S. Kuriyama[2], N. Fuse[2], N. Yaegashi[2], S. Hiyama[1], M. Nagasaki[2], J. Sugawara[2], Maternity-log Project Team. 1) Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan.*

A major issue in perinatal medicine is that personalized protocols for preventing various diseases related to pregnancy have not yet been established. Over two hundred thousand pregnant women each year are affected by pregnancy related disorders, including pregnancy induced hypertension, gestational diabetes, and premature deliveries in Japan. Those multifactorial diseases are caused by a complex interaction of genetic factors and acquired environmental factors such as lifestyle and living environment. To date, previous studies have been conducted to clarify pathogenesis of multifactorial diseases by using data on individual combinations of genomic analysis and changes in circulating factors, as well as subjective data on environmental factors collected through questionnaires. As described, environmental factors have been usually analyzed on the basis of questionnaires targeting an individual lifestyle, however, the accuracy, frequency, and currency of these data have been a major issue. To resolve these issues, we have designed a prospective cohort study "maternity-log study" for pregnant women. In this study, pregnant women are recruited in the first stage of pregnancy at routine obstetric visit in Tohoku University Hospital under satisfactory informed consent. Study participants upload daily information including morning sickness, abdominal pain, and uterine contractions, as well as physiological data such as body weight, blood pressure, heart rate, and body temperature. This study protocol obtained approval from the ethics review committee established by Tohoku University Graduate School of Medicine (ID 2014-1-704: application date 28 January 2015). To the best of our knowledge, this study will be the first to integrate multi-omics data of genome, transcriptome, proteome, metabolome, and metagenome, with objective data on environmental factors, including daily lifelog data of pregnant women. The present study is expected to elucidate the causal relationship between perinatal diseases and maternal lifestyle information, physiological changes, and multi-omics data and to establish methods for personalized early prediction of pregnancy related diseases. Precise protocol and expected results of this study will be discussed.

1582W

Genome-wide gene-environment interaction analysis for brain structure in healthy Korean women. *H. Kim[1], H. Kang[2], S. Kim[1], Y. Yun[1], Y. Chang[3], S. Ryu[3], D. Lee[2], H. Shin[4], H. Kim[1].* 1) Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, South Korea; 2) Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, South Korea; 3) Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea; 4) Department of Family Medicine and Health Screening Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea.

Neuroticism is known to be associated with the volume of different brain regions as well as emotion and health. Although recent genome-wide association studies have identified some novel loci for brain structure, few addressed genome-wide gene-environment interactions. To determine gene-personality interactions in brain volume, we found 11 different brain regions associated with neuroticism in 49 women scoring in the extremes of neuroticism (>90th and <10th percentile) and conducted genome-wide gene-environment interaction analyses for the brain regions. Based on analyses of 226,705 single nucleotide polymorphisms (SNPs) we found evidence for an interaction in which the associations between the genotype and brain volume depend on scores of neuroticism. The SNPs showed associations of the opposite direction with brain volume in low- and high-scored groups of neuroticism. Total 19 SNPs reached the genome-wide statistical significance (<2.21X10^{-7}). They included the interaction of neuroticism with *MIR548AB* for culmen, *PWRN1* and *SMOC2* for cingulate gyrus, *FBXO17*, *ZNF20* and *ZNF136* for inferior frontal gyrus, *METTL22* for inferior frontal gyrus, *RAPGEF3*, *VPS53* and *RPAP3* for middle frontal gyrus, *CTTNBP2NL* for middle occipital gyrus, *CTNNA2* for posterior cingulate, and *KCNK2* and *SASH1* for precentral gyrus. Most of these genes are known to be implicated in behavior and psychiatric disorders. We believe that our analysis is a pilot study that fist describes the gene-personality interaction in brain structure at genome-wide approaches. Our findings suggest that genes with effects on behavior and mental disorders may be involved in neuroticism-associated brain structure. [This research was supported by the National Research Foundation of Korea (NRF) funded by ICT & Future Planning (NRF-2014R1A2A2A04006291) and the Ministry of Education (NRF-2013R1A1A2062702), and Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare (HI14C0072).].

1583T

A new world of biobanking: CARTaGENE, an open-access resource for genomic epidemiology studies. *J. Tcherkezian, C. Boileau, S. Jacquemont, A.M. Nuyt.* CARTaGENE, CHU Sainte-Justine, Montréal, Québec, Canada.

Genome-based studies have provided powerful research tools for identifying genetic variants that contribute to chronic diseases. Recognition is growing, however, that chronic diseases are caused by a combination of an individual's genetic predisposition and their exposure to certain environmental risk factors. CARTaGENE was created to support the scientific community in identifying the determinants of chronic diseases of environmental and/or genetic origin. It was also created to accelerate the process of translational medicine through the identification of biomarkers for early diagnosis, disease treatment and prevention. CARTaGENE is conducting the largest prospective health study on men and women in Québec. Since 2007, the project has recruited 42 000 individuals, aged 40-69 years, representing six metropolitan regions of the province and collected detailed lifestyle, health and medical data on these individuals. The program also gathered detailed physical measurements, clinical and biochemical measures at baseline. Epigenomic, genotyping, and exome sequencing have been collected in subsets of CARTaGENE participants and have resulted in both preliminary data and high profile studies, while supporting over 40 research programs. Several of these projects offer unique opportunities for genomic epidemiology studies and some will be highlighted. These ongoing partnerships, our interoperability with other national biobanks, and our plan to develop other sectors of activity support our mission to provide meaningful high-quality data and biospecimens for years to come.

1584F

Comprehensive evaluation of disease- and trait-specific enrichment for eight functional elements among GWAS-identified variants. *C.A. Markunas, E.O. Johnson, D.B. Hancock.* RTI International, Research Triangle Park, NC, USA.

The majority of disease-associated variants identified from genome-wide association studies (GWAS) are present in functional elements. Studies of functional enrichment have focused on the NHGRI-EBI Catalog of published GWAS (GWAS Catalog) as a whole, or on a small number of functional elements across specific diseases. We sought to understand the relative importance of different functional elements across human diseases and traits, which may provide insight into biological mechanisms that disproportionately contribute to different diseases. We examined variability in functional enrichment of disease-associated variants (P≤5×10⁻⁸) in the GWAS catalog (v1.0.1) across 17 diseases/traits, including neurological and psychiatric disorders, cancers, autoimmune diseases, cardiovascular disease, metabolic disorders, eye disease, lung diseases and traits, viral diseases, drug response, and height. SNPs were annotated using HaploReg (v.4.1) for 8 functional elements: DNase I hypersensitivity sites, expression quantitative trait loci (eQTL), sequence conservation, enhancer marks (H3K4me1, H3K27ac), promoter marks (H3K9ac, H3K4me3), missense variants, sequence motifs, and protein binding sites. Functional enrichment analyses were conducted using logistic regression. Each disease/trait SNP set was compared to a set of 3,587 SNPs that were randomly drawn using SNPsnap, matching our collective set of GWAS SNPs (N=1,294) on allele frequency, gene density, distance to nearest gene, and linkage disequilibrium at a 3:1 ratio. After applying a conservative Bonferroni correction for 136 tests (p<0.0004), significant enrichment was observed for all functional annotations, except sequence motifs. The strongest enrichment signals were seen for missense variants (eye disease, OR [95% confidence interval] = 53.6 [21.1-136.4]), followed by eQTLs (autoimmune diseases, OR = 6.4 [4.9-8.2]). Significant differences in the relative enrichment of functional annotations were observed across disease categories (e.g. eQTL enrichment for autoimmune diseases OR = 6.4 [4.9-8.2] vs. cancer OR = 2.3 [1.7-3.0]) but were limited among diseases within categories. Understanding these commonalities and differences may help to generate hypotheses on common and distinctive mechanisms underlying variant-disease associations, guide the interpretation of GWAS results, and increase power to detect disease variants by further utilizing specific types of regulatory information.

1585W

***GPR161* variants identified in infants with neural tube defects.** *Y. Lei[1], E. Ross[2], G. Shaw[3], R. Finnell[1].* 1) University of Texas at Austin, Austin, TX; 2) Center for Neurogenetics, Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY; 3) Department of Pediatrics, Division of Neonatology, Stanford University School of Medicine, Stanford, CA.

Neural tube defects (NTDs) result from failure of neural fold fusion along the length of the neural tube during early embryogenesis. These are among the most severe of any viable human malformations. The most frequent NTDs are spina bifida and anencephaly. The etiologies of NTDs are complex and include both genetic and environmental factors. Variants in over 400 genes have been identified in the mouse that result in NTD phenotypes. *GPR161*, which encodes a primary cilia protein, has recently been identified as a Tulp3/IFT-A-regulated GPCR. Homozygous *Gpr161* mutant mice are believed to have NTDs due to Shh signaling dysregulation. In a whole genome sequencing study of 125 spina bifida cases, we identified what was predicted to be loss of function variant (p.Pro18Ter) in 7 NTD cases, but the LOF variant was not observed in 158 matched nonmalformed controls. To validate the association between *GPR161* variants and NTDs, we resequenced a validation population of 384 spina bifida cases and 190 nonmalformed controls. Six rare missense variants were identified in 6 NTD cases (MAF<0.01), four of which did not appear in the non-malformed controls studied, or in dbSNP, EVS or ExAC (http://exac.broadinstitute.org). Of those four novel rare variants in NTD cases, two (p.Trp202Gly and p.Leu428Phe) were not found in any database and were predicted to be damaging by SIFT and/or PolyPhen analysis. A third probably damaging rare variant was found in an additional NTD case (MAF=0.003 in ExAC). Among 190 control, 6 of 7 rare *GPR161* variants identified had been previously reported in the EVS or ExAC databases, that were either synonymous or predicted to be benign by both SIFT and Poly-Phen. The frequency of rare (MAF<0.01) missense *GPR161* variants was significantly higher in NTDs (6/384) compared with ExAC data plus our own controls (472/60,898) (p<0.01). Our data indicate that damaging missense GPR161 variants could contribute to the etiology of spina bifida.

1586T

Ultra-rare disruptive mutations in highly constrained genes influence the cognitive and developmental spectrum in the general population.

A. Ganna[1,2,3], K. Satterstrom[1,2], S.M. Zekavat[2,4,5], G. Genovese[1,2,6], A. Byrnes[1,2], M. Artomov[1,2], M. Kurki[1,2,7], C.W. Whelan[2,6], M. Rivas[1,2], P.B. Mortensen[8], S. Kathiresan[2,4], M.J. Daly[1,2], J. Flannick[2], P.F. Sullivan[3,9], A. Palotie[1,2,7], T. Esko[10], C. Hultman[3], B.M. Neale[1,2], MIgen Consortium, iPsych Consortium, AMP T2D Consortium, IBD Consortium, STARR cancer Consortium. 1) ATGU, Massachusetts General Hospital; Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; 6) Department of Genetics, Harvard Medical School, Boston, MA , USA; 7) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland; 8) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 9) Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA; 10) Estonian Genome Center, University of Tartu, Tartu, Estonia.

Large-scale sequencing studies allow to investigate the role of rare exonic variants in explaining observed phenotypic variation. We explored the impact of a subset of rare variants that we hypothesized would be more likely to drive phenotypic variation in the general population and differences in disease risk in case/control cohorts. Specifically, we considered only variants that were (1) singletons in the study and not observed in the Exome Aggregation Consortium, (2) in evolutionarily constrained genes and (3) predicted to modify the gene by either truncating the protein or causing non-sense mediated decay. These damaging ultra-rare variants in highly constrained genes (dURV-HCs) have been shown to be enriched in severe neurodevelopmental disorders, but the extent to which this class of rare variation explains the genetic risk for psychiatric disorders, and indeed a broader range of diseases and traits, is unknown. In this analysis, we included over 80,000 European individuals that were part of population-based studies or case-control studies of common diseases, and for which either whole-exome or whole-genome sequencing data was available. We investigated 16 diseases and 10 quantitative traits and observed enrichments of dURV-HC in neurodevelopmental and psychiatric disorders (including a previously unreported association with attention-deficit/hyperactivity disorder) and identified novel associations with HDL-cholesterol, height and years of education (YOE). For YOE specifically, an analysis of 18,329 individuals from four Northern European studies demonstrated that the presence of one or more dURV-HCs was associated with a decrease in YOE (2.7 fewer months for each additional mutation; $P=5\times10^{-7}$) and the effect was stronger in genes enriched for brain expression (6.1 fewer months, $P=6\times10^{-5}$). To place the effect of dURV-HCs into context, we also examined the impact of previously reported genetic influences on YOE, including a polygenic score from common variants, runs of homozygosity and a burden of rare pathogenic copy number variants. We found that the effect of dURV-HC interacted ($P=0.006$) with a common variant polygenic score for YOE and was more pronounced than the estimated effects of runs of homozygosity and pathogenic copy number variation. Our findings suggest that the impact of dURV-HCs is not confined to severe neurodevelopmental disorders, but that these variants influence the cognitive and developmental spectrum in the general population.

1587F

The TMM BirThree Cohort Study: World-first genome cohort study design of birth and three generation.

M. Ishikuro[1], H. Metoki[1,2], T. Obara[1,3], M. Kikuya[1], Y. Sato[1], M. Miyashita[1], C. Yamanaka[1], S. Mizuno[1], A. Hozawa[1], I. Tsuji[1], F. Nagami[1], S. Kure[1], N. Yaegashi[1], S. Kuriyama[1]. 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Tohoku Medical and Pharmaceutical University, Sendai, Miyagi, Japan; 3) Tohoku University Hospital, Sendai, Miyagi, Japan.

The Tohoku Medical Megabank Organization (ToMMo) was established by Tohoku University along with Iwate Medical University after the Great East Japan Earthquake to reconstruct the disaster area with introducing an advanced medical system. The mission of ToMMo is to identify health problems in the aftermath and to facilitate the solution with developing personalized health care and medicine. One of the projects to achieve its mission is "The Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study (the TMM BirThree Cohort Study)". The TMM BirThree Cohort Study is a unique, large population based birth cohort study with genome analysis that recruits pregnant women and their babies as well as their family members, including relatives. The purpose of this cohort study is to identify health status of people in the aftermath of the disaster and transfer them to the medical facilities for receiving essential medical support, and to analyze the origin or causes of disease with their genomic information to establish personalized medicine in the disaster area. Recruitment of the BirThree Cohort Study was started in July, 2013 and would be continued until the end of March, 2017. The study collects blood or saliva, urine, questionnaires and medical or health records from all the participants, and additionally collect breast milk and cord blood from mothers and their babies. Extra physiological measurements are also obtained for participants who are interested in. All participants will be followed up at least every one year by questionnaires and medical or health records, furthermore, the study plans to provide a set of examination as same as baseline examination after five years. The number of the desirable participants is 70,000, and so far on May 18, 2016, 54,766 people excluding fetuses are involved in the study. There are varieties of pedigree types. At present, total of 3,551 trios with babies, 272 trios with siblings, and 1,931 quartet with babies and their siblings were participated. The number of pedigrees with whole three generations including grandparents, parents and babies has been reached 126 sets. Even more, there are 284 big families consists of more than 7 participants including relatives. Those varieties of pedigrees are enable us to elucidate the origin or causes of disease and contribute to build more quality-oriented care.

1588W

Genome wide association studies of transcriptomic endotypes of asthma. *J. Chu[1], X. Yan[1], J. Gomez[1], L. Cohn[1], B. Raby[2], G. Chupp[1].* 1) Section of Pulmonary, Critical Care, and Sleep Medicine and Department of Internal Medicine, Yale University School of Medicine, New Haven, CT; 2) The Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA.

Background: Previously we have identified from gene expression in sputum and blood three transcriptomic endotypes of asthma (TEA) clusters with distinct clinical and physiological characteristics of asthma. These TEA clusters help explaining the pathophysiologic heterogeneity in asthma patients. This study aims to identify genetic variants associated with these TEA clusters. **Methods:** Genome wide association studies for the three TEA clusters were performed using SNP data from Yale Center for Asthma and Airway Diseases (YCAAD) cohort (asthmatic adult population, n=93) and Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE, asthmatic children population, n=728). The significant loci were furthered analyzed for associations with gene expression and clinical phenotypes. **Results:** We have identified 1 locus at genome-wide significance (FDR<0.05) at one cohort (rs1984564, LP-CAT3, previously linked to macrophages inflammatory response), along with many other loci that were below genome wide significant threshold and can be mapped to biologically plausible genes (eg. rs4525262, TRIM22, p=2.9E-05, previously linked to the airway epithelial response to viral infections). There were also 3 genes that map to genetic variants that were nominally significant (p<0.01) for both cohorts (OPRM1, MROH7, KBTBD12). LPCAT3, TRIM22 and OPRM1 loci were also significantly associated with many clinical phenotypes characterized by TEA clusters, including intubation, hospitalization, and atopy (p<1E-04). **Conclusion:** Our study identifies several genetic variants associated with TEA clusters that located in genes that may underlie key characteristics in the TEA clusters. These genes may be contributing to heterogeneity in asthma pathogenesis.

1589T

Tests of association for asthma among > 10,000 individuals of mixed African ancestry. *M. Daya[1], N.M. Rafaels[1], S. Chavan[1], M. Boorgula[1], H.R. Johnston[2], J.G. Wilson[3], L.K. Williams[4,5], C. Rotimi[6], C. Ober[7], D.A. Meyers[8], V.A. Ortega[8], J. Knight-Madden[9], N.N. Hansel[10], H. Watson[11,12], M.U. Faruque[13], G.M. Dunston[13,14], L. Caraballo[15], M.I. Araujo[16], E.G. Burchard[17,18], M. Taub[19], T.H. Beaty[20], I. Ruczinski[19], Z.S. Qin[2], R.A. Mathias[20,10], K. Barnes[1], CAAPA Consortium.* 1) Department of Medicine, University of Colorado, Denver, CO; 2) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 3) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 4) Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, MI; 5) Department of Internal Medicine, Henry Ford Health System, Detroit, MI; 6) Center for Research on Genomics & Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 7) Department of Human Genetics, University of Chicago, Chicago, IL; 8) Center for Human Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 9) Tropical Medicine Research Institute, The University of the West Indies; 10) Department of Medicine, Johns Hopkins University, Baltimore, MD; 11) Faculty of Medical Sciences Cave Hill Campus, The University of the West Indies; 12) Queen Elizabeth Hospital, Queen Elizabeth Hospital, The University of the West Indies; 13) National Human Genome Center, Howard University College of Medicine, Washington, DC; 14) Department of Microbiology, Howard University College of Medicine, Washington, DC; 15) Institute for Immunological Research, Universidad de Cartagena, Cartagena.; 16) Immunology Service, Universidade Federal da Bahia, Salvador, BA; 17) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 18) Department of Medicine, University of California, San Francisco, San Francisco, CA; 19) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD; 20) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD.

Asthma is a complex disease with striking disparities between racial and ethnic groups, which may be partly attributable to genetic factors. The first asthma genome-wide association study (GWAS), revealing association between asthma and markers near the *ORMDL3* gene on chromosome 17q21 among European populations, has been widely replicated in several European, Asian and Hispanic cohorts. However, SNPs that significantly associated risk to asthma in these populations showed no evidence of association in several independent African ancestry cohorts, although SNPs elsewhere in the gene cluster were modestly associated, underscoring the challenges in replicating SNP-for-SNP findings from GWAS in European populations. One of the main goals of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to discover genes conferring risk to asthma in populations of African descent. In this study, we aimed to increase statistical power to detect genetic variants by performing a meta-analysis across CAAPA cohorts (using a total of 2,281 cases and 3,704 controls), genotyped on the African Diaspora Power Chip (ADPC) and imputed using the whole genome sequence reference panel from CAAPA itself. The ADPC is a SNP array consisting of tagging SNPs useful in comprehensively identifying African specific genetic variation when combined with a commercially available GWAS chip, and the 883 high coverage (~30X) CAAPA genomes provide a much better representation of African haplotypes compared to the standard 1000 Genomes panel. A large proportion of ADPC SNPs have been integrated into Illumina's Multi-Ethnic Genotyping Array (MEGA), and an additional 2000 cases and 2302 controls will be genotyped on this MEGA chip, which will further boost statistical power. We present our quality control, imputation and analysis pipeline, and showcase preliminary results on chromosome 17q21. Within CAAPA, the top 17q21 association signals were observed in a Puerto Rican cohort, consistent with previous reports, whereas signals in two of the African ancestry cohorts analyzed thus far were not suggestive of association. We show this can be explained by the relatively low European ancestry in our African ancestry cohorts compared to the Puerto Rican cohort, resulting in a decreased frequency of putative protective variants. In conclusion, we demonstrate the importance of considering specific ethnic background when conducting genetic association studies of asthma.

1590F

Genome-wide multivariate analyses of anthropometric traits in over 120,000 individuals. *A.W. Drong[1], T. Karaderi[1], P.K. Albers[1], R. Magi[2], A.P. Morris[1,3], C.M. Lindgren[1,4].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 3) Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK; 4) Big Data Institute, University of Oxford, Oxford OX3 7BN, UK.

Recent studies suggested that adjusting for heritable covariates correlated with an outcome can introduce collider bias in genome-wide association studies (GWAS). Here, we apply multivariate models to examine the effect of correlation between phenotypes on the power to detect genetic associations in GWAS of anthropometric traits. We simulated 1,000 genotypes for 10,000 individuals, with effects for two traits (A/B) ranging from β=-0.03 to β=+0.03. Residual variation was then added allowing for non-genetic correlation of $\rho(A,B)$=0.5 or -0.5. The simulations results show good control of type I error rates at in a reverse regression multivariate model, while confirming a bias of $\rho(A,B)*\beta_B$ in β_A in traditional linear regression adjusted analyses. We filtered UK Biobank Interim I variants from European ancestry samples for minor allele frequency (MAF)>0.5% and info score>0.4, resulting in ~12M variants in up to 120,288 individuals. Height, weight, waist, hip, body mass index (BMI), waist-hip ratio (WHR), Body Fat% and Trunk Fat% were adjusted for the age, 15 principal components to adjust for population structure, assessment centre (batch effects) and array type. We fitted multivariate models using reverse regression and traditional univariate linear models for M1: height+weight+waist+hip; M2: M1+BMI+WHR; M3: BMI+WHR+BodyFat%+TrunkFat%. Of 49 previously-reported WHRadjBMI SNPs, we replicate 38 signals using M1 alone, with no improvement in adding BMI and WHR (M2). When selecting the best combination of variables by Bayesian Information Criterion (BIC) selection, the BMI+WHR model is enriched (19/49, P<1x10^{-9}). When adding the choice of BodyFat% and TrunkFat% (M3), for 11 out of these 19 SNPs, WHR+TrunkFat% is the best model for (P<3x10^{-9}), highlighting that these WHRadjBMI loci are indeed primarily associated with central adiposity For M1, we find 13 novel multi-trait associations at genome-wide significance (p<5x10-8), with strongest signals at loci near *LINC01214* (p=7.55x10^{-11}), *MCHR2-AS1* (p=4.83x10^{-10}), 9 loci for M2 (a subset of the 13 M1 loci)*;* and 11 for M3: *ADAMTS14* (p=1.89x10^{-11}), *DNM1P41* (p=6.52x10^{-10}) and *EEF1G* (p=5.76x10^{-10}). None of these signals were detected in our single-trait analyses in UK Biobank. In conclusion, we find that the 49 loci published by Shungin et al are unlikely to be false positives caused by collider bias. For future studies we recommend multivariate analyses as an alternative to adjusted analyses of multiple heritable traits.

1591W

The SHARE study: Meta-analysis of GWAS including 260,000 individuals identifies 40 new loci for allergic disease. *M. Ferreira[1], J. Esparza-Gordillo[2], A. Tillander[3], C. Tian[4], B. Brumpton[5], C. Almqvist[3], D. Hinds[4], D. Boomsma[6], G. Koppelman[7], L. Yi[3], G. Willemsen[6], V. Ullemar[3], P. Magnusson[3], H. Baurecht[8], I. Marenholz[2], J. Vonk[9], L. Paternoster[10], R. Karlsson[3], S. Grosche[3], S. Weidinger[8], Y.A. Lee[2] on behalf of the SHARE study.* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) Max Delbrück Center for Molecular Medicine, Berlin, Germany; 3) Karolinska Institute, Stockholm, Sweden; 4) 23andMe Inc, Mountain View, California, USA; 5) Norwegian University of Science and Technology, Trondheim, Norway; 6) VU University Amsterdam, Amsterdam, Netherlands; 7) University of Groningen, Groningen, Netherlands; 8) University of Kiel, Kiel, Germany; 9) University Medical Center Groningen, Groningen, Netherlands; 10) University of Bristol, England, UK.

Asthma, hay fever and eczema are common, highly polygenic allergic diseases that are thought to share ~50% of their genetic make-up. Despite this large genetic overlap, which translates into hundreds of shared genetic risk variants, almost all GWAS of allergic disease published to date have only considered one disease at a time. A consequence of this is that the control group for a given disease will include a large number of cases from a genetically-correlated disease, which considerably reduces power. Motivated by this limitation, we established the SHARE study to identify genetic risk variants shared between asthma, hay fever and eczema. In this study, we first used the LD-score regression approach to estimate the genetic correlation between the three diseases; estimates were consistent with those reported by twin and family studies: 0.45 between asthma and eczema (P=10^{-20}), 0.69 between asthma and hay fever (P=10^{-188}), and 0.55 between eczema and hay fever (P=10^{-60}). We then performed a GWAS in 13 studies with cases defined as individuals suffering from one or more of these three diseases, and controls as those who have never suffered from any allergic disease. GWAS results were available for 260,000 individuals of European descent. Meta-analysis of these results identified 166 independent (linkage disequilibrium [LD] r<0.1) variants located in 83 loci, of which 40 have not been previously implicated in allergic disease. For 21 of these novel loci, the sentinel SNP was in high LD (r>0.8) with one or more regulatory variants identified in eQTL studies conducted in 12 cell types/tissues relevant to asthma, thereby pointing to likely candidate genes underlying the association. We complemented this list of putative target genes by intersecting the location of risk SNPs with putative enhancers known to be involved in chromatin interactions with gene promoters. We also developed a novel gene-based test of association that focus on regulatory variants and identified additional risk genes missed by the single-SNP analysis. Collectively, our study provides the most comprehensive analysis of genetic risk variants for allergic disease reported to date, and demonstrates the improvement in power that can be obtained by considering information from all three allergic diseases in a GWAS.

1592T

A novel gene identified for alcohol drinking in more than 100K Europeans. *H. Gao[1], C. Liu[2,3,4], P. Song[5], S. Kliewer[5,6], D. Mangelsdorf[6], C.P. Müller[7], G. Schumann[8], E. Elliott[1], AlcGen and CHARGE+ consortia.* 1) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperical College London, UK; 2) The Framingham Heart Study, Framingham, MA, USA; 3) The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; 4) Boston University School of Public Health, Boston, MA, USA; 5) Department of Pharmacology and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA; 6) Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA; 7) Department of Psychiatry and Psychotherapy, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 8) Medical Research Council - Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK.

Excessive alcohol consumption is a major public health problem worldwide. We conducted a genome-wide association meta-analysis and replication study among 105,898 individuals of European ancestry, and identify a novel locus associated with alcohol consumption calculated as log grams/day (SNP with smallest P=9.2x10^{-12}). We show genotype-dependent alcohol preference in this gene in brain-specific knock-out mice compared with controls. Our experiments in mice suggest that effect of alcohol is due to reinforcing behavior, not alleviation of anxiety. Whereas most previous studies investigating the mechanisms underlying alcohol drinking behavior have focused on investigations of brain functions alone, our results suggest the possibility of a coordinated action across two organ systems, liver and brain. The newly identified gene may play a part in the regulation of complex adaptive behaviors involving alcohol drinking.

1593F

GWAS of over 20,000 Generation Scotland: Scottish Family Health Study participants: Expansion of quantitative trait availability using linkage to electronic health records in Scotland. *C. Hayward, MRC QTL in Health and Disease group and Generation Scotland.* Human Geneics Unit, MRC Human Genetics Unit,IGMM, University of Edinburgh, Edinburgh, United Kingdom.

We are using a combination of genome-wide genotyping, sequencing and imputation in population cohorts with high kinship to search for quantitative trait loci (QTL) variants of clinical and biological relevance. The Generation Scotland Scottish Family Health Study (GS:SFHS) is a family-based cohort with biological samples, socio-demographic, clinical and genetic data from approximately 24,000 adult volunteers across Scotland (www.generationscotland.org). Although data collection was cross-sectional, GS:SFHS becomes a prospective cohort as a result of the ability to link to routine NHS electronic health record (EHR) data. GS:SFHS participants were analysed by genome-wide chip genotyping using the Illumina OMNI+exome chip. QC analyses were performed, data cleaned using quality scores and proportions typed, sample identity verified against recorded gender and pedigree, and data checked for unknown relationships based on estimated identity-by-descent. Genome-wide association studies (GWAS) were run on 20,032 individuals, correcting for family relationships using a polygenic kinship matrix and for population stratification using the first three principal components. Data was imputed to the Haplotype Reference Consortium dataset allowing accurate imputation of rare variants into all the genotyped samples. A number of separate programmes of activity are using this genetic data to investigate phenotypes relevant to pain, lung function, psychiatric disorders, cognition, cardiovascular and kidney disease. Here, we present GWAS of a range of biochemical traits obtained from data linkage to EHRs collected by the NHS in Scotland. Scotland has some of the best health service data in the world. Few other countries have information which combines high quality data, consistency, national coverage and the ability to link data to allow patient-based analysis and follow up. Proof of principle was achieved using serum uric acid measures on 2,238 participants after data cleaning, records dating from Oct 1988 to August 2015. A GWAS taking into account the sex of the participant and adjusting for age at the time of the test resulted in a top hit in the urate transporter SLC2A9 gene ($-\log_{10}$(p) value of 7.2 x 10^{-17}). The GWAS of uric acid, and other less studied laboratory measures, will be repeated using HRC-imputed data to investigate rarer variants and elucidate the genetic architecture of the traits.

1594W

A customized genotype array to investigate the genetic architecture of tuberculosis in Peru. *Y. Luo[1,2], L. Lecca[3], R. Calderon[3], C. Conteras[3], R. Yataco[3], Z.B. Zhang[5], J. Galea[3], B. Moody[8], M. Murray[3,4,5,6,7], S. Raychaudhuri[1,2].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Socios En Salud, Lima, Peru; 4) Partners In Health, Boston, MA, USA; 5) Division of Global Health Equity, Brigham and Women's Hospital, Boston, MA, USA; 6) Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA; 7) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 8) Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

As a part of the Tuberculosis Research Unit program (TBRU) that aims to integrate basic and clinical research to study tuberculosis (TB) in countries where the disease is endemic, we are conducting the first TB genome-wide association study (GWAS) in a Latin American cohort in Lima, Peru consisting of 1,500 TB cases and 1,500 controls. While previous TB GWA studies have not always controlled for *Mycobacterium tuberculosis* (*Mtb*) infection status, here we will leverage a well-described cohort of cases and *Mtb* infected household contacts, thereby improving the power of the study to identify genetic variations associated with active TB. To optimize the capture of genetic variants in Peruvians, particularly those that might be missed from the current commercial genotype arrays which are often biased towards European populations, we designed a customized Affymetrix Axiom array based on whole-exome sequencing (WES) data from 116 Peruvian TB cases and 85 Peruvian individuals from the 1000 Genomes Phase 3 project whole-genome sequenced (WGS) at 7x. Our goal in conducting WES in 116 TB cases is to detect rare and low frequency coding variants that are population specific, and might also be predisposing to TB risk. The WGS data will also allow us to ascertain non-coding common variants, particularly those are specific to Peruvian populations only. Using available population genetics tools to infer demographic history both on a global scale and on an individual level, we demonstrate a unique heritage of our Peruvian samples. In total, the WES and WGS data identified 1.5 million variants that were not observed in any other populations included in 1000 Genomes. In particular, 156,162 variants with minor allele frequencies (MAF) ≥1% are private or show significant differences in allele frequencies compared to other populations included in the 1000 Genomes dataset. The customized array includes ~800K variants. After imputing using the 1000 Genomes Phase 3 reference panel, ~8 million high quality (info $r^2 \geq 0.4$) variants with MAF≥1% were detected. Compared to the Affymetrix UK Biobank array, this novel array improves the genome-wide coverage for common variants by 20%, and imputation quality by 35% with a mean imputation $r^2 = 0.901$. This highlights further the value of having a customized array that is tailored for the Peruvian population. The customized array is currently being manufactured and we expect to present the initial findings of the GWAS at the time of the meeting.

1595T

Comprehensive understanding of associations between HLA class II genes and chronic hepatitis B infection in Japanese individuals. *N. Nishida[1,2], J. Ohashi[3], S-S. Khor[2], M. Sugiyama[1], H. Sawai[2], K. Tokunaga[2], M. Mizokami[1].* 1) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan.

Associations of variants located in the *HLA* class II region with chronic hepatitis B (CHB) infection have been identified in Asian populations. Moreover, previous reports revealed that polymorphisms within the *HLA-DP* and *HLA-DQ* loci were independently associated with CHB infection in the Japanese population. Here, *HLA* imputation method was applied to determine *HLA* alleles using genome-wide SNP typing data of 1,975 Japanese individuals (1,033 HBV patients and 942 healthy controls). Together with data of an additional 1,481 Japanese healthy controls, association tests of six *HLA* loci including *HLA-A*, *C*, *B*, *DRB1*, *DQB1*, and *DPB1*, were performed. To assess the association of each *HLA* allele or haplotype with CHB infection, Pearson's chi-square test was applied to a two-by-two contingency table based on the allele or haplotype frequencies. Although the strongest association was detected at a SNP located in the *HLA-DP* locus in a SNP-based GWAS using data from the 1,975 Japanese individuals, HLA genotyping-based analysis identified *DQB1ˑ06:01* as having the strongest association, showing a greater association with CHB susceptibility (OR = 1.76, P = 6.57×10^{-18}) than any one of five *HLA-DPB1* alleles that were previously reported as CHB susceptibility alleles. Moreover, *HLA* haplotype analysis showed that, among the five previously reported *HLA-DPB1* susceptibility and protective alleles, the association of two *DPB1* alleles (*DPB1ˑ09:01*, and *ˑ04:01*) had come from linkage disequilibrium (LD) with *HLA-DR-DQ* haplotypes, *DRB1ˑ15:02-DQB1ˑ06:01* and *DRB1ˑ13:02-DQB1ˑ06:04*, respectively. The present study showed that SNP-based GWAS does not necessarily detect the primary susceptibility locus in the *HLA* region due to the strong LD.

1596F

Association between *STOML2* and allergic diseases. *J.A. Revez[1,2], M.A.R. Ferreira[1].* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) School of Medicine, The University of Queensland, Brisbane, Australia.

A regulatory variant (rs7039317) in the Stomatin-like protein 2 (*STOML2*) gene was recently found to associate with asthma risk in two independent genome-wide association studies (GWAS). However, the evidence for association (combined $P = 2\times10^{-6}$; N = 48,154) was not genome-wide significant. Here, we studied this association in an additional independent GWAS (N=132,860) and examined the possible effect of this and other *STOML2* functional variants in asthma, hay fever and eczema – three allergic diseases that share a large fraction of their genetic make-up. rs7039317 had a significant association with allergic disease in the UK Biobank, with comparable effect sizes across asthma, hay fever and eczema. *STOML2* might represent a genetic risk factor shared between asthma, hay fever and eczema. This research has been conducted using the UK Biobank Resource.

1597W

Genetic ancestry and genome-wide associations with bronchopulmonary dysplasia in preterm infants treated with Inhaled Nitric Oxide (iNO). *D. Torgerson[1], S. Oh[1], R.L. Keller[2], S. Huntsman[1], D. Hu[1], C. Eng[1], D. Black[3], E.G. Burchard[1,4], P.L. Ballard[2], R.A. Ballard[2], the TOLSURF Study Group.* 1) Department of Medicine, University of California, San Francisco, San Francisco, CA; 2) Department of Pediatrics, University of California, San Francisco, San Francisco, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 4) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Bronchopulmonary dysplasia (BPD) is a severe yet common disease in preterm infants. In the Trial of Late Surfactant (TOLSURF) study, in which all infants received iNO, those with maternal self-reported Black/African American race/ethnicity had increased survival without BPD at 36 weeks postmenstrual age as compared to Non-Hispanic White. However, when untreated no difference between racial/ethnic groups has been observed in prior studies, suggesting there may be population-specific genetic contributions to respiratory outcomes following iNO treatment. We performed ancestry analyses and a genome-wide association study for survival without BPD at 36 weeks postmenstrual age on 387 preterm infants from mothers of three racial/ethnic groups. Association testing using genome-wide SNP genotypes and local ancestry were performed within each racial/ethnic group using logistic regression, and combined in a meta-analysis. All analyses were adjusted for birth weight, gestational age, sex, multiple gestation, and global ancestry. Gene-based statistics were calculated using VEGAS, and pathway analyses performed using GREAT and IPA. Genomic African ancestry was associated with increased survival without BPD within infants of maternal self-reported Hispanic White race/ethnicity (OR=4.5, p=0.01), but not within infants of Black/African American race/ethnicity (OR=1.01, p=0.97). Admixture mapping found suggestive associations with local African ancestry at 18q21 and 10q22 and survival without BPD, and our top associated individual SNPs were within the intron of *NBL1* (meta-analysis $p=7.4\times10^{-7}$). Variants in NO pathway genes that differed in frequency between African Americans and non-Hispanic Whites were more likely to be associated with survival without BPD as compared to variants at similar frequencies ($p=1.6\times10^{-4}$), indicating a potential genetic contribution to differential response to iNO treatment between racial/ethnic groups. Furthermore, integration of genes upregulated in BPD lungs found a significant association with variation in *CCL18*. Pathway analyses implicated sets of genes that are transcriptionally regulated by histone methylation, and those involved in immune and inflammatory processes in response to infection and mechanical ventilation. Overall our results suggest that genetic variation contributes to individual differences in survival without BPD, and may contribute to racial/ethnic differences in respiratory outcomes following iNO treatment.

1598T

A genome-wide association study for gut microbiota populations. *Y. Yun[1], H. Kim[1], S. Kim[1], Y. Chang[2], S. Ryu[2], H. Shin[3], H. Kim[1].* 1) Department of Biochemistry, Ewha Medical Research Institute, School of Medicine, Ewha Womans University, Seoul, South Korea; 2) Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea; 3) Department of Family Medicine and Health Screening Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea.

The human gut microbiota has been studied extensively due to its vital functions that affect nutritional efficiency and overall health. However, it is still no clear how the host genome gives a selective pressure to build certain microbial populations. We analyzed the genome-wide association (GWA) to bacteria harboring in human intestine using GWA datasets in a cross-section of 873 adults (45.11±9.22 years of mean age, 535 male/338 female) having metagenomic microbial data from fecal samples. We selected two microbes existing in samples, 'Christensenellaceae' which is reported to be the most heritable especially in lean population and '*Akkermansia*' which is supposed to be a beneficial mucin-utilizing microbe in relation with host immunity. In the discovery phase, we identified variants of *AQP9* (aquaporin 9) that associated with Christensenellaceae colonization (P=1.29×10^{-7}). This gene encodes the protein that can facilitate the uptake of glycerol in hepatic tissue and also play a role in specialized leukocyte functions such as immunological responses and bactericidal activity. Although *Akkermansia* had no significant association with host gene, the pathway involved with beta subunit of IL-2 receptor was significantly associated with *Akkermansia* population (P<0.001, FDR=0.083). This subunit of receptor is essential for IL-2 signaling involved in T cell differentiation and proliferation, which activity is crucial to the regulation of the immune response. These findings will pave the way to understand the host genetic influences on gut microbial composition that codeveloped with the host from birth. [This research was supported by the National Research Foundation of Korea (NRF) funded by ICT & Future Planning (NRF-2014R1A2A2A04006291) and the Ministry of Education (NRF-2013R1A1A2062702), Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare (HI14C0072), and Intramural research support program of Ewha Womans University School of Medicine.].

1599F

Genes for good: Engaging the public in genetics research using social media. *A. Pandit[1], W. Li[1], E.M. Schmidt[1], G.J.M. Zajac[1], C.P. Clark[1], K. Brieger[1], L.G. Fritsche[1,2], J.R. Forster[1,2], G.R. Abecasis[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway.

Genes for Good is a new kind of genetic study which aims to engage 10,000s of individuals in genetic research through social media, with the goal of understanding the genetic basis of human health and disease. Participants sign up through a Facebook app and answer health-related questionnaires, including one-time health histories and longitudinal health tracking surveys that can be completed as often as once per day. After answering a minimum number of surveys, participants are eligible to request a spit kit and send in their saliva sample for genotyping. After processing, participants can access ancestry information through the app, compare the survey answers to aggregate summaries of other participants, and download their raw genetic information. We are developing new ways to visualize genetic results in order to provide participants with meaningful, interesting, and accessible insights into their own genomes.As of May 2016, >11,500 individuals from all 50 U.S. states have enrolled. Growth has steadily increased since the start of the study primarily through word of mouth, and being featured on local news and web articles has further increased participation. We observe a higher percentage of female participants (over 65% of our participants are female) and also a shortage of older participants (for example, 10% of participants are >61 years old versus 18% in the US Census).Preliminary analyses demonstrate that we can reproduce population-level associations between health traits (for example, replicating associations between diabetes and obesity or between smoking and cardiovascular disease) and also known genetic association signals (for traits ranging from skin color to obesity).Genes for Good is a collaborative study, and we encourage researchers to contact us about potential collaborations.

1600W

Pharmacogenomics of ritodrine as a tocolytic medication in preterm labor. *K. Lee[1], J. Park[2], Y. Kim[3], H. Hwang[4], H. Gwak[2].* 1) Chungbuk National University, Cheongju, South Korea; 2) Ewha Womans University, Seoul, South Korea; 3) Ewha Womans University School of Medicine, Seoul, Korea; 4) Konkuk University School of Medicine, Seoul, Korea.

Background Preterm labor is the leading cause of neonatal mortality and morbidity around the world. The beta-2 agonist ritodrine has been used as a tocolytic agent due to its mechanism to cause myometrial relaxation. The purpose of this study was to evaluate the association of genetic polymorphisms with therapeutic outcomes of ritodrine. Methods Forty-one single nucleotide polymorphisms of 10 genes were genotyped in 216 preterm labor patients: *ADRB2, PDE4D, PDE4B2, GRK5, GRK6, ARRB2, RGS2, RGS5, GNAS, and CACNA1C.* Baseline characteristics and blood samples of patients were collected for analyses. The primary endpoint was time to delivery (hours) or therapeutic failure defined by using alternative tocolytics. Results Among demographics of patients, gestational age at admission, modified Bishop score, and placental abnormality showed statistical significance with time to delivery or therapeutic failure in multivariate analysis. *ADRB2* rs1042719 (adjusted hazard ratio (AHR) 2.7, 95% CI 1.5–5.0), *PDE4D* rs983280 (AHR 2.1, 95% CI 1.2–3.7), *ADE4B2* rs2180335 (AHR 0.3, 95% CI 0.1-0.9), *RGS2* rs1152746 (AHR 0.1, 95 CI% 0.0-0.7), and *RGS5* rs4657251 (AHR 0.5, 95% CI 0.3-0.9) showed significant associations with time to delivery or therapeutic failure. Conclusion Our findings demonstrate the impact of mutations of pharmacodynamic genes of ritodrine on its pharmacologic response in preterm labor.

1601T

Genome-wide scans identify novel genetic variants on *SOX2*, *FOXD1* and *EDAR* influencing eyebrow thickness. *S. Wu[1], M. Zhang[1,2], J. Tan[2], Y. Yang[2,3], Q. Peng[1], J. Li[1], Y. Liu[1], H. Lou[1], D. Lu[1], Y. Gun[4], K. Adhikari[5], A. Ruiz-Linares[5], K. Tang[1,2], L. Jin[1,2], S. Xu[1,2,6], S. Wang[1,2].* 1) CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Shanghai, China; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China; 3) Fudan-Taizhou Institute of Health Sciences, Taizhou, Jiangsu 225300, China;; 4) Department of Biochemistry, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, China; 5) Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London WC1E 6BT, UK; 6) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China;.

The eyebrow is an area of thick, delicate hairs above the eye that follows the shape of the lower margin of the brow ridges of some mammals. It was thought to plays a main function to prevent moisture, mostly sweat, rain and other debris, from falling down into the eye. It is also important to facial expression, human communication and face recognition. A recent genome-wide scan found common variants in the Forkhead Box L2 (*FOXL2*) gene that are associated with eyebrow thickness in Latin Americans, but different genes can be responsible for eyebrow thickness in other populations. We firstly performed a genome-wide scan for eyebrow thickness in 2,961 Han Chinese. Two novel loci were found to be significantly associated with thicker eyebrow: rs1345417 at 3q26.3 ($P=6.51\times10^{-10}$; $\beta=0.092\pm0.015$) near *SOX2*, a gene reported to specify hair follicle type and control hair growth; rs12651896 at 5q13.2 ($P=1.73\times10^{-8}$; $\beta=0.085\pm0.015$) near *FOXD1*, a gene that was enriched in dermal condensate cell. Then we validated the findings by performing a second genome-wide scan in 721 Uyghurs (rs1345417: $P=3.78\times10^{-4}$; $\beta=0.105\pm0.029$; rs12651896: $P=0.042$; $\beta=0.061\pm0.030$). Further, we found our findings can be replicated in published data of Latin Americans (rs1345417: $P=1.04\times10^{-7}$; $\beta=0.98\pm0.018$; rs12651896: $P=7.44\times10^{-6}$; $\beta=0.080\pm0.018$). However, the previously reported signal at *FOXL2* in the Latin Americans do not affect eyebrow thickness in Han Chinese ($P=0.730$) and Uyghur ($P=0.494$), suggesting distinctive mechanisms are affecting eyebrow thickness in different populations. Finally we performed a trans-ethnic meta-analysis with 2,961 Han Chinese, 721 Uyghurs, and 2,421 Latin Americans. One additional novel loci were found to be significantly with thicker eyebrow: rs1866188 at 2q12.3 ($P=4.96\times10^{-11}$; $\beta=0.099\pm0.025$) near *EDAR*, a gene that plays an important role in the development of ectodermal tissues such as the hair, teeth and other ectodermal derivatives. All of the above signals locate at regulatory region that exhibits distinct active enhancer signatures, suggesting an important way to affect the phenotype by regulating the expression of genes. These findings further expand our knowledge on the genetics of eyebrow thickness. .

1602F

Homocysteine and IgA nephropathy: A Mendelian randomization approach. *Y. Zhang[1], X. Zhou[1], B. Li[2], F. Yu[2], S. Shi[1], L. Liu[1], J. Lv[1], H. Zhang[1].* 1) renal department, peking university first hospital, Beijing, China; 2) School of International Trade and Economics, Central University of Finance and Economics, Beijing, China.

Objectives: Hyperhomocysteinemia (HHcy) occurs almost uniformly in patients with end stage renal disease (ESRD). IgA nephropathy (IgAN) is the most common form of immune related glomerulonephritis and cause of ESRD. However, whether the elevated plasma Hcy is related to IgAN is still unclear. Due to confounding factors, we used a novel analysis tool, namely 'Mendelian randomisation (MR) approach' for testing whether elevated plasma Hcy is causally related to IgAN. **Methods:** The study comprised 1686 patients with sporadic IgAN and 1442 healthy controls. Serum Hcy was measured and patients with minimal change disease (MCD) were enrolled as patient controls. For MR analysis, the missense variant rs1801133 in the gene *MTHFR* with the most consistent effect on plasma Hcy concentrations was selected as the instrument. It was genotyped by TaqMan allele discrimination assays and verified by Sanger sequencing. **Results:** Serum levels of Hcy were significantly elevated in patients with IgAN compared to those in MCD patients (medians: 18.32 umol/l vs 11.15 umol/l, $P = 1.26 \times 10^{-7}$) and healthy controls (medians: 18.32 umol/l vs 10.00 umol/l, $P = 1.96 \times 10^{-18}$). About 93.52% patients with IgAN showed elevated plasma levels of Hcy ($\geq$10umol/l, HHcy). And HHcy was associated with higher Scr levels ($P = 0.07$), lower estimated glomerular filtration rates (eGFRs) ($P = 6.38 \times 10^{-3}$), higher systolic blood pressures (SBPs) ($P = 0.01$) and diastolic blood pressures (DBPs) ($P = 0.07$), and a higher proportion of Oxford pathologic classification of tubular atrophy and interstitial fibrosis (T) ($P = 4.88 \times 10^{-3}$). In further, the risk allele of rs1801133-T was significantly associated with increased serum Hcy level ($P = 3.89 \times 10^{-4}$) and patients with the risk allele tended to have higher Scr levels ($P = 0.03$), higher SBPs ($P = 0.02$) and DBPs ($P=0.01$), and higher ratio of T lesion ($P = 0.01$). For long-term renal survival, patients with the risk allele of rs1801133-T were independently associated with kidney disease progression (log rank test $P = 0.04$). On instrumental variable analysis, genetically elevated Hcy resulted in effects of 0.76 ($P = 0.02$), 0.26 ($P = 0.01$), 0.20 ($P = 0.01$), and 0.011 ($P = 0.01$) for Scr, SBP, DBP, and T in IgAN. **Conclusions:** In a large cohort of IgAN patients, we tested the causality between Hcy and IgAN, which could be of great significance in predicting the pace of disease progression, developing targeted therapeutic strategies in preventing the progression of IgAN.

1603W

A genome-wide association study identifies *SIGLEC5* and the *DEF1A* locus as genetic risk factors of aggressive and chronic periodontitis. *M. Munz[1], G.M. Richter[1], Y. Jockel-Schneider[2], K. Berger[3], P. Hoffmann[4,5], J. Meyle[6], P. Eickholz[7], C. Doerfer[8], C. Bruckmann[9], M. Laudes[10], A. Uitterlinden[11], A. Teumer[12], B. Holtfreter[13], T. Kocher[13], J. Erdmann[14], W. Lieb[15], A. Franke[16], S. Schreiber[10,16], S. Jepsen[17], B.G. Loos[18,19], H. Dommisch[1], A. Schaefer[1].* 1) Department of Periodontology and Synoptic Dentistry, Charité, Berlin, Berlin, Germany; 2) Department of Periodontology, Clinic of Preventive Dentistry and Periodontology, University Medical Center of the Julius-Maximilians-University, Würzburg, Germany; 3) Institute of Epidemiology and Social Medicine, University Münster, Germany; 4) Institute of Human Genetics, University of Bonn, Germany; 5) Human Genomics Research Group, Department of Biomedicine, University Hospital of Basel, Switzerland; 6) Department of Periodontology, University Medical Center Giessen and Marburg, Germany; 7) Department of Periodontology, Centre for Dental, Oral, and Maxillofacial Medicine (Carolinum), Johann Wolfgang Goethe-University, Frankfurt am Main, Germany; 8) Department of Operative Dentistry and Periodontology, University Medical Center Schleswig-Holstein, Campus Kiel, Germany; 9) Department of Conservative Dentistry and Periodontology, Clinic of Dentistry, Bernhard Gottlieb University, Vienna, Austria; 10) Clinic of Internal Medicine I, University Clinic Schleswig-Holstein, Kiel, Germany; 11) Erasmus Medical Centre in Rotterdam, Netherlands; 12) Institute of Community Medicine, University of Greifswald, Germany; 13) Department of Periodontology, University of Greifswald, Germany; 14) Institute of Integrative and Experimental Genomics, University Medical Center Schleswig-Holstein, Campus Lübeck, Medical Clinic II, Lübeck, Germany; 15) Institute of Epidemiology, Christian-Albrechts-University, Kiel, Germany; 16) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 17) Department of Periodontology, Operative and Preventive Dentistry, University of Bonn, Bonn, Germany; 18) Department of Periodontology and Oral Biochemistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam; 19) VU University Amsterdam, Netherlands.

Periodontitis (PD) is a heritable common inflammatory disease with worldwide prevalence rates of 11% for the severe forms. It is the major cause of tooth loss in adults above 40 years and is characterized by destruction of the alveolar bone, resulting from the inflammatory process. To better understand the molecular mechanisms of PD, we performed a genome-wide association study in the worldwide largest sample of the severe early-onset phenotype aggressive periodontitis (AgP). A case-control sample (N=896 cases, N=7,090 controls) of German and Dutch descent was genotyped using OmniExpress BeadChips (Illumina). The genotypes were imputed (1000G Phase3 EUR haplotype reference), followed by a separate analysis of the German (GER) and Dutch (NL) samples (allelic model, adjusted for the covariates sex and smoking). Subsequently, a meta-analysis on the results of both samples was performed (random effects model) with subsequent filtering by p-value (p_{meta}GER, p_{NL}; $p_{meta}<10^{-5}$; p<0.05 for LD SNPs with $r^2 > 0.8$ [linkage disequilibrium] in both imputed and un-imputed data). 16 SNPs were selected for replication in a German case-control sample of the moderate but widespread phenotype chronic periodontitis. Variants at two loci, within the protein coding gene *SIGLEC5,* and at the annotated pseudogene *DEFA9P,* reached genome-wide significance in the pooled samples. Best associated were the GWAS lead SNPs rs4284742, intronic of *SIGLEC5* ($p_{meta}=3.83^{-8}$, odds ratio [OR]=0.75), and rs2738058, 4kb upstream *DEFA9P* (p=1.04^{-9}, OR=1.28). *In silico* functional analysis (Haploreg, GTEx) revealed expression quantitative trait loci effects of these SNPs on the genes *SIGLEC5* (p=7.7^{-14} in whole blood) and *DEFA4* (p=1.7^{-3} in blood), respectively. *SIGLEC5* inhibits the activation of several cell types including monocytes, macrophages and neutrophils. *DEFA9P* and *DEFA4* belong to the family of Defensins, which are abundant in the granules of neutrophils and the epithelia of mucosal surfaces such as the gingiva. They encode antimicrobial peptides and are involved in host defense. This study identifies, for the first time, shared risk loci of AgP and CP with genome-wide significance, and highlights the role of innate and adaptive immunity in the etiology of periodontitis.

1604T

Genome-wide study of parental origin-specific allelic associations with birth weight. *R.N. Beaumont[1], O. Helgeland[2,3], J.P. Bradfield[4], M. Nodzenski[5], X. Liu[6], A.E. Jonsson[7], F. Day[8], V. Karhunen[9,10], M. Horikoshi[11,12], M.N. Kooijman[13,14,15], R.B. Prasad[16], T. Ahluwalia[17], G. Zhang[18,19,20], Z. Kutalik[21,22], D.M. Evans[23,24,25], R.M. Freathy[1,23], Early Growth Genetics Consortium.* 1) Institute of Biomedical and Clinical Science, University of Exeter Medical School, Royal Devon and Exeter Hospital, Exeter, UK; 2) KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 3) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway; 4) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 5) Department of Preventive Medicine, Division of Biostatistics, Feinberg School of Medicine, Northwestern University, Chicago, USA; 6) Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark; 7) Novo Nordisk Foundation Center for Basic Metabolic Research and Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 8) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK; 9) Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland; 10) Biocenter Oulu, University of Oulu, Finland; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 12) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 13) The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, the Netherlands; 14) Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, the Netherlands; 15) Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, the Netherlands; 16) Lund University Diabetes Centre, Department of Clinical Sciences, Lund University CRC, Skåne University Hospital Malmö, SE-205 02, Malmö, Sweden; 17) Copenhagen Prospective Studies on Asthma in Childhood, Faculty of Health and Medical Sciences, University of Copenhagen and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark; 18) Department of Pediatrics, Vanderbilt University School of Medicine and Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, Tennessee, USA; 19) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; 20) Vanderbilt Kennedy Center for Human Development, Vanderbilt University, Nashville, Tennessee, USA; 21) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 22) Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 23) Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK; 24) School of Social and Community Medicine, University of Bristol, Bristol, UK; 25) The University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia.

Rare genomic anomalies at a number of known imprinted loci are associated with disorders of birth weight. However, to date, no common genetic variants have reported robust parent-of-origin associations with birth weight. We analysed 10,915 mother-child pairs and 583 parent-offspring trios from 7 studies to determine parental origin of alleles at 6.1M SNPs genome-wide (MAF >0.01; imputed to 1000 Genomes). We tested the associations between birth weight and (i) maternally-inherited alleles, (ii) paternally-inherited alleles and (iii) maternally- versus paternally-inherited alleles in heterozygotes. We highlight SNPs with associations $P<5\times10^{-8}$ in model (iii) AND $P<0.05$ in either (i) or (ii). For 59 loci identified in recent birth weight GWAS, we were able to include additional samples, resulting in a total N of 18,719 pairs/trios. To complement this approach, we tested for differences in birth weight variance between heterozygotes and homozygotes at 12.2M variants (MAF>0.01) in 103,619 unrelated individuals from 12 studies. Effect sizes on birth weight differed between maternally- and paternally-inherited alleles for SNPs at two loci (near *TENM4* and in *ARHGAP25*; model (iii) $P=3\times10^{-11}$ and $P=2\times10^{-8}$, respectively); for both SNPs only the maternally-inherited allele was associated with birth weight at $P<0.05$: 0.58 SD (95%CI: 0.27, 0.89 SD, $P=2\times10^{-4}$) change in birth weight per maternally-inherited allele near *TENM4* and 0.29 SD (95%CI: 0.08, 0.51 SD; $P=8\times10^{-3}$) in *ARHGAP25*. The paternally-inherited alleles showed no association with birth weight ($P>0.2$). One of the 59 known birth weight loci, *ANK1*, showed association in model (iii) after Bonferroni correction ($P=8\times10^{-4}$): 0.07 SD (95%CI: 0.04, 0.11 SD, $P=8\times10^{-5}$) change in birth weight per paternally-inherited allele; maternally-inherited $P=0.2$. Among unrelated individuals, no variant was associated with variance differences between heterozygotes and homozygotes at $P<5\times10^{-8}$, and there was no difference at either *TENM4* or *ARHGAP25* ($P>0.2$). However, there was some support for greater variance in heterozygotes than homozygotes at *ANK1* ($P=0.002$). Our initial analysis provides evidence of possible parent-of-origin effects on birth weight at 3 loci, but these require replication in further samples. The upcoming addition of further data, including GWAS in 11,000 parent-offspring trios, will greatly enhance our power to detect robust associations.

1605F

Parent-of-origin effect in asthma - GWAS meta-analysis in three Canadian cohorts. *A. Eslami[1], L. Akhabir[1], G. Ellis[1], A.B. Becker[2], A.L. Kozyrskyj[3], P.D. Paré[1], A.J. Sandford[1], C. Laprise[4], D. Daley[1].* 1) University of British Columbia, Vancouver, BC, Canada; 2) Department of Pediatrics and Child Health, Faculty of Medicine, University of Manitoba, Winnipeg, MB, Canada; 3) Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada; 4) Université du Québec à Chicoutimi, Saguenay, QC, Canada.

   Background: Asthma is a complex disease caused by a combination of genetic and environmental factors. Heritability is estimated to range between 0.48-0.79. To date, 44 significantly associated SNPs have been identified by 23 asthma genome-wide association studies (GWAS) (at $p<10^{-6}$). The consensus is that the main genetic effects of these common SNPs (with modest effects) do not fully explain the heritability of asthma. Genomic imprinting is a potential mechanism which may explain some of the 'missing heritability'. Imprinting is an epigenetic phenomenon where the expression of genes depends on their parental origin (parent-of-origin effect). Imprinting effects have been reported in the development of many complex diseases. **Hypothesis**: Imprinting is involved in the etiology of asthma. **Methods:** To identify candidate genomic regions for imprinting we used GWAS data from three family-based studies (two parents and one offspring). These studies are: 1) the Canadian Asthma Primary Prevention Study (CAPPS), a high-risk asthma birth cohort and, 2) the Study of Asthma Genes and Environment (SAGE), a population-based asthma birth cohort 3) the Saguenay-Lac-Saint-Jean Québec Familial Collection (SLSJ), a founder population of French-Canadians. We used a likelihood-based variant of the Transmission Disequilibrium Test. Parent-of-origin effects in SLSJ as well as the combined CAPPS and SAGE were performed by including parental sex as a modifier in the analysis. Meta-analysis was conducted using the results of SLSJ and the joint analysis of CAPPS and SAGE weighted by the number of informative transmissions for each study. **Results**: In SLSJ, 7 SNPs showed significant parent-of-origin effects with $p<10^{-5}$ (252 trios with asthmatic children). In the joint analysis of CAPPS and SAGE, 13 SNPs showed significant parent-of-origin effects with $p<10^{-5}$ (148 trios with asthmatic children). Notably, in the joint analysis of CAPPS and SAGE, we identified a parent-of-origin effect at a known imprinted gene, *CTNNA3*. This gene was previously identified in a GWAS study of occupational asthma. Of all significant results in those two analyses ($p<10^{-5}$), 12 out of 20 of the SNPs were in or near Long non-coding (lnc)RNA genes. LncRNAs are known to be involved in genomic imprinting and gene regulation. Meta-analysis resulted in two SNPs with significant parent-of-origin effects with $p<10^{-5}$ in the genes *SLC39A10* and *LNX2/POLR1D*. .

1606W

Cumulative impact of red blood cell polymorphisms in resistance and susceptibility to severe malaria. *G.M. Clarke[1], G. Band[1], C.M. Ndila[2], C.A.C. Spencer[1], K.A. Rockett[1,3], D.P. Kwiatkowski[1,3].* 1) Wellcome Trust Centre for Human, University of Oxford, Oxford, United Kingdom; 2) Department of Epidemiology and Demography, KEMRI-Wellcome Trust Research Programme, P.O. Box 230 Kilifi 80108, Kenya; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge. CB10 1SA.

   Individually, polymorphisms in *HBB*, *ABO*, *ATP2B4*, *G6PD* and the glycophorin region have been reliably associated with protection from severe malaria due to *Plasmodium falciparum*. All are red blood cell (RBC) polymorphisms. Given the extensive changes in the host red cells brought about by invasion of the malaria parasite, it is natural to ask whether epistasis between these loci could be responsible for variation in malaria susceptibility. Here we assess interaction between combinations of these RBC loci in a case control study of 11,890 cases of severe malaria due to *Plasmodium falciparum* and 17,441 controls from 11 countries in Africa, Asia and Oceania. We find no evidence for any departure from independence of effects, indicating that protection accumulates cumulatively with each protective genotype carried. Cumulative genetic protective scores were calculated using unweighted and weighted approaches. We report on the predictive power of the genetic scores and show that as the genetic protective scores increase, the risk of severe malaria increases: individuals carrying protective genotypes at multiple loci have up to a 20-fold increase in protection compared to those carrying none. .

1607T

Variation in locus 5q31.1 confers resistance to *Mycobacterium tuberculosis* **infection in immunosuppressed individuals from hyper-endemic regions of East Africa.** *R.S. Sobota[1,2], C.M. Stein[3,4], N. Kodaman[1,2,3], S.A. Tishkoff[5], W.K. Scott[6], J.H. Moore[2,7], T. Lahey[2], W.H. Boom[3,4], C.F. von Reyn[2], S.M. Williams[2,3], G. Sirugo[8].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Geisel School of Medicine at Dartmouth, Hanover, NH; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 4) Tuberculosis Research Unit, Case Western Reserve University, Cleveland, OH; 5) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 6) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 7) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 8) Centro di Ricerca, Ospedale San Pietro Fatebenefratelli, Rome, Italy.

One in three people in the world have been infected with *Mycobacterium tuberculosis* (MTB). Individuals immunosuppressed by HIV infection are more likely both to be infected with MTB and to develop subsequent clinical disease. We hypothesized that HIV-positive individuals who live in areas endemic for MTB but do not get infected are genetically resistant. We performed a genome-wide association study of tuberculin skin test positivity using 479 HIV-positive patients in two recently concluded prospective cohorts of tuberculosis from Tanzania and Uganda. In our screening, 244 tested positive for MTB infection either at enrollment or throughout the >8 year follow up, while 235 did not. We identified a genome-wide significant association between the dominant model of rs877356 and binary TST status in the combined cohort (OR = 0.2671, 95% CI 0.1696-0.4207, p = 1.22×10^{-8}). The variant lies in the 5q31.1 region, 57kb downstream from *IL9*. Two-locus analyses of association of variants near rs877356 showed that haplotype comprised of rs877356 and an *IL9* missense variant rs2069885 had the most significant association. *IL9* encodes for a cytokine produced by mast cells and T_H2 cells in inflammatory responses. It has been previously associated with bronchial hyperresponsiveness, a heightened bronchoconstriction and airway inflammation in response to external stimuli. Observational studies have reported decreased prevalence of MTB infection in people with chronic bronchial hyperresponsiveness, as seen in asthma. Our results identify a novel locus of MTB infection resistance, propose a possible link between airway inflammation and protection from MTB infection, and indicate that studying those who remain disease-free in the face of extensive exposure increases the power to detect associations in complex infectious disease.

1608F

Genome-wide association study of multivariate measures of human facial morphology. *M. Lee[1], E.J. Leslie[1], J.C. Carlson[2], E. Orlova[3], E. Feingold[3], M.L. Marazita[1], J.R. Shaffer[3], S.M. Weinberg[1].* 1) Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

A small number of studies have reported genome-wide associations with quantitative aspects of normal human facial morphology. While these studies have each used different phenotyping approaches, significant associations have generally been restricted to univariate measures of facial morphology. Such measures capture only limited information about facial structure. In an effort to expand on previous efforts, we applied factor analysis to 276 measures (linear distances) covering the entire face. These distances were derived from 3D facial images of 2187 healthy White individuals of European ancestry and were adjusted by sex, age and body size. The samples were genotyped on the Illumina OmniExpress+Exome v1.2 array and imputed to the 1000 Genomes reference panel. Genome-wide association analyses were conducted on 23 extracted facial factors in PLINK adjusting for ancestry principal components. We observed a total of six genome-wide significant associations (p<5e-8) with different facial shape factors. Most notably, variants in *PARK2* (top SNP: rs9456748; 6q26) were associated with a factor defined by midfacial height (p=4.99e-8). Furthermore, variants in *FREM1* (top SNP: rs72713618; 9p22.3) were associated with a factor defined by height of the upper lip (p=2.02e-8). Mutations in *FREM1* have been associated with facial dysmorphology in humans and strong expression has been described in mouse embryos at the site of medial nasal process fusion. The role of *PARK2* in craniofacial development is currently unclear, but the gene is expressed in the epithelium of the nasal cavity in mice. These associations have not been reported previously. Our results provide further insight into the genetic underpinnings of normal-range variation in human facial morphology and suggest that multivariate data reduction methods may be useful for defining relevant facial phenotypes. .

1609W

Clinical, social and genomic factors associated with obesity at 12 months of age. *S. Hazrati[1], W. Wong[1], N. Prem[1], Y. Yui[1], K. Huddleston[1], S. Hourigan[1,2], J. Niederhuber[1].* 1) Inova Health System, Falls Church, VA; 2) Johns Hopkins University.

Purpose: To examine genomic, social and clinical predictors of obesity at 12 months of age Methods: Over 3000 families from various races or ethnicities have been recruited in prenatal stage, in the longitudinal study of genomics and child health, at Inova Translational Medicine Institute. 760 children had clinical and genomic data available at age 12 months. Whole genome sequencing was performed on blood. Weight for length at 12 months was calculated using WHO gender specific growth charts with the following definitions: overweight as greater or equal to 85th, obese as greater or equal to 95th, severely obese as greater or equal to 99th percentiles. A supervised genetic analysis was conducted using the SKAT for the association of rare variants (MAF less than or equal to 0.1) with obesity. Gene-set level p-values were computed for 363 obesity related genes from the human obesity gene map. Chi-square and paired t-test were used to test for association of clinical and social factors with obesity. Results: Of the 760 infants, 15% were overweight, 15% obese and 8% severely obese. Clinical or social factors significantly associated with being overweight, obese or severely obese at 12 months were ethnicity, maternal education, juice consumption and parental BMI (p -value less than 0.05). After adjusting for multiple testing with Bonferroni correction, only two genes were significant at the 0.1 level, namely, WT1 (P=0.0033) and CNR1 (P=0.090), for the severely obese group. Conclusion: Many genomic, clinical and social risk factors are associated with obesity. None of the genes were significant after Bonferroni correction for the overweight or obese groups. Clinical and socials factors that were significantly associated (p -value less than 0.05) with being overweight and obese in all 3 groups were Hispanic origin, lower maternal education and any juice consumption at 12 months.

1610T

Comparison of collection methods and whole genome amplification for pharmacogenetics. *J.S. Ang[1], H. Katzov-Eckert[1], M. Aloise[1], D. Dawes[1,2], A. Gonzalez-Suarez[1,3], A. Paterson[1,4,5], P. Stanley[1], R. Fraser[1,6], P. Cullis[1,6], M. Dawes[1,7].* 1) GenXys Health Care Systems, Vancouver, BC; 2) University of British Columbia Physical Therapy, Vancouver, BC; 3) University of British Columbia Department of Botany, Vancouver, BC; 4) University of British Columbia Faculty of Pharmaceutical Sciences, Vancouver, BC; 5) Clinicare Pharmacists Inc, Vancouver, BC; 6) University of British Columbia Personalized Medicine Initiative, Vancouver, BC; 7) University of British Columbia Family Practice, Vancouver, BC.

A simple, non-invasive sample collection method is key to the uptake and integration of pharmacogenetic testing by clinicians. We tested two different types of swabs as an alternative to saliva collection, which was used in the Implementation of Pharmacogenomics in Primary Care study. Swabs have the potential to be more cost effective and convenient for sample collection. Our goals were to compare the DNA yield and quality, to assess the feasibility of whole genome amplification (WGA) of low quantity DNA samples, and to evaluate sample performance on clinically relevant pharmacogenetic assays. One OraCollect DNA (OCR-100, DNA Genotek) and four PurFlock Ultra swabs (Puritan) were collected from thirty-one study volunteers. Genomic DNA (gDNA) was extracted using a bead-based extraction kit (Ambion, MagMAX, Applied Biosystems). To simulate postal delays, PurFlock swabs were extracted at two different time points: 3 days and 7 days after sample collection. A WGA step was added to the Puritan swabs extracted at day 3 using the REPLI-g kit (Qiagen). The DNA yield (mean±SD) and A260/280 (mean±SD) ratio was found to be 1.9±1.4µg and 1.85±0.39 for OraCollect DNA and 0.25±0.35µg and 1.72±0.78 for PurFlock swabs, respectively. Results indicated no significant differences in the yield and quality between PurFlock swabs extracted at two different time points. WGA resulted in a mean increase of 0.34 µg of initial DNA. SNP genotype call rate on the OpenArray for the WGA samples was low (21%) and only 45% of samples had copy number variant (CNV) confidence scores >95% for a TaqMan CNV assay. In comparison, samples obtained from OraCollect genotyped on the same platform had a genotyping call rate of 97% and CNV confidence scores of 100%. We conclude that WGA from gDNA extracted from buccal swabs is not suitable for genotyping on the OpenArray and may introduce bias for CNV assignment. The OraCollect DNA swab is a reliable source of genomic DNA that can be effectively used for pharmacogenetic testing.

1611F

Implementation of pharmacogenetics with clinical decision support in primary care. *H. Katzov-Eckert[1], M. Aloise[1], J.S. Ang[1], D. Dawes[1,2], A. Gonzalez-Suarez[1,3], G. Liknaitzky[1], A. Paterson[1,4,5], P. Stanley[1], R. Fraser[1,6], P. Cullis[1,6], M. Dawes[1,7].* 1) GenXys Health Care Systems, Vancouver, BC; 2) University of British Columbia Physical Therapy, Vancouver, BC; 3) University of British Columbia Department of Botany, Vancouver, BC; 4) University of British Columbia Faculty of Pharmaceutical Sciences, Vancouver, BC; 5) Clinicare Pharmacists Inc, Vancouver, BC; 6) University of British Columbia Personalized Medicine Initiative, Vancouver, BC,; 7) University of British Columbia Family Practice, Vancouver, BC.

Inappropriate prescribing increases patient morbidity and death due to adverse drug events. The inclusion of pharmacogenetics into primary care medication practices is one solution. We conducted a cohort study in six primary care settings, enrolling 191 adults with at least one of ten common diseases. To develop a pharmacogenetic test, evidence for genotype-guided medication dosing was compiled for drugs commonly prescribed by family physicians. Genetic variants that influence drug response were ranked according to clinical annotations from PharmGKB, Clinical Pharmacogenomics Implementation Consortium and FDA guidelines. Finally, a panel of genetic variants was selected and validated including *CYP2D6, CYP2C19, CYP2C9, VKORC1 SLC01B1, G6PD* and *HLA-B*58:01*. Saliva samples were collected and genomic DNA was isolated using a magnetic bead based extraction method (Ambion® MagMAX™, Applied Biosystems). Genotyping of SNPs and CNV was carried out by Openarray™ and TaqMan™ probe assays on the QuantStudio 12K Flex System (ThermoFisher Scientific). To provide personalized treatment options, annotated genotypes were integrated into a medication decision support system called TreatGx. Findings indicated that 96.8% of samples had at least one actionable genotype for medications included in TreatGx. Moreover, TreatGx facilitated the use of genetic information in the decision-making processes of health care professionals working in primary care. In conclusion, a decision support system that includes genetics can optimize personalized prescribing in primary care and may lead to a reduction in inappropriate prescribing.

1612W

The global spectrum of coding region pharmacogenomic diversity. *G. Wright[1], B. Carleton[2], M. Hayden[1], C. Ross[3].* 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Pediatrics, University of British Columbia, Vancouver, BC, Canada; 3) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada.

Background Inter-individual differences in response to medications are known to have a strong genetic component. By leveraging publically available genomic data, the spectrum of genetic variation in genes that influence either response to medications or adverse drug reactions can be investigated extensively. **Methods** Genes were selected based on relevance to pharmacogenomics according to the PharmGKB database and the literature. Genetic variants in these regions of interest were extracted from the 1000 Genomes Project data from 26 global populations (*n*=2504). Variants were then annotated using the Ensembl VEP and analysed in further detail utilising the program, VCFtools and the software environment, R. **Results** A total of 11902 genetic variants were found in the 58 genes that met inclusion criteria, with over 80% of the dataset comprising of rare variants. Seven genes were excluded due to accessibility to sequencing technologies and there was an overrepresentation of cytochrome *P450* genes found in segmental duplications (*P*=0.006). Pharmacogenomic variation tended to be most similar within continental populations and 16 gene regions contained a highly differentiated marker. Samples carried a median of five clinical variants with a high level of evidence and 36% of samples carried at least one loss of function (LOF) variant. LOF variants were found in 64% of the pharmacogenes, and *CYP3A5*, which influences tacrolimus dose, displayed the highest number of unique LOF variants. **Conclusion** These data demonstrate that current sequencing technologies can successfully identify large amounts of clinical pharmacogenomic variation. Since the relative frequency of deleterious variants is inversely correlated with allele frequency, future approaches need to consider this class of variation to fully understand the full spectrum of genetic diversity contributing to pharmacogenomic traits.

1613T

Introducing "SG10K": Cataloging genetic diversity and population structures in 10,000 South Asians. *C. Bellis[1], I.D. Irwan[1], C. Wang[2], W.W. Soon[3], A. Wilm[4], C.C. Shih[4], R.S. Foo[5,6], A.M. Richards[5], C.S. Lam[7], H.H. Ng[8], J. Liu[1].* 1) Human Genetics 2, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 2) Computational and Systems Biology, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 3) Next Generation Sequencing Platform, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 4) Scientific and Research Computing, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 5) Cardiovascular Research Institute, National University Health System, Singapore; 6) Human Genetics 3, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 7) National Heart Centre Singapore, Singapore; 8) Gene Regulation Laboratory, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore.

Genetic variation plays an important role in a variety of human diseases and quantitative traits. Due to different underlying genetic architecture and contrasting environmental exposures, many genetic findings have shown population-specific characteristics, highlighting the importance of population diversity in human genetic studies. The Singapore population consists of three major ethnic groups, Chinese, Malay, and Indian, which together represent >80% of the Asian population. To empower biomedical and human genetic studies in Asian populations, the SG10K project will perform 12-13× WGS of 10,000 Singaporeans. Coupled with powerful bioinformatics tools, our study design will enable high-quality genotype calling for a full frequency spectrum of genetic variants segregating in the population. Our main objectives are to (1) comprehensively characterize genetic variation in Singapore population; (2) create a WGS reference panel for accurate genotype imputation in Asian population; and (3) generate a large control dataset for WGS-based genetic association study of diseases. Our extensive calculations have provided substantial preliminary evidence allowing us to confidently proceed with our plan to survey our 10,000 samples at a depth of coverage between 12-13× in our WGS strategy. Because the phased variant calling will become more powerful with the increased sample size (via more accurate haplotype information), 10× WGS will allow us to characterize the full spectrum of germ-line genetic variants (except the private ones) in 10K individuals with the similar accuracy and sensitivity as 30× WGS. To achieve our goals, we will be employing the Illumina TruSeq Nano DNA Library Preparation Kit and sequencing study participants on the Illumina HiSeq 4000 instrument, in-house at GIS. However, this project is a fully national effort across multiple institutions; our collaborative partners include SingHealth Duke-NUS Institute of Precision Medicine, Singapore Eye Research Institute, Centre for Personalised and Precision Health, National University Health System and several Translational and Clinical Research Flagship Programmes (Heart failure, Parkinson disease). To date we have taken possession of the complete SG10K cohort and sequenced 576 samples. Upon completion, this study will provide valuable genetic information to facilitate clinical and pharmaceutical research in Singapore populations and will empower genetic studies of Singapore and Asian-centric diseases.

1614F

Genome-wide association study of sleep duration in the 14,000 Japanese general population. *T. Nishiyama[1,3], M. Nakatochi[2], A. Hosono[3], Y. Tamai[3], S. Suzuki[3], A. Hishida[4], K. Wakai[4], H. Tanaka[5], K. Matsuo[6], The Japan Multi-Institutional Collaborative Cohort Study.* 1) Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan; 2) Center for Advanced Medicine and Clinical Research, Nagoya Univ. Hospital, Nagoya, Japan; 3) Dept. of Public Health, Nagoya City Univ. Graduate School of Medicine, Nagoya, Japan; 4) Dept. of Preventive Medicine, Nagoya Univ. Graduate School of Medicine, Nagoya, Japan; 5) Division of Epidemiology and Prevention, Aichi Cancer Center, Nagoya, Japan; 6) Division of Molecular Medicine, Aichi Cancer Center, Nagoya, Japan.

Usual sleep duration is heritable with heritability estimates being around 40%. Candidate gene association studies for sleep duration identified a number of potential loci, though the results were inconsistent. A GWAS of sleep duration have identified and replicated two SNPs near *PAX8* and *CBWD2* [Gottlieb DJ 2015], but in the other GWAS, any locus have neither reached genome-wide significance nor been replicated in independent samples presumably due to limitations in sample size. Therefore, we conducted a GWAS of sleep duration using data from the Japan Multi-Institutional Collaborative Cohort Study (J-MICC study). The J-MICC study is one of the largest population-based cohort studies in Japan [Wakai K 2011]. The participants were enrolled in 10 study areas throughout Japan except for the northern region. Phenotype information was obtained from self-administered questionnaires at study entry. The ethics committees of all participating institutions approved the protocol for the J-MICC study, and all participants provided written informed consent. Genotyping was performed with the Illumina Infinium OmniExpressExome-8 BeadChip. In total, 513,776 SNPs in 14,088 participants passed our quality control criteria (MAF ≥ 5%, HWE p-value ≥ 0.001 and call rate ≥ 95% for both individuals and markers). Single SNP association analysis of sleep duration as a quantitative trait was performed using PLINK, adjusted for age and gender. Population structure was also adjusted by the first principal component from principal components analysis. A gene-based association test was performed using the SKAT package in R. In the single SNP analysis, two SNPs (rs12229654 and rs2188380) were significantly associated with sleep duration with genome-wide level (p-value = 5.E-09 and 1.E-08, respectively). In the gene-based analysis, a gene, *PTPLAD2* was significantly associated with sleep duration after adjusting for multiple comparisons (p-value = 0.047 in Bonferroni method). The two associated SNPs are located 28kb apart from each other and display very strong linkage disequilibrium (r2 = 0 .94). These genes are located between two genes, *MYL2* and *CUX2* on chromosomes 12. These genes were not identified in the former GWAS of sleep duration and warrants further exploration.

1615W

The scope and mechanism of inbreeding depression in humans. *D.W. Clark[1], P.K. Joshi[1], I. Gandin[2], T. Esko[3,4,5], J.F. Wilson[1,6], ROHgen consortium.* 1) Centre for Global Health Research, University of Edinburgh, Edinburgh, Midlothian, United Kingdom; 2) Department of Medical Sciences, University of Trieste, Strada di Fiume 447 - Osp. di Cattinara, Trieste, 34149, Italy;; 3) Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia;; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge Center 7, Cambridge, 02242, MA, USA;; 5) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Cambridge, 02141, MA, USA;; 6) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland.

Homozygosity caused by consanguineous union has long been associated with an increased prevalence of rare Mendelian disorders. In contrast, the role of homozygosity in relation to quantitative traits and complex disease susceptibility is less well established. Recent work has shown that an increased burden of homozygous DNA segments (defined as the fraction of each genome in any Run of Homozygosity > 1.5 Mb, or F_{ROH}) is associated with reduced height and cognitive ability in diverse human populations. Here, we assess the scope of inbreeding depression in humans by analysing the effect of F_{ROH} on 41 quantitative traits of public health and evolutionary interest, in >500,000 individuals from >100 cohorts worldwide. Phenotypes studied include anthropometry, lipids, liver, lung and kidney function, cognition, glycaemia, fertility, inflammation, electrocardiography and haematology. We confirm the findings of previous work, and also show that increased F_{ROH} is associated with reduced values of several components of evolutionary fitness. The constancy of these effects across different levels of homozygosity and across diverse populations in which social confounding more plausibly engenders the opposite effect, strengthens the argument that F_{ROH} has a causal effect. In addition, we investigate the magnitude of inbreeding depression between men and women to explore whether sex-specific dominance may provide a mechanism, independent of the sex-chromosomes, by which males and females might adapt to different selective pressures. We investigate the mechanism of inbreeding depression by exploring different measures of homozygosity. Long homozygous segments arise from recent common ancestry and bring all variants, including rare causal variants, into a homozygous state. In contrast, shorter homozygous segments may arise from more ancient common ancestry and thus capture only the dominance effects of more common variants. Using multivariate linear models we show that F_{ROH} is more predictive of inbreeding depression than traditional estimates of inbreeding coefficient based on individual SNP homozygosity.

1616T

Analysis of expression quantitative trait loci in the human eye. *L.D. Orozco[1], J. Hackney[1], C. Cox[2], Y. Li[2], M. Jeanne[2], M. van Lookeren Campagne[2].* 1) Bioinformatics, Genentech, South San Francisco, CA; 2) Immunology, Genentech, South San Francisco, CA.

Age-related Macular Degeneration (AMD) is the leading cause of vision loss in the elderly. AMD affects over 10 million people in the US, and 20-25% of caucasians over 75 years of age. Risk factors of AMD include smoking, age, ethnicity, and genetics. In the most recent GWAS study, Fritsche and colleagues identified a total of 34 loci that account for 46% of the genetic variance, and observed a heritability of 44% for "wet" AMD, and 52% heritability for "dry" AMD. Anti-VEGF drugs such as *Lucentis* have revolutionized treatment for *neovascular* or "wet" AMD, but there are currently no available treatments for patients suffering from *geographic* or "dry" AMD. Risk variants identified through GWAS are a valuable source for new target identification, and eQTL studies allow for causal gene identification in complex traits. However, a focused eQTL study in the eye tissues affected by AMD has not been performed. In order to better understand the disease pathogenesis and prioritize candidate genes for AMD, we performed QTL analysis to identify tissue specific *cis*-eQTL. We profiled human eye donors for genotypes in 2 million SNPs, and expression patterns using RNAseq in macular and extra-macular retina and RPE/choroid. To identify eQTL, we filtered genes with normalized RPKM greater than 0.1 in at least 10 donors, and transformed expression levels using quantile normalization. The covariates in the model included 15 PEER factors estimated from the normalized expression, sex, age, and 3 genotyping eigen vectors. We selected SNPs with minor allele frequency greater than 10%, and missing frequency in less than 10% of individuals. Finally, we fit a linear regression model using the package MatrixEQTL, and called *cis*-eQTL based on 1 Mb distance from the gene. Overall, we followed the analysis pipeline used by the GTEx project. Here we describe, for the first time, eQTL in multiple tissues of the human eye. Our findings will aid our understanding of the biology behind AMD pathogenesis, and can be used to identify novel therapeutic targets and biomarkers.

1617F

Identifying genetic associations with variability in metabolic health and blood count laboratory values: Diving into the quantitative traits in an EHR. *S. Verma[1,2], A. Lucas[2], J. Leader[2], C. Bauer[2], R. Metpally[2], S. Krishnamurthy[2], F. Dewey[3], A. Lopez[3], J. Overton[3], J. Reid[3], G. Breitwieser[2], S. Pendergrass[2], M. Ritchie[2,4].* 1) The Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA; 2) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 3) Regeneron Genetics Center, Tarrytown NY; 4) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

Metabolic and complete blood count (CBC) panels provide a comprehensive picture of health and disease in patients during routine clinical care. These blood panels are useful in diagnosis of many common, complex diseases, and can be used as quantitative traits to elucidate the genetics of disease. We extracted 25 clinical laboratory measures from the metabolic and CBC panels in ~40,000 patients from Geisinger Health System MyCodeâ Discov-EHR cohort to perform association studies of both common and rare variants associated with these traits. We identified 2,453 unique common loci (minor allele frequency >1%) above genome-wide significance (p-value < $5x10^{-8}$) that are associated with one or multiple traits. This analysis validated many known variants and revealed novel genes associated with variability among patients in these 25 phenotypes. For example, we replicate several known variants in the *UGT1A1* gene associated with bilirubin levels and also found many variants in genes *CASR*, *CCDC58*, *KPNA1*, *CSTA* associated with serum calcium levels. Our study also identified 22 associations with platelet counts where the odds ratio ranges from 172.19-11.22. These include some known associations (such as variants in *NF-E2* and *GP1BA*) as well as novel association such as variants in *PTPN11* and *ITPR3* that have not been identified by previous GWAS. We also explored pleiotropic associations from this study and identified variants in *ABO* and *GCKR* genes to be associated with multiple traits. Next, we explored the trait variability in patients based on the longitudinal data from the EHR, we tested patients with high and low variance for the 25 labs for association with ICD-9 based case/control diagnoses. Lastly, we performed gene-based rare variant collapsed analysis for all variants with minor allele frequency < 1% from whole exome sequencing data for same individuals. Rare variant binned SKAT-linear analysis also resulted in many known and novel associations. There were 235 results that pass bonferroni significance ($8.74x10^{-8}$) at alpha 0.05 (25 phenotypes and 22,874 genes tested). Among the top results is the association between rare variants in *HTR1F* and white-blood cell count (p-value $1.54x10^{-10}$). This study shows the utility of exploring clinical lab measures for association testing, identifying both common and rare variants related to trait variability among patients, and present a novel approach for exploring of associations between trait variability and disease diagnoses.

1618W

Genetic overlap between dizygotic twinning and BMI, height and smoking: Results from the Twinning GWAS Consortium. *H. Mbarek[1], M. Nivard[1], E.A. Ehli[2], G.E. Davies[2], D.I. Boomsma[1].* 1) Department of Biological Psychology, Netherlands Twin Register, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; 2) Avera Institute for Human Genetics, Sioux Falls, SD 57108, USA.

Spontaneously becoming the mother of dizygotic twins (MoDZ) is a complex trait, which aggregates in families. MoDZ might index fertility and has distinct health implications for both the mother and the child. Recently we reported for the first time that sequence variation at the FSHB and SMAD3 loci increases the odds of MoDZ, based on three discovery cohorts from the Netherlands, Australia and Minnesota (USA). The findings were replicated in a large cohort from Iceland by deCODE (1).Our prior work on the role of body composition and smoking in DZ twinning (2) established that MoDZ is significantly associated with increasing height, an increased BMI, and smoking before the twin pregnancy. Here we used our meta-analysis results and the LD score regression method (3) to estimate genetic correlations between MoDZ and BMI, height and smoking. MoDZ shows positive genetic correlations with BMI, height and smoking (ever smoking and age at onset). All results are in the direction predicted by our prior work, however the correlations did not pass the significance threshold of 0.05 (corrected for multiple testing), likely due to low statistical power (our discovery sample included only ~2000 cases and 14,000 controls) and suggests that polygenic risk score analyses will give additional insights into the nature of these associations. 1. Mbarek, H., Steinberg, S., Nyholt, D.R., Gordon, S.D., Miller, M.B., McRae, A.F., Hottenga, J.J., Day, F.R., Willemsen, G., de Geus, E.J., et al. (2016). Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and Female Fertility. Am J Hum Genet 98, 898-908. 2. Hoekstra, C., Willemsen, G., van Beijsterveldt, C.E., Lambalk, C.B., Montgomery, G.W., and Boomsma, D.I. (2010). Body composition, smoking, and spontaneous dizygotic twinning. Fertil Steril 93, 885-893. 3. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.

1619T

Genetic variants do not predict efficacy response with mepolizumab in severe asthma. *L. Condreay[1,2], M. Chiano[3], H. Ortega[2,4], N. Buchan[3], E. Harris[1,2], S.W. Yancey[2], S. Ghosh[5]. 1) PAREXEL International, Research Triangle Park, NC; 2) GSK, Research Triangle Park, NC, USA; 3) GSK, Stevenage, UK; 4) Genentech, South San Francisco, CA, USA; 5) GSK, King of Prussia, Pennsylvania, USA.*

Purpose: Treatment with mepolizumab, a humanized monoclonal antibody to interleukin-5, reduces the rate of asthma exacerbations and the requirement for systemic glucocorticoids while maintaining asthma control. Treatment benefit is guided by predictors of response such as blood eosinophil thresholds in patients with frequent exacerbations despite intensive anti-inflammatory and controller treatment. We investigated genetic variants that may predict efficacy response to mepolizumab treatment. **Methods:** Candidate genetic variant and an exploratory genome-wide association analyses (GWAS) were performed in 1053 patients with severe eosinophilic asthma treated with at least one dose of mepolizumab who provided consent for pharmacogenetic (PGx) research enrolled in phase 3 clinical studies, DREAM and MENSA. Evaluation of baseline factors indicated that PGx cohorts for DREAM (n=589) and MENSA (n=464) were representative of the intent-to-treat study populations (n=616 and 576, respectively). Analysis was post-hoc and included non-Hispanic white patients (n=820). A total of 105 genetic variants in 53 gene regions were selected for analysis based on prior history of association with relevant traits or pathways. Efficacy endpoints included: rate of clinically significant exacerbation (primary endpoint, tested by generalized linear regression), time to first exacerbation (tested with a Cox proportional hazards model), and change in exacerbation rate, change in eosinophil count, or change in IgE level (tested by linear regression). For the primary endpoint, candidate variant analysis had 80% power to identify exacerbation rate ratios of ≥ 1.6 (or ≥ 2.0 for GWAS) for common genetic variants with a minor allele frequency >20%. **Results:** No genetic variant was significantly associated with the primary endpoint, clinically significant exacerbation rate in candidate variant or GWAS analyses. The intronic variant, rs1021621, in *POU2F1* was significantly associated with time to first clinically significant exacerbation (p=2×10^{-5}) but the association was driven by DREAM data and not supported in additional sensitivity analyses by treatment regimen/dose. The available sample size had limited statistical power to detect small genetic effects from common variants or effects from rare variants using GWAS. **Conclusion:** No genetic effect on mepolizumab-treatment response was identified in this population with severe eosinophilic asthma. **Funding:** GSK (NCT01691859, NCT0169521).

1620F

Gene expression response to rhinovirus infection differs between Chronic Rhinosinusitis (CRS) Patients and non-CRS controls. *M. Soliai[1], A. Kato[3], J. Norton[3], A. Klinger[3], B. Tan[4], R. Kern[4], R. Schleimer[3], J. Pinto[2], C. Ober[1]. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Surgery, University of Chicago, Chicago, IL; 3) Department of Surgery, Northwestern University Feinberg School of Medicine; 4) Department of Medicine, Northwestern University Feinberg School of Medicine.*

Chronic rhinosinusitis (CRS) is an inflammatory disorder of the nose and paranasal sinuses that affects upwards of 12% of the U.S. population. Microbes in the upper respiratory system can influence the development of CRS in individuals at risk for this disease, but little is known about inter-individual responses to infection. In this analysis, we interrogated genes that are differentially expressed in response to human rhinovirus (RV) infection in nasal uncinate tissue epithelial cells (UTEC) obtained from 47 cases with CRS with nasal polyps (CRSwNP) and 52 non-CRS controls. UTECs were cultured with RV or a vehicle control for 48 hours, followed by sequencing (Illumina RNA-Seq). Technical sources of variation were identified and adjusted by using principle components analysis and ComBat, respectively. A paired analysis of gene expression response to RV was conducted to identify different responses between cases and controls using limma, and including age, sex, and race as covariates. This was followed by Ingenuity Pathway Analysis (IPA) of the differentially expressed genes to identify multi-gene networks. Overall, 805 of the 11,341 genes that passed quality control were differentially expressed in response to RV in the cases, controls or combined sample (FDR 10% in each sample). 113 genes were responsive to RV in both CRS cases and controls, 428 were responsive to RV in the cases only, and 264 genes were responsive to RV in the controls only. Response genes unique to the controls, as well as response genes common to both cases and controls, are well known genes in viral response pathways, and are involved in networks associated with antimicrobial and inflammatory responses and infectious diseases (network score p-values < 10^{-40}). In contrast, response genes that were unique to the cases formed networks with functions associated with nucleic acid metabolism and small molecule biochemistry. Among these uniquely expressed genes in CRSwNP, *HDAC5* is a major gene hub in the top-scored network (p = 10^{-47}). Moreover, 10 response genes unique to CRSwNP cases are implicated in the cholesterol biosynthesis pathway (Fisher Exact Test, p-value = 6.14×10^{-11}), including *DHCR7*, *FDFT1*, and *HMGCR*. Overall, these data show different gene expression responses to RV between CRSwNP cases and non-CRS controls, and implicate different biological pathways in such responses, informing our understanding of underlying disease biology. Supported by AI106683.

1621W

Whole genome sequence analysis of a multi-ethnic population of children with asthma reveals novel pharmacogenetic associations. *M. White[1], A.C.Y. Mak[1], Z. Szpiech[1], D. Hu[1], M. Pino-Yanes[3], S. Huntsman[1], C. Eng[1], S. Oh[1], J. Galanter[1], D. Torgerson[1], M. Seibold[2], E. Ziv[1], N. Zaitlen[1], R. Hernandez[1], E.G. Burchard[1], New York Genome Center.* 1) Department of Medicine, University of California, San Francisco, San Francisco, CA; 2) Center for Genes, Environment, and Health, Department of Pediatrics, National Jewish Health, Denver, Colorado; 3) Research Unit, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain.

Asthma is a complex disease characterized by chronic inflammation of the airways, and is the most common chronic disease of childhood worldwide. Latinos and African Americans have the highest asthma morality and the lowest drug response to asthma therapies. Specifically, there are ethnic disparities to albuterol bronchodilator drug response (BDR) among children with asthma. Previous GWAS studies have failed to reveal common variants associated with BDR that account for the majority of the variation in BDR. We analyzed whole genome sequence (WGS) data to evaluate the impact of common variant effects on BDR in a multi-ethnic pediatric asthma study. We leveraged the increased statistical power obtained from performing an extreme phenotype study design to identify common variants associated with BDR. Our study consisted of 1500 children (500 Mexican, 500 Puerto Rican, 500 African American) sampled from the extremes of each group's BDR distribution. We performed association tests separately in each population and then performed a trans-ethnic meta-analysis across all populations to identify cosmopolitan effects. We identified one variant near *DNAH5* that reached genome-wide significance across populations (meta-analysis β= 0.44, p = 8.11 x 10⁻⁸), after correcting for the number of independent tests. This finding is of particular interest, as *DNAH5* has been associated with lung capacity in chronic obstructive pulmonary disease. To our knowledge, this is the first study to comprehensively investigate the effects of common variants in three admixed populations of children collected from the extremes of the bronchodilator drug response phenotypic distribution.

1622T

SNPs, linkage disequilibrium and transcriptional factor binding sites associated with chronic mountain sickness among Tibetan Chinese at the Qinghai-Tibetan Plateau. *N. Buroker[1], X-H. Ning[1,2], Z-N. Zhou[4], K. Li[3], W-J. Cen[3], X-F. Wu[4], W-Z. Zhu[5], C.R. Scott[1], S-H. Chen[1].* 1) Pediatrics, 356320, University of Washington, Seattle, WA; 2) Division of Cardiology, Seattle Children's Hospital. Institute. Foundation, Seattle, WA; 3) Lhasa People Hospital, Tibet; 4) Laboratory of Hypoxia Physiology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China; 5) Center for Cardiolvascular Biology and Regenerative Medicine, University of Washington, Seattle, WA.

Chronic mountain sickness (CMS) or Monge's disease is estimated at 15% in Quechua Andeans and 1.2% in Tibetans living at the same high altitude greater than 3000 meters above sea level. A CMS Tibetan Chinese and a normal Tibetan group were compared. Eighteen simple nucleotide polymorphisms (SNPs) within 9 genes that have been associated with CMS were analyzed with respect to linkage disequilibrium (LD) between intra- and intergenic SNP alleles and alterations in transcriptional factor binding sites (TFBS). Included in the study was the angiotensin-converting enzyme (ACE) (rs4340), the angiotensinogen (*AGT*) (rs699) and the angotensin II type 1 receptor (*AGTR1*) (rs5186) SNPs from the renin-angiotension system (RAS) and a LDL apolipoprotein B (*APOB*) (rs693) SNP. The endothetal Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) SNP and an egl nine homolog 1 (*EGLN1*) SNPS (rs480902) from the hypoxia-inducible factor (HIF) oxygen signaling pathway were included. SNPs analyzed in the vascular endothelial growth factor (VEGF*)* signaling pathway are the v-akt murine thymoma viral oncogene homolog 3 (*AKT3*) (rs4590656 and rs2291409), the endothelial cell nitric oxide synthase 3 (*eNOS3*) (rs1007311 and rs1799983) and the (*VEGFA*) (rs699947, rs34357231, rs79469752, rs13207351, rs28357093, rs1570360, rs2010963 and rs3025039). These SNP alleles alter the TFBS for transcriptional factor (TF) binding. Pair-wise LD was computed between SNPs. An increase in LD occurred in 40 pair-wise comparisons while a decrease was found in 55 pair-wise comparisons between the CMS and controls. Increases and decreases in LD pairs were found within and between signaling pathways and systems indicating the interaction of SNP alleles or potential TFBS from different areas of the genome. The most drastic change in TFBS occurs with ACE (I/D) SNP (rs4340) where the ACE-I allele generates 84 potential TFBS while the ACE-D allele generates only four binding sites. The alteration in TFBS generated by the 18 SNPs is discussed with respect to CMS.

1623F

Real-life prosocial attitudes in North America: A genome-wide association approach. *G. Primes, M. Fieder.* University of Vienna, Austria.

   In humans, prosocial behaviour is essential for social functioning. Twin studies suggest this distinct human trait to be partly hardwired. In the last decade research on the genetics of prosocial behaviour focused on neurotransmitters and neuropeptides, such as oxytocin, dopamine, and their respective pathways. Recent trends towards large scale medical studies targeting the genetic basis of complex diseases such as alzheimer's disease and schizophrenia pave the way for new directions also in behavioural genetics.Based on data from 10,713 participants of the American Health and Retirement Study we estimated heritability of helping behaviour – its total variance explained by 1.2 million single nucleotide polymorphisms – to be 11%. Two approaches of genetic association analyses identified *RS11697300*, an intergene variant on chromosome 20, as a candidate variant moderating this particular prosocial attitude. We assume that this so far undescribed area is worth further investigation in association with human prosocial behaviour.

1624W

Elucidating the complex relationship between pleiotropy and disease risk in the human genome. *K. Chesmore[1], J. Bartlett[2], S. Williams[2].* 1) Department of Genetics, Dartmouth College, Hanover NH; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland OH.

   In recent years several studies have investigated pleiotropy in order to better understand the complex relationship among diverse human diseases. However, one facet of this relationship that has largely been unexplored is whether pleiotropic genes (those that associate with multiple disease phenotypes) have larger effects sizes than those that appear to only associate with a single disease. Taking advantage of the wealth of information gathered from the GWAS catalog, we explored this relationship using the number of disease phenotypes associated with each gene and the Odds Ratios of case/control studies as a measurement of effect size. We identified 10,153 genes that associate with 847 diseases and tested whether the number of diseases associates with a gene's effect size, using linear regression. Our results showed that genes associating with more disease phenotypes have significantly larger effect sizes (p = 2*10^-8). Additionally, we explored how the collection of more data through the years has helped to resolve this relationship. As more data was added to the GWAS catalog over the years we found evidence of more examples of pleiotropy for more genes (at present 46% of genes in the GWAS catalog are pleiotropic up from 10% in 2010). The shift of this distribution is resolving the relationship between pleiotropy and effect size, resulting in higher levels of correlations with more significant p-values. As more data is collected we expect this relationship to be resolve even further. Elucidating this relationship and the underlying mechanisms is crucial for understanding the potentially common between multiple diseases, and may allow for the development of predictive models of the molecular mechanisms that drive human diseases.

1625T

Polygenic load for neuroticism predicts chronic pain in the UK Biobank cohort. *M.E.S. Bailey[1], J. Ward[2], D.J. Smith[2], B.I. Nicholl[2], J.P. Pell[2].* 1) School of Life Sciences, CMVLS, University of Glasgow, Glasgow, United Kingdom; 2) Institute of Health and Wellbeing, CMVLS, University of Glasgow, Glasgow, United Kingdom.

Chronic pain affects up to 20% of adults in Europe and accounts for more disability in the developed world than any other condition. It is highly heterogeneous and influenced by a complex mixture of social, psychological and environmental exposures, yet its heritability is estimated at 20-40%, with a substantial genetic contribution. The basis for the known correlation between chronic pain and personality/psychological traits (e.g. neuroticism) is unclear. We have investigated the connection between chronic pain and neuroticism in the UK Biobank (UKB) population cohort using a polygenic risk score (PRS) approach. Data from a recent meta-GWAS of neuroticism (Smith et al., 2016), which included >90,000 individuals from UKB, were used to derive a series of PRSs with meta-GWAS-derived thresholds from $p<5x10^{-8}$ up to $p<0.5$. SNPs underwent full QC and were filtered after clumping for linkage disequilibrium. PRSs were calculated, using GWAS β weighting, for unrelated white European individuals in UKB that had not been included in the meta-GWAS. PRSs were ranked and divided into 13 quantiles, which was used as a quasi-quantitative variable in case/control logistic regression models in R. The main outcome tested was presence of chronic pain (>3 months in duration) at any of several pre-specified sites in the body (Case n=10,722). Controls (n=13,114) included only those reporting no pain >1 month in duration. Models were adjusted for 8 principal genetic components, age, sex, genotyping chip and data collection centre. Increasingly inclusive PRSs defining polygenic load for neuroticism were notable predictors for chronic pain (best $p=6.6x10^{-10}$) with a per-PRS level O.R. of 1.02. Number of pain sites was also predicted by neuroticism PRS, and individual sites predicted most strongly were back, neck, shoulder and head pain. Our findings support a biological basis for the known correlation between neuroticism and chronic pain. The polygenic load for neuroticism in the general population correlates with chronic pain prevalence and with number of pain sites reported. Mendelian Randomisation approaches may help distinguish whether this association is due to shared pleiotropic variants or to a causal influence of neuroticism-related psychology on propensity to experience or report pain. Further understanding of the biology of chronic pain will inform the development of new treatment approaches that help individuals maintain both good quality of life and economic productivity.

1626F

Discovery and replication of rare variant associations using a knowledge-driven PheWAS approach in eMERGE and Geisinger Health System. *A.O. Basile[1], A.M. Lucas[2], A. Verma[2], A.T. Frase[2], M. de Andrade[3], M.H. Brilliant[4], J.C. Denny[5], D.R. Velez Edwards[5], T. Edwards[5], M.M. He[4], S.J. Hebbring[5], B. Namjou-Khales[6], A. Lopez[7], J. Overton[7], J. Penn[7], J. Reid[7], L.J. Rasmussen-Torvik[8], O.J. Veatch[5], M.D. Ritchie[1,2].* 1) Department of Biochemistry and Molecular, The Pennsylvania State University, University Park, PA; 2) Department of Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 3) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 4) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 5) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville,TN; 6) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 7) Regeneron Genetics Center, Tarrytown, NY; 8) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.

Rare genetic variation has recently been implicated in various multifactorial conditions such as Alzheimer's disease and obesity. Elucidating the influence of rare variants has the potential to identify novel associations and expand our current understanding of the heritability of complex traits. As these variants are individually uncommon, standard approaches to association studies are statistically underpowered. We addressed this challenge by applying a knowledge-driven approach using BioBin to bin variants into biological features, thereby increasing the composite allele frequency. Herein, we applied Bio-Bin-SKAT (Bin-KAT) in a gene-based analysis to examine the influence of rare and low-frequency variants (MAF< 5%) in 8 phenotypes, asthma, benign prostatic hyperplasia, cataracts, diverticulosis, gastroesophageal reflux disease, hypertension, hypothyroidism and uterine fibroids, using two distinct EHR-linked biorepositories. Discovery analysis was performed in 7,000 subjects from the electronic Medical Records and Genomics (eMERGE) Network as part of the eMERGE-PGX study. Subjects were sequenced using PGRNseq, an 82 pharmacogene targeted sequencing platform. Case-control designation was determined by ICD-9 codes. All significant models from eMERGE (p< 0.05) were then tested in a replication set of approximately 39,000 individuals from the Geisinger Health System MyCode® Community Health Initiative sequenced at Regeneron Genetics Center as part of DiscovEHR. Associations in both datasets were evaluated using the SKAT implementation in BioBin, adjusting for sex, age, median BMI, and principal components. Our gene analysis, replicated 8 results with a p< 0.01, and 1 of these associations replicated with high significance (p< 10-4). The most significant replicating associations include rare variants in *CYP2C19* with asthma, *CRHR1* and *NR3C1* with uterine fibroids and *CES2* with diverticulosis. Notably, *CYP2C19* has previously been implicated in the development of asthma, as genetic variation is observed to affect the enzyme's ability to metabolize pollutants. We also applied Bin-KAT to 2 other biological feature analyses (pathway and SNPEff functional predictions) to provide further insight into the rare variant influence on these phenotypes. Our findings support a role for low-frequency variants in susceptibility to multiple complex phenotypes and demonstrate the utility of Bin-KAT for the biologically-inspired binning and effective analysis of rare variants.

1627W

Complex traits analysis with big data: A landscape of the effects of sample size, number of SNPs and the statistical model on prediction accuracy. *G. de los Campos[1,2], A. Grueneberg[1], H. Kim[1], A. Vazquez[1], S. Hsu[3].* 1) Epidemiology & Biostatistics, Michigan State University, East Lansing, MI; 2) Statistics & Probability, Michigan State University, East Lansing, MI; 3) Department of Physics, Michigan State University, East Lansing, MI.

The continued reduction of genotyping costs has lead to an increase in the availability of data sets comprising genotypes linked to phenotypes and disease outcomes. In the last decade GWAS studies have reported large number of variants associated with important human traits and diseases; however, our ability to use DNA information to predict phenotypes and health outcomes with complex genetic architecture remains limited. Prediction accuracy (PA) depends on the genetic architecture of the trait or disease being analyzed, the number of markers used (P) and how these were selected (S), the size (N) of the training (TRN) data set and the statistical method (M) used to estimate effects. Until recently, the investigation of how these factors interact in determining PA has been limited by the size of standard GWAS cohorts. Initiatives such as the UK Biobank and US Precision Medicine Initiative® will soon deliver very large biomedical datasets comprising genetic, clinical and phenotypic information; these resources will enable empirical research at a scale that was not previously possible. In this study we use data from the UK Biobank corresponding to unrelated individuals of Caucasian background to produce a landscape of PA in testing (TST) data as a function of N (from 5K to 80K records in TRN, K=thousand), P (from 0.5K to 50K SNPs, selected from GWAS done with TRN data), S (independent vs. linked markers) and M (shrinkage vs. variable selection methods) for anthropometric traits with complex genetic architecture. Using N=80K (20K) from model TRN (TST) and 50K SNPs we achieved levels of PA in TST data higher than what has been previously reported for the traits analyzed. The estimated response surface indicates that PA increases with P for all traits; however the increase in PA occurs at diminishing marginal rates; consequently, PA plateaus at a level that is largely determined by heritability and N. Our results also indicate that increasing N beyond 80K is likely to yield further gains in PA. Models that combine variable selection and shrinkage performed slightly better than shrinkage-only methods. We also find that the use of several markers per LD block increases PA relative to the use of 'independent' markers when P is large. We conclude that our ability to predict complex traits will continue to improve as more data becomes available; however, even with hundreds of thousands of records a sizable fraction of missing heritability will still remain.

1628T

Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *H.K. Finucane[1,2], Y.A. Reshef[1,3], A. Saunders[4], E. Macosko[4], K. Slowikowski[5], A. Byrnes[6,7], V. Anttila[4,7], S. Gazal[1], A. Gusev[1], G. Genovese[7], S. Pollack[1], J.R.B. Perry[8], S. Raychaudhuri[4], B. Neale[6,7], S. McCarroll[4], A.L. Price[1].* 1) Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Mathematics, MIT, Cambridge, MA; 3) Department of Computer Science, Harvard University, Cambridge, MA; 4) Harvard Medical School, Boston, MA; 5) Bioinformatics and Integrative Genomics, Havard University, Cambridge, MA; 6) Massachusetts General Hospital, Boston, MA; 7) Broad Institute, Cambridge, MA; 8) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK.

Genetics can provide an unbiased approach to discovering causal tissues and cell types for disease. Identifying these tissues is critical for developing ex vivo human systems to explore gene regulatory mechanisms that may be crucial to disease, and to follow up non-coding allelic function. Here, we analyze gene expression data from the GTEx Consortium and other sources, together with GWAS summary statistics for 52 diseases and traits with an average sample size of 83,751, to identify disease-relevant tissues and cell types. To accomplish this, we use stratified LD score regression to test whether disease heritability is enriched within 100kb of the top 10% of specifically expressed genes in a given tissue. We detected tissue-specific enrichments at FDR < 5% for 37 traits. We find that regions containing genes specifically expressed in arteries are enriched for heritability of migraine without aura, supporting a theory that migraine etiology has a vascular basis. Similarly, regions containing brain-specific genes are enriched for epilepsy heritability; this is particularly notable as there are only two genome-wide significant loci for this trait. For smoking behavior, we find the strongest heritability enrichment in regions containing genes that are specifically expressed in the nucleus accumbens, which is known to be important in addiction. Additionally, we find heritability enrichment for age at menarche in regions containing pituitary-specific and brain-specific genes; for schizophrenia, bipolar disorder, anorexia, Tourette syndrome, and major depressive disorder in brain-related regions; for rheumatoid arthritis and several other immunological traits in immune-related regions; for LDL and triglycerides in liver-related regions; and for waist-hip ratio in adipose-related regions, strengthening confidence in our method. We also applied this approach to single-cell expression data in two human and three mouse brains generated using Drop-seq. For schizophrenia, Tourette syndrome, smoking status, and other brain-related traits, we detected significant enrichments for neuronal cells relative to other brain cell types. Overall, our results suggest that gene expression data provides a valuable source of tissue-specific and cell-type-specific information for interpreting GWAS signals, and that a polygenic approach to detecting such enrichments can be more powerful than previous approaches based only on top SNPs.

1629F

Ygen: The first systematic assessment of the influence of human Y chromosome variation on a wide range of health-related traits. *I. Gandin[1], P.K. Joshi[2], T. Esko[3,4,5], J.F. Wilson[2,6], The Ygen consortium.* 1) Department of Medical Sciences, University of Trieste, Strada di Fiume 447 - Osp. di Cattinara, Trieste, 34149, Italy; 2) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland; 3) Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge Center 7, Cambridge, 02242, MA, USA; 5) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Cambridge, 02141, MA, USA; 6) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland.

Functions beyond sex determination and spermatogenesis have recently been suggested for the Y chromosome, however in terms of complex traits it remains the most poorly studied large block of DNA in the human genome. This is in large part due to the minimal representation of Y variants on genome-wide and custom SNP arrays, and further that these haploid data are routinely ignored in genome-wide association studies. The Ygen consortium leverages the inclusion of variants capturing the major branches of the Y chromosome gene genealogy into the widely genotyped exome chip, to perform the first systematic dissection of the influence of Y-linked variation on complex traits. Over 100 population-based and other cohort studies from 5 continental ancestries participate in Ygen, delivering the required statistical power with over 300,000 men. Genotypes at 68 markers, which represent ~4000 variants in perfect LD, are used to categorise men into 65 haplogroups, thus capturing the majority of global Y chromosome variability. 39 traits of public health and evolutionary interest are considered – including anthropometry, lipids, liver, lung and kidney function, cognition, glycaemia, fertility, inflammation, electrocardiography and haematology – and analysed in a single-SNP association analysis. We detail in high-resolution the spectrum of diverse haplogroup frequencies across the different cohorts. There are many challenges to the analysis of such strongly structured data, but initial meta-analyses suggest Y chromosome variants associate with adult height, the first evidence from large scale studies that the non-recombining region of the Y chromosome influences a complex trait. The yield per megabase of common "GWAS variants" on the Y chromosome appears to differ from that on the autosomes, suggesting the different selective exposures, effective population size and recombinational environments to the autosomes have influenced the function of Y DNA. .

1630W

Joint fine mapping of GWAS and eQTL detects target gene and relevant tissue. *F. Hormozdiari[1], A. Segre[2], M. van de Bunt[3,4], X. Li[2], J. Wha J Joo[1], M. Bilow[1], J. Sul[5,6], S. Sankararaman[1], B. Pasaniuc[7,8], E. Eskin[1,8].* 1) Department of Computer Science, University of California, Los Angeles, CA; 2) Cancer Program, The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA; 3) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 5) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA; 6) Semel Center for Informatics and Personalized Genomics, University of California, Los Angeles, CA; 7) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA; 8) Department of Human Genetics, University of California, Los Angeles, CA.

The vast majority of genome-wide association studies (GWAS) risk loci fall in non-coding regions of the genome. One possible hypothesis is that these GWAS risk loci alter the disease risk through their effect on gene expression in different tissues. In order to understand the mechanisms driving a GWAS risk locus, it is helpful to determine which gene is affected in specific tissue types. If the same variant responsible for a GWAS locus also affects gene expression, the relevant gene and tissue may play a role in the disease mechanism. Identifying whether or not the same variant is causal in both GWAS and eQTL studies is challenging due to the uncertainty induced by linkage disequilibrium (LD) and the fact that some loci harbor multiple causal variants. However, current methods that address this problem assume that each locus contains a single causal variant. In this paper, we present a new method, eCAVIAR, that is capable of accounting for LD while computing the quantity we refer to as the colocalization posterior probability (CLPP). The CLPP is the probability that the same variant is responsible for both the GWAS and eQTL signal. eCAVIAR has several key advantages. First, our method can account for more than one causal variant in any loci. Second, it can leverage summary statistics without accessing the individual genotype data. We use both simulated and real datasets to demonstrate the utility of our method. Utilizing data from the Genotype-Tissue Expression (GTEx) project, we demonstrate that computing CLPP can prioritize likely relevant tissues and target genes for a set of Glucose and Insulin-related traits loci.

1631T

Replication of genetic associations with height as a proof of concept for the Million Veteran Program (MVP) dataset. *H. Hunter-Zinck[1], Y. Shi[1], M. Li[1,5,11], N. Sun[2,6], J. Huang[1], C. Pan[3,7], J. Gelernter[2,10], P.S. Tsao[3,7], C.J. O'Donnell[1,9], E.R. Hauser[4,8], H. Zhao[2,6,10], T.L. Assimes[3,7], S. Pyarajan[1,9] for MVP Genomic Working Group.* 1) VA Boston Healthcare System, Boston, MA; 2) VA Connecticut Healthcare System, West Haven, CT; 3) VA Palo Alto Health Care System, Palo Alto, CA; 4) Durham VA Medical Center, Durham, NC; 5) University of Utah, Salt Lake City, UT; 6) Yale School of Public Health, New Haven, CT; 7) Stanford University School of Medicine, Stanford, CA; 8) Duke University, Durham, NC; 9) Harvard Medical School, Boston, MA; 10) Yale School of Medicine, New Haven, CT; 11) Johns Hopkins University, Baltimore, MD.

Height is a complex trait with polygenic architecture. Much effort has been spent to assemble large samples to better understand the genetic basis of this trait. We performed a genetic association analysis with height using approximately 110,000 samples of inferred European ancestry from the Million Veteran Program (MVP), from the U.S. Department of Veterans Affairs, to replicate previous associations and serve as a proof of concept for genome-wide association studies (GWAS) of other conditions using this dataset. We used data derived from the electronic health record of veteran participants, collected at the time of a clinical visit, for height and other covariates including sex and age. To assess replicability with the MVP dataset, we calculated the phenotypic variance explained by associated markers as well as the correlation in the direction of effect for previously associated markers in height meta-studies from Lango Allen, et al. in 2010 and Wood, et al. in 2014. Within the European ancestry cohort, we explain 8.1% of the variance in the height phenotype, a level comparable to previous studies. Using the overlap between markers on the MVP chip and markers analyzed in previous meta-studies, we replicate the direction of effect for 146 of 146 previously associated markers in the Lango Allen, et al. height GWAS and 135 of 136 markers associated in the Wood, et al. study. The Pearson correlation in effect sizes between the MVP study and the two previous analyses is 0.97 and 0.95 for the Lango Allen, et al. and Wood, et al. study, respectively. We also compared power and performance between self-reported and measured height to construct the phenotype as well as self-reported and inferred genetic ancestry to construct sample cohorts on which to conduct the GWAS. Finally, given the large numbers of individuals from other ancestry groups, we replicated associations found in the MVP cohort of European ancestry in a cohort of individuals admixed for European and African ancestry. The results strongly replicate previous genetic associations with height and serve as a demonstration of the utility of the MVP dataset for GWAS.

1632F

Providing a comprehensive GWAS-based genetic risks to healthy individuals for disease prevention in an academic hospital. *K. Ishikawa[1], K. Numakura[2], K. Aisaka[2], H. Kikuchi[2], T. Tanaka[2], T. Tanaka[3], A. Araki[1], S. Tanaka[4], T. Furukawa[5], M. Yoshida[6], E. Kanda[6], H. Mizusawa[7], Y. Ogawa[8], R. Bouchi[8], K. Ono[9], K. Moroboshi[9], K. Saito[10], H. Sakashita[11], T. Ishikawa[12], T. Nakagawa[13], T. Asakage[14], Y. Nakajima[15], M. Inokuchi[15], T. Saito[16], Y. Ebana[5,6], H. Numabe[17], J. Inazawa[3].* 1) Center for Personalized Medicine for Healthy Aging, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 2) P5, Inc., Tokyo, Japan; 3) BioResource Center, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 4) Dept. Molecular Oncology, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 5) Dept. Bio-informational Pharmacology, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 6) Dept. Life Science and Bioethics, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 7) National Center for Neurology and Psychiatry, Kodaira, Tokyo, Japan; 8) Dept. Molecular Endocrinology and Metabolism, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 9) Dept Ophthalmology and Visual Science, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 10) Dept. Urology, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 11) Dept. Clinical Oncology, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 12) Dept. Surgical Specialities, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 13) Dept. Surgery, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 14) Dept. Head and Neck Surgery, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 15) Dept. Gastrointestinal Surgery, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 16) Dept. Rheumatology, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 17) Faculty of Core Research Natural Science Division, Ochanomizu University, Tokyo, Tokyo, Japan.

Background: As one of the preventive healthcare services, direct-to-consumer genetic tests are widely provided. Most of these tests are based on risk SNP markers from genome-wide association studies (GWAS) and genetic counseling of family histories are not usually provided. Although common diseases are frequently observed in family history, GWAS based genetic risks are tended to be weak. Thus the frequency and heritability of common diseases are need to be considered in interpretation. In this presentation, we will report the results of a clinical research about preventive healthcare program using GWAS based genetic test at Medical Hospital, Tokyo Medical and Dental University, Japan. **Methods:** 101 Japanese adults were recruited for this study subjects. Genome-wide SNP genotyping, family history questionnaire and health examination were performed. Then, we created reports about genetic risks of about 30 diseases and health examinations, in which congenital hereditary diseases are not included. All the reports were disclosed by physicians. Selections of genetic risk markers and creating a consultation guideline for high risk groups were discussed at the clinical advisory board meetings, where specialists from each clinical department are attended. **Results:** In relatively rare diseases such as ulcerative colitis, Graves' disease and gout, GWAS based genetic risks were significantly high in subjects with family histories compared to subjects without family histories (p-value < 0.05). On the other hand, in relatively common diseases such as hypertension and type 2 diabetes, GWAS based genetic risks were not significantly different by family histories. **Conclusion:** Subjects with family histories and high GWAS based genetic risks especially need follow-ups as high risk groups. For example, in Graves' disease, one of the subjects whose genetic risk was high commented that his results were acceptable with his family history and looked convinced. We expect those subjects keep motivated to change their behavior for prevention. Meanwhile, subjects with no family histories but high GWAS based genetic risks or subjects with abnormal results in health examination are also need follow-ups as high risk groups. These results suggest the combination of genetic risks, family histories and health examination results are the challenges of how to disclose genetic risks and also the potentials to better follow-up interventions.

1633W

Pleiotropy in complex diseases estimated from family health history collected in primary care. *R.A. Myers[1], L.A. Orlando[1], R.R. Wu[1,2,3], A.H. Buchanan[4], G.S. Ginsburg[1], E.R. Hauser[5,6].* 1) Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, NC, USA; 2) Duke-National University of Singapore Medical School, Singapore; 3) National Heart Centre Singapore, Singapore; 4) Genomic Medicine Institute, Geisinger Health System, Danville, PA , USA; 5) Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, USA; 6) Cooperative Studies Program Epidemiology Center, Durham VAMC, Durham, NC, USA.

Complex diseases typically do not occur in isolation and are often accompanied by other conditions in individuals. Likewise, these sets of pleiotropic conditions may occur in families. The extent to which complex diseases are co-inherited and have shared genetic etiologies is informative in a number of applications such as patient disease-risk identification. PheWASs have been successful at identifying pleiotropic loci but do not quantify the extent to which pleiotropy exists in pairs or sets of complex diseases. We used family health histories (FHH) collected in primary care as a source of unascertained pedigrees to explore evidence for pleiotropy within families. We hypothesize that systematically collected, self-reported FHH can be used to quantify complex disease pleiotropy and describe the genetic relationships between complex diseases, thereby informing clinically relevant disease risk. MeTree® was used to collect FHH of 1608 patients for 48 different diseases. Patients averaged 25 relatives per family, 2.4 diseases for themselves, 1.3 for 1st degree relatives, and 0.6 diseases for 2nd degree relatives. We estimated heritability (h^2) using generalized linear mixed effect models and identified 39 diseases with h^2 >0, with values generally consistent with published h^2 estimates. We used logistic regression to estimate the odds ratio (OR) of a patient having disease A and a 1st, or 2nd degree relative having disease B, adjusted for age, sex, race, and family size, for all pairwise disease combinations. This measure of co-inheritance was positively correlated with h^2 for within-disease ORs (adj-r = 0.59, p = 0.0003). Genetically related sets of diseases were identified by using the ORs from disease-pairs with false discovery rate ≤ 5% to construct a network of diseases. The resulting network averaged 5.1 connections per disease and ranged from 0-13 connections per disease. There were several expected connections within physiologically related cancers, metabolic syndrome diseases, and autoimmune disorders (e.g. rheumatoid arthritis (RA), and lupus). The diseases with the largest number of connections were thyroid disease (13), RA (12), polyp count (11), and hypertension (11), suggesting the potential for novel pleiotropic relationships. Through analysis of unascertained FHH, we identify complex diseases with shared genetic etiologies that may help clinicians more comprehensively ascertain complex disease risk in families.

1634T

Steady-state gene expression explains 5-15% of heritability for many complex traits. *L.J. O'Connor[1,2], S. Gusev[1], X. Liu[1], H.K. Finucane[1,3], A.L. Price[1].* 1) Harvard University, CDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MAambridge, MA; 2) Program in Bioinformatics and Integrative Genomics, Harvard University, Cambridge, MA; 3) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

Disease risk variants identified by GWAS are predominantly noncoding, suggesting that gene regulation plays an important role. eQTL studies in healthy individuals have often been used to link disease-associated variants with the genes they regulate, relying on the hypothesis that pathogenic patterns of gene expression include changes in steady-state expression levels. To test this hypothesis, we developed a method, GE co-score regression, to estimate the proportion of complex trait heritability explained by the cis-genetic component of assayed gene expression levels. This quantity measures the extent to which genetic effects on complex traits are mediated by cis-regulation of gene expression. Our method is based on the idea that, for a gene whose expression is associated with a phenotype, SNPs with similar effects on the expression of that gene will have similar phenotypic effects. Specifically, we regress products of trait z-scores, for pairs of SNPs in linkage equilibrium, on products of cis-eQTL z-scores. We verified that GE co-score regression produces unbiased estimates in simulations.We applied GE co-score regression to eQTL data in 42 tissues from the GTEx consortium (N=79-362) in conjunction with GWAS summary statistics for 24 complex traits (median N=70,000), estimating the proportion of heritability explained by the union of all tissues. Across traits, the median estimate was .10 (median s.e.=.05, median p=.02, aggregate p < 10^{-10}); for 18/24 traits, the estimate was between .05 and .15. There was little evidence of tissue specificity, consistent with high cis-genetic correlations across tissues. We also analyzed splicing QTL data in brain (N=452), which explained limited additional heritability (median 0.036, median s.e.=0.02, median p=.02, aggregate p< 10^{-10}). These results suggest that pathogenic expression profiles do include a steady state component but are primarily transient or context- and cell type-specific; more targeted eQTL study designs will be required before GWAS results can routinely be translated into biological insights using gene expression data.

1635F

Enrichment of regulatory variants in the PheWAS Catalog: An implication of phenome-wide association studies. *Z. Zhao[1,2], J. Zhao[1], F. Cheng[3,4], P. Jia[1].* 1) School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX; 2) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 3) Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; 4) Center for Complex Networks Research, Northeastern University, Boston, MA.

   Genome-wide association study (GWAS) has evolved the primary approach to study the association of common genetic variants, typically in the form of single nucleotide polymorphisms (SNPs), with the complex diseases and traits during the last decade. So far, GWAS have identified thousands of SNP-trait associations, most of which are deposited in the GWAS Catalog. Recently, thank to the rapid accumulation of enriched phenotypic data from electronic medical records (EMRs), phenome-wide association studies (PheWAS) has become a feasible and a complementary approach to GWAS in order to detect genetic variants that are associated with multiple disease risks. A PheWAS Catalog that collected the large PheWAS dataset was made available by Denny *et al.* in 2013, which included 3,144 SNPs in the GWAS Catalog. These two catalogs allow us to investigate whether PheWAS variants are more likely to have functional impacts leading to develop multiple human diseases or traits. In this study, we used the functional annotation data, including transcription factor (TF) based motifs, promoters, enhancers, eQTLs from major databases such as FANTOM5, ENCODE, Roadmap and GTEx to curate the disease SNPs in PheWAS Catalog and GWAS Catalog. We found that the disease-associated loci in PheWAS Catalog were significantly enriched with functional SNPs. Based on these functional annotations, we constructed a comprehensive gene regulatory network using TF – Promoter/Enhancer – Target Gene (TF-PE-TG) interactions. Then, we investigated the trait-SNP associations in PheWAS Catalog at the tissue-specific regulatory circuit level. Our results indicated that using functional annotations could significantly improve the power of detecting novel associations in PheWAS Catalog. We found a number of novel associations in PheWAS Catalog that showed strong functional evidence even though they only had moderate significance in the original study. Our work enhanced the understanding of the associations in PheWAS Catalog in terms of their mechanisms on affecting multiple diseases and traits.

1636W

Rare copy number variants in over 100,000 subjects reveal novel disease and genomics element associations. *Y.R. Li[1,2,3], J.T. Glessner[1,4,5], B.P. Coe[6], J. Li[1], X. Chang[1], C. Kao[1], A. Cederquist[1], C. Kim[1], M. Khan[1], F. Mentch[1], M. Garris[1], D. Abrams[1], F.G. Otieno[1], P.A. Sleiman[1,2,7], E.E. Eichler[8], H. Hakonarson[1,2,7].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA; 3) Medical Scientist Training Program, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA; 4) Psychiatric and Neurodevelopmental Genetics Unit, Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 5) Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; 6) Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, 98195, USA; 7) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 8) Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington; Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, Washington, 98195, USA.

   Copy number variations (CNVs) have a significant impact on the diversity of the human genome and loci bearing CNVs are known to affect gene function and are associated with common and rare human diseases. We report the 'CNV landscape' across the genome of 100,028 unrelated individuals of European ancestry, based on genome-wide SNP and CGH arrays. We observed over 1.7×10^6 CNVs, averaging a 650kb CNV burden/individual and mapping to 11,314 deletion, 5,625 duplication, and 2,746 homozygous-deletion (hd) CNV Regions (CNVRs). Most CNVRs are rare (freq<0.01) and recurrent (>98.5%), although 25% of hdCNVRs are private. Overall, 3.25% of identified CNVs are novel. Over 58% of CNVRs overlap at least one gene, and were enriched (P<1x10-4) for OMIM morbid genes (enrichment ratio 'ER'=2.94), GWAS loci (ER=1.52) and non-coding RNAs (ER=1.44). The CNVR-bearing loci were also strongly enriched (35.3%) for recombination hotspots (ER>10; P<1x10-4), underscoring a potential impact of natural selection in driving human genome diversity and CNV distribution. In contrast, rates of overlap with microsatellites and segmental duplications were the same as observed at random. The CNVRs identified associated with 4 major disease categories including autoimmune/ inflammatory (n=11,489), oncologic (n=9,105), congenital heart/metabolic (n=2,581), developmental delay (n= 29,085)/neurological (n=14,756) diseases. Of 131 significant disease-associated CNVRs (DA-CNVRs) identified, 18% overlap GWAS hits (ER=2.40 and 1.20, P<1x10-4); these include a novel deletion CNVRs associated with autoimmune disease mapping to chr7p15.3 (*ITGB8*), which encodes the cell surface glycoprotein beta-8 integrin ($P < 6.87 \times 10^{-19}$). Another deletion CNVR ($P < 6.53 \times 10^{-47}$) associated with neurological diseases maps to chr19p13.3 (*HCN2*), which encodes a hyperpolarization-activated, cyclic nucleotide-gated K+ channel expressed in the brain. Finally, we observed multiple DA-hdCNVRs, including an autoimmune disease-associated locus at chr2q34 interrupting the exonic region of *ERRB4*, a cell-surface receptor tyrosine kinase whose expression is attenuated in microglia of patients with relapsing-remitting MS. Our work encompassing the genome wide CNV landscape of >100K individuals underscore the value of large-scale, genome-wide CNV analysis and the need to consider variants common and rare CNVs as well as SNVs in understanding the genetic contribution to complex disease.

1637T

Integration of functional annotation with GWAS to build disease-specific pathogenic scores. *K. Burch[1], G. Kichaev[1], B. Pasaniuc[1,2,3].* 1) Bioinformatics IDP, University of California Los Angeles, Los Angeles, CA, USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA.

Genome-wide association studies (GWAS) have provided invaluable insight into the genetic architecture of many complex traits, with recent works showing that functional elements of the genome are enriched with GWAS signals. The large compendium of functional annotations that describe regulatory features, conservation, and other sequence-based metrics can be used to construct integrative genetic models of disease architecture. CADD is one example of a variant scoring method that combines multiple, diverse annotations into a single score of deleteriousness (Kircher et al. 2014). However, current measures of functionality are calculated from information that is independent of GWAS data, and therefore cannot provide scores that are specific to a complex trait under investigation. Here, we describe a systematic, integrative analysis of 39 complex traits to produce disease-specific functionality scores. For each trait, we construct functional models by integrating GWAS summary data and functional annotations to deliver measures of pathogenicity readily interpreted as the prior probability that a SNP impacts the trait. We observe that pathogenicity indices correlate across related traits and find a number of regions in the genome that are enriched for pathogenic variants for many traits. We believe these scores will be useful for follow-up study design – particularly in the context of fine-mapping. We illustrate this by conducting integrative fine-mapping on over 2000 GWAS loci and observe that leveraging these disease-specific scores yields a marked reduction in the size of the 90% credible SNP set. Notably, we find hundreds of regions that have experimentally tractable credible set resolution of less than 10 SNPs that can serve as candidates for empirical validation in functional studies. In conclusion, we believe our atlas of disease-specific pathogenicity scores and fine-mapping credible SNPs will be a valuable resource for the genetics community.

1638F

Combination of promoter and coding region of *N-acetyltransferase 2* **(*NAT2*) have stronger effect in predicting anti-tuberculosis drug induced liver injury (DILI).** *R. Yuliwulandari[1,2], K. Prayuni[2], R.W. Susilowati[1], M. Muslikh[3].* 1) Faculty of Medicine, YARSI University, Indonesia, DKI Jakarta, Indonesia; 2) YARSI Research Institute, YARSI University, DKI Jakarta, Indonesia; 3) Faculty of Economic, YARSI University, DKI Jakarta, Indonesia.

Tuberculosis (TB), as one of global public health problems, needs to be treated and eradicated well. TB treatment needs a combination of several drugs, including isoniazid, and long term treatment, which can lead to severe side effect such as anti-tuberculosis drug induced-liver injury (AT-DILI). N-acetyltransferase 2 (NAT2) has been reported to have a role in isoniazid metabolism. NAT2 metabolic activity is classified into slow, intermediate and rapid acetylators. Our previous study reported that polymorphisms in the promoter of NAT2 may affect the NAT2 activity. Therefore, in the current study, we would like to examine the effect of the promoter polymorphisms in the susceptibility to develop AT-DILI, alone and in combination with the coding region haplotype. In total, 95 AT-DILI cases and 192 TB tolerant controls with Javanese, Sundanese and Betawi ethnics of Indonesia participated in this study. The DNA was extracted using Qiamp blood mini kit. The selected promoter SNPs was rs4646246 and was genotyped using Taqman assay (this SNP has been reported to affect NAT2 activity in the previous study). The result showed that rs4646246 has a significant association with AT-DILI (P=0,0001, odds ratio (OR)=2.27 (1.49-3.47)). As our previous study reported that among coding region haplotypes, only NAT2*6A showed significant association with AT-DILI (P=7.7 × 10[-4], OR=4.75 (1.8–12.55)), then we perform combined promoter and coding region haplotype analysis with focus on the combination between rs4646246 with NAT2*6A. We found that the combination of promoter and coding region of NAT2 showed stronger effect to the development of AT-DILI (P=0.0016, OR=5.02 (1.81-13.93)). This result is in accordance to the hypothesis in the previous reported study that there is an additional role of promoter of NAT2 in the prediction of AT-DILI. Therefore, we propose that screening of AT-DILI in TB patient before administering anti-tuberculosis drugs needs to be done, not only for the haplotype in the coding region, as commonly reported, but also include promoter region.

1639W

Prevalence of allelic heterogeneity in complex traits. *E. Eskin[1,4], F. Hormozdiari[1], A. Zhu[1], G. Kichaev[2], B. Pasaniuc[3,4], S. Shifman[5].* 1) Department of Computer Science, University of California, Los Angeles, California; 2) Bioinformatics IDP, University of California, Los Angeles; 3) Department of Pathology & Laboratory Medicine, University of California, Los Angeles; 4) Department of Human Genetics, University of California, Los Angeles, California; 5) Department of Genetics, The Life Sciences Institute, Edmond J. Safra Campus, The Hebrew University.

   Allelic heterogeneity (AH) is a phenomena where more than one variant in the same gene affects a phenotype. The Genotype-Tissue Expression project data provides an unprecedented opportunity to quantify the amount of allelic heterogeneity across tissues. The current approach for detecting the allelic heterogeneity is based on conditional analysis used in the context of fine-mapping. The conditional approach relies on an iterative selection of the most associated variants followed by re-computation of the statistical score for the remaining variants conditional on the already selected variants. The conditional process continues until statistical score for all the variants falls below a pre-defined significant threshold. As we show in our experiments, the conditional approach is sensitive to complex patterns of linkage disequilibrium and can often result in inaccurate predictions. In our work, we propose a novel probabilistic method to quantify allelic heterogeneity. Our method is an extension of our fine-mapping method, CAVIAR. The benefit for our method is that it provides a probability for different number of variants that affect the phenotype. Thus, we can assign probability to each gene which it harbors allelic heterogeneity and as a result we can have a confident level for our prediction of genes which harbors allelic heterogeneity. We assess the performance of our method using simulated and real datasets. Our simulation results show that our method has high confidence in detecting the allelic heterogeneity loci. In addition, we apply our method to GTEx datasets that consist of nine tissues, we observe multiple allelic heterogeneity loci in different tissues. We considered genes which are detected as allelic heterogeneity for both methods. We observe both tissue-specific and cross-tissue allelic heterogeneity. The software for our method is freely available for download at http://genetics.cs.ucla.edu/caviar/.

1640T

GWAS identifies two novel loci associated with daily consumption of spicy food in Chinese adults. *R.G. Walters[1], M.V. Holmes[1], W.X. Mei[1], Y. Guo[2], K. Lin[1], S. Rafiq[1], J. Lv[3], I.Y. Millwood[1], Z. Biang[2], L. Lee[2,3], Z. Chen[1] on behalf of the China Kadoorie Biobank Collaborative Group.* 1) Nuffield Department of Population Health, University of Oxford, Oxford, Oxon, United Kingdom; 2) Chinese Academy of Medical Sciences, Beijing, China; 3) School of Public Health, Peking University Health Science Center, Beijing, China.

   Background: A large cohort study in China recently showed that regular consumption of spicy food may have significant health benefits. However, the causal relevance of the observed associations is unclear. Capsaicin, the active component of chilli peppers, is responsible for a burning sensation following contact with mucous membranes. We hypothesised that genetic variation may affect sensing of or response to capsaicin (e.g. through changes in neural development or taste/olfactory receptor function/signalling), leading to a liking for or aversion to spicy foods. Such variation might enable the construction of a genetic score for use in Mendelian randomisation, to investigate whether spicy food consumption is causally protective against disease. **Methods:** The China Kadoorie Biobank is a prospective study of 512,000 adults recruited from 10 geographically diverse areas of China. At baseline, 30.1% of participants reported daily consumption of spicy food. We conducted genome-wide association analyses of the frequency of spicy food consumption, among 33,000 subjects genotyped using a custom Affymetrix Axiom® array. Additive per-allele effects were tested using linear regression against days-per-week of spicy food consumption, with sex and 10 principal components as covariates. To control for differences in local cuisine, the 10 regions were analysed separately and the results combined using fixed-effects meta-analysis with inverse-variance weighting. **Results**: Of the 517,443 autosomal markers passing quality control, two SNPs were close to genome-wide significance and reached array-wide significance ($P < 7 \times 10^{-8}$). One of these lies within the *PYGO1* gene, which has a potential role in the development of taste buds and papillae through its involvement in Wnt/β-catenin signalling. A further 18 independent SNPs were significant at a 5% false discovery rate. An unweighted gene score constructed from these 20 SNPs explained 0.4% of variance in the frequency of spicy food consumption (F-statistic=124). **Conclusions:** This first large GWAS of spicy food consumption has identified novel SNPs potentially impacting on food preference. Construction of a genetic score from these SNPs yields a strong instrument that will enable future Mendelian randomization analyses.

1641F

Functional properties of genomic region associated with multiple traits in genome-wide association studies. *Y.F. Wang, M. Guo, T.Y. Wang, Y. Zhang, J. Yang, Y.L. Lau, W. Yang.* Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.

More than 10,000 associations with over 600 human complex traits have been identified by genome-wide association studies (GWAS) in the past years. An interesting observation is that many genetic loci emerged as associated with multiple distinct traits. Although the pleiotropic phenomena have been examined on certain variants or genes, comprehensive characterization of these loci is still lacking. In this study, we collected a plethora of trait-associated variants from the National Human Genome Research Institute (NHGRI) Catalogue of Published Genome-wide Association Studies, and defined cross-phenotype (CP) loci as a region within which variants are associated with at least two phenotypes. We observed that nearly 25.36% of the GWAS loci are associated with multiple traits, accounting for 7.79% of human genome, suggesting a widespread existence of CP loci across the human genome. Besides, we showed that the functional markers such as DNase-I hypersensitive sites, transcription factor binding sites and histone modification regions are highly enriched in CP regions, much more so than single trait-associated loci or random genomic controls, suggesting that these loci may contain more functional elements and these elements may play a broader role in genetic regulation. Further study observed the variants within the CP loci are more likely to be eQTLs and are more likely to be located in early DNA replication regions, suggesting that distinct regulatory and functional elements may exist within these loci. Finally, through comparing the genetic overlapping between diseases, we have explored the shared and distinct mechanistic patterns and pathways among different diseases, which may have significant clinical implications and be informative on drug repositioning.

1642W

Identification of individuals by trait prediction from the genome. *C. Lippert[1], M.C. Maher[1], O. Arikan[1], E.Y. Kang[1], S. Lee[1], R. Sabatini[1], A. Harley[1], A. Bernal[1], P. Garst[1], V. Lavrenko[1], K. Yocum[1], T. Wong[1], S. Ramakrishnan[1], H. Tang[1], C. Xie[3], J. Piper[3], S. Brewerton[3], Y. Turpaz[2,3], A. Telenti[2], R.K. Roby[2,4], J.C. Venter[2,4], F.J. Och[1].* 1) Human Longevity, Inc., Mountain View, CA, USA; 2) Human Longevity, Inc., San Diego, CA, USA; 3) Human Longevity Singapore, Pte. Ltd., Singapore; 4) J. Craig Venter Institute, La Jolla, CA, USA.

Accurate prediction of human physical traits and demographic information from genomic data promises to revolutionize forensic genomics. To explore the capabilities of phenotype-based genomic identification, we applied whole-genome sequencing, detailed phenotyping, and statistical modeling to predict biometric traits in an ethnically diverse cohort of 1,061 participants. For a large fraction of traits, their complex genetic architecture limits predictive accuracy. However, we developed a novel framework to integrate these predictions for multiple phenotypes and re-identify a high percentage of held-out individuals in our cohort, as well as the first ten participants of the Personal Genome Project. We developed state-of-the-art predictive models for prediction of 3D facial structure and color, eye color, skin color, height, weight, BMI, age, and voice from genomic features. On this cohort, we achieved the following cross-validated fraction of variance in the trait explained by the predictive model (R^2): for face color R^2 of 0.74, for face depth R^2 of 0.30, for eye and skin color R^2 of 0.78-0.82 (for RGB), for age R^2 of 0.46, for height R^2 of 0.59 and weaker performance for the prediction of weight and BMI ($R^2=0.21$ and $R^2=0.12$, respectively). Next, we integrated each of the individual phenotypic prediction models to obtain a model for identification from genomic information. Given an individual's genomic sample, we sought to identify that individual out of a lineup of 10 individuals. We refer to this problem as *select at 10* (s_{10}). We found that for re-identification problem 3D face alone is highly informative, with a s_{10} value of 58%. We found that ethnicity was the second most informative feature, with a s_{10} performance of 50%. Voice had comparable performance to ethnicity, while age, height/weight/BMI, gender, and age all yielded s_{10} performance of around 20%. Finally, we integrated these variables to obtain s_{10} performance of 77%. The presented results have important implications for personal genomic privacy. We have shown that, by using our algorithms, it is possible to extract a genome from a database and identify the individual. If this is conducted for unethical purposes and for inappropriate uses, it compromises the identity of individuals who contributed their genomes. The scientific community is in need of dedicated programs to evaluate privacy in light of today's technological advances and ensure basic rights of all contributors.

1643T

Neurodevelopmental indications of increased vulnerability for executive dysfunction in boys with 47,XXY. *C. Samango-Sprouse[1,2], C. Keen[3], C. Cappello[3], T. Sadeghin[4], A. Gropman[5,6].* 1) Pediatrics, George Washington University, Washington, DC; 2) Molecular Genetics, Florida International University, Miami, FL; 3) The Focus Foundation, Davidsonville, MD; 4) Neurodevelopmental Diagnostic Center for Children, Crofton, MD; 5) Neurogenetics and Neurodevelopmental Pediatrics, Children's National Medical Center, Washington, DC; 6) Neurology, George Washington University, Washington, DC.

Introduction: Neuroimaging studies have found consistent neuroanatomical abnormalities in the frontal lobes of males with 47,XXY. Frontal lobes are critically pertinent to executive functions (EF), such as time management, planning, flexibility and rapidity of response. EF is an area of known dysfunction in 47,XXY, but neurodevelopmental indicators of vulnerability are less understood. Statement of Purpose: A small prospective pilot study was conducted to begin to characterize an algorithm of risk factors associated with greater vulnerability to impaired EF in males with 47,XXY. Methods: 15 boys with 47,XXY were referred for neurodevelopmental evaluations (median age=144mo.). Performances from the Comprehensive Trail Making Test (CTMT), visuomotor planning (VP) and motor planning (MP) of the Amsterdam Neuropsychological Tasks (ANT), Motor Coordination (MC) of the Beery-Buktenica Test of Visual Motor Integration, Fine Motor Control (FM) of the Bruininks-Oseretsky Test of Motor Proficiency, and Processing Speed Index (PSI) of the Weschler Scales of Intelligence for Children were evaluated to investigate EF and selected areas of neurodevelopment within EF. Results: 11 of the 15 47,XXY boys had composite scores less than 42 on the CTMT, indicating 73.3% of the study group had below average, mildly impaired, or severely impaired performance. Of these 11 boys, the majority also had performance below two standard deviations (SD) of average in FM (72.7%), MC (54.5%) and MP (63.9%). 45.5% were two SD below average in PSI and VM. Discussion: Our results suggest that impaired EF in 47,XXY may be highly associated with FM, MC and MP, but that no single assessment is able to diagnose executive dysfunction by itself. These results further suggest that comprehensive assessment is necessary to determine both capabilities and deficits in EF in boys with 47,XXY. These results further enhance our understanding of the most vulnerable children with 47,XXY.

1644F

LD-dependent architecture of human complex traits reveals action of negative selection. *S. Gazal[1], H.K. Finucane[1,2], P.R. Loh[1], P.F. Palamara[1], X. Liu[1], A. Schoech[1], B. Bulik-Sullivan[3,4], A. Gusev[1], B. Neale[3,4], A.L. Price[1,5].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Understanding the polygenic architecture of complex traits, and the underlying biological processes, is a fundamental goal of human genetics. Recent work has hinted at LD-dependent architectures for several complex traits, with larger per-SNP heritability for genetic variants with low levels of linkage disequilibrium (LD) to other variants after conditioning on minor allele frequency (MAF). However, this has not been formally assessed, quantified, or biologically interpreted. We quantified the relationship between MAF-stratified LD percentile (level of LD; LLD) and per-SNP heritability by analyzing GWAS summary statistics from 22 human complex traits (average N = 62,341). To enable this analysis, we extended stratified LD score regression, a method for partitioning heritability across discrete annotation categories, to continuous-valued annotations. We determined that SNPs with low LLD have significantly larger per-SNP heritability ($P = 4.15 \times 10^{-31}$). Roughly half of the LLD signal can be explained by functional annotations that are negatively correlated with LLD, such as DHS, histone marks and enhancers. By investigating a broad set of LD-related annotations, including allele age as predicted by the ARGweaver method, we determined that the remaining LLD signal is largely driven by MAF-stratified allele age ($P = 5.65 \times 10^{-56}$), with the youngest 20% of common SNPs explaining 3.8x more heritability than the oldest 20%. This is even larger than MAF-dependent effects, in which the 20% of common SNPs with largest MAF explain 1.9x more heritability than the 20% with smallest MAF. In a joint analysis, we also inferred significant effects for other annotations related to negative selection, including recombination rate, McVicker B statistic, and nucleotide diversity. In particular, we inferred a smaller causal effect size variance for SNPs in high recombination rate regions, opposite to the direction of the LLD effect but consistent with the Hill-Robertson effect. We used forward simulations to confirm that all of these findings are consistent with the action of negative selection on deleterious variants that affect complex traits. Finally, we analyzed 11 phenotypes from the UK Biobank (N = 124,388) to replicate our results ($P = 1.80 \times 10^{-54}$ and $P = 2.40 \times 10^{-77}$ for LLD and allele age, respectively). Our empirical findings complement efforts to learn about negative selection by analyzing much smaller rare variant data sets.

1645W

Shared genetic basis of complex and rare disease. *M. Kumar[1], V. Arboleda[2], N. Mancuso[2], B. Pasaniuc[1,2].* 1) Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA.

The biological mechanisms underlying complex genetic disease and monogenic Mendelian disease are typically considered to be distinct. This mindset is reflected in the statistical approaches used to link single nucleotide polymorphisms (SNPs) with the disease of interest. Traditionally, SNPs associated with complex disease are found by genome-wide association studies (GWAS), while genetic variants causing Mendelian diseases have been identified through linkage analysis and, more recently, exome sequencing. In this work, we investigate the shared genetic basis of traits that show both complex and Mendelian presentations (e.g., height). We used publically-available GWAS summary statistics from more than 30 complex traits to identify risk genes for complex traits. We quantified the overlap of this gene set with trait-matched Mendelian disease genes from the Online Mendelian Inheritance in Man (OMIM) catalogue. For example, in the case of height, we constructed a complex-trait gene list based on distance to GWAS risk variant and quantified the degree of overlap with OMIM genes showing height-specific pathogenicity (e.g., undergrowth, short stature, small for gestational age, dwarf, tall stature, failure to thrive, hemihypertrophy). We find a significant overlap between these gene sets consistent with shared genetic basis between complex and Mendelian forms of disease. Furthermore, we expanded this approach to an array of 30 complex traits and matched Mendelian disorders and find a consistent and significant overlap across all studied traits. Overall, our results support a shared genetic basis of complex and Mendelian traits and provide insight into genetic mechanisms of complex traits. These results can be leveraged to prioritize candidate genes surrounding significant GWAS SNPs and identify modifiers of rare disease phenotypes in both large-scale association studies and clinical exome sequencing.

1646T

Targeted pharmacogenetic sequencing among 9,000 eMERGE participants with linked electronic health records. *A. Gordon[1], D.R. Crosslin[2], P. Devi[2], A.A. Burt[1], B. Almoguera[3], B.L. Cobb[4], M. de Andrade[5], J.C. Denny[6], H. Hakonarson[7], J.B. Harley[8], M.G. Hayes[9], K.M. Kaufman[4], E.B. Larson[10], N.D. Leslie[8], B. Namjou-Khales[4], C.A. Prows[8], L.J. Rasmussen-Torvik[13], M.D. Ritchie[11], D.M Roden[6], S.L. Van Driest[6], W.Q. Wei[6], Q.S. Wells[6], K. Zhang[12], G.P. Jarvik[1], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 3) The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 5) Biomedical Statistics and Informatics, Mayo Clinic and College of Medicine, Rochester, MN; 6) Department of Medicine, Vanderbilt University, Nashville, TN; 7) Center for Applied Genomics, The Children's Hospital of Philadelphia, PA; 8) Department of Pediatrics, Cincinnati Children's Hospital, University of Cincinnati, Cincinnati, OH; 9) Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 10) Group Health Research Institute, Seattle, WA; 11) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 12) Biomedical Informatics, Vanderbilt University, Nashville, TN; 13) Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL.

As genome and exome sequencing becomes routine in clinical care, pharmacogenetic results are often cited as a motivation for implementing these tests. To this end, the Clinical Pharmacogenetics Implementation Consortium (CPIC) offers guidelines, denoted in 'star allele' terms, describing the evidence for specific gene-drug pairs. Although large sequencing efforts such as the NIH's Precision Medicine Initiative (PMI) will uncover many novel, deleterious variants, the phenotypic effects of these variants in combination with known alleles are unclear. To explore the extent and phenotypic effects of such variation, the Electronic Medical Records and Genomics Network (eMERGE) has deployed PGRNseq, a next-generation sequencing panel targeting 82 pharmacogenes, among 9000 diverse patients across 9 different sites. As of writing we have completed preliminary analysis of the first 5000 samples. As these data revealed that >95% of patients carried at least one variant designated as actionable by CPIC (Level A), analysis of the full dataset will yield accurate allele frequencies for alleles of known effect among individuals of diverse ancestry; as these frequencies are key components of pharmacogenetic interpretation algorithms, their accuracy in specific populations is critical. In addition to common alleles, we found at least one rare, predicted deleterious variant (CADD > 20) in 78 of 82 targeted genes, suggesting that the full dataset will even further expand this pool. To explore how these rare variants affect alleles of known function, we will use the phasing algorithm BEAGLE to construct individual diplotypes from 12 well-characterized genes for which CPIC guidelines exist. By assigning these haplotypes star allele names using commercially employed tag SNPs, we will identify individuals who carry a novel deleterious variant in combination with a level A allele for further analysis. As a strength of the eMERGE network is the linked electronic health record data linked to each participant, we will explore potential clinical consequences of these novel variants using approaches such as ICD code-based Phenome-Wide Association and expert chart review. The data presented here to quantify and evaluate individuals carrying rare variation with common CPIC alleles is an important model for massive clinical studies like PMI, which will deploy clinical decision support for pharmacogenetic results that is unlikely to provide accurate interpretations for these individuals.

1647F

Establishment of a pharmacogenomics testing platform using next-generation sequencing (NGS). *Y.S. Liu[1], Y.H. Lin[1], L.J. Shen[2,3], P.L Chen[1,4].* 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan; 2) Graduate Institute of Clinical Pharmacy/School of Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan; 3) Department of Pharmacy, National Taiwan University Hospital, Taipei, Taiwan; 4) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan.

Medical drugs show different efficacy and/or adverse drug reactions (ADRs) on patients. This is mostly due to the variations of DNA sequence. Pharmacogenomics is the study and applications about how genetic variations in individuals influence the drug response, which is composed of both pharmacokinetics and pharmacodynamics. Pharmacogenomics plays an important role in optimal drug choice and drug dosing. There are numerous genes involved in pharmacogenomics, which imposes a big challenge because of the complexity and high cost using conventional techniques like Sanger sequencing. In the present study, we set up a genetic testing platform through capture-based target enrichment followed by next-generation sequencing (NGS). Our panel covered approximately 100 major pharmacogenomics genes, including pharmacokinetics genes (for example, ADME genes regarding to absorption, distribution, metabolism and excretion) and pharmacodynamics genes, such as *ABCB1, CFTR, CYPs, DRYP, EGFR, HLAs, KRAS, NAT2, RYR1, TPMT, UGT1A1* and *VKORC1*. A great proportion of our genes overlapped with those listed on FDA labeled biomarkers and Pharmacogenomics Knowledgebase (PharmGKB). We applied this panel to 24 individuals, including 6 controls with whole genome sequencing data for technical validation and 18 individuals with ADR, trying to find specific ADR gene biomarkers. We inferred that dose-dependent adverse drug effects might have higher correlation with pharmacokinetics genes. Also, we analyzed Taiwan Biobank data to find out the reference allele frequencies in our population. We used various NGS software packages, including BWA, SAMtools, Picard, GATK, Variant Tools, ANNOVAR, IGV, etc. In conclusion, the NGS-based pharmacogenomics panel could be beneficial for precision medicine in clinical applications and academic research.

1648W

Clinical significance of genetic variation among genes involved in drug response: Pharmacogenomic interpretation in the ClinVar database. *S.A. Scott[1], G.C. Bell[2], J.S. Berg[3], U. Broeckel[4], K.E. Caudle[5], C.E. Haidar[5], E.M. Ramos[6], M.V. Relling[5], E.A. Rivera-Munoz[3], M. Whirl-Carrillo[7], K. Wiley[6], M.S. Williams[8], T.E. Klein[7], H.L. McLeod[9].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Mission Health, Fullerton Genetics Center, Asheville, NC 28803; 3) Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; 4) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI 53226; 5) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105; 6) Division of Genomic Medicine, National Human Genome Research Institute, NIH, Rockville, MD 20892; 7) Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305; 8) Genomic Medicine Institute, Geisinger Health System, Danville, PA 17822; 9) DeBartolo Family Personalized Medicine Institute, Moffitt Cancer Center, Tampa, FL 33612.

The ClinVar database archives reports of genomic variants and phenotypes submitted by clinical laboratories, researchers, clinicians, expert panels, practice guidelines, and other organizations. This repository enables more accurate and consistent variant interpretation among submitters, in part by providing detailed annotation and transparent supporting evidence for each catalogued variant. The clinical significance of variants implicated in Mendelian phenotypes in ClinVar is appropriately based on the standard terminology recommended by the American College of Medical Genetics and Genomics (ACMG); however, these widely used five categories do not effectively translate to genetic variants implicated in drug response phenotypes. To facilitate the evaluation of clinical significance for pharmacogenomic variants, the Pharmacogenomics Working Group of the Clinical Genome Resource (ClinGen), in collaboration with ClinVar, has proposed three categories of standard terminology to be used when evaluating genetic variation among genes involved in drug response – clinically actionable, clinically informative, and variants of uncertain clinical significance. Like the ACMG terminology for Mendelian variants, these terms are defined by levels of evidence supporting the association between a pharmacogenomic variant and a drug response phenotype. Variants previously curated by the Pharmacogenomics Knowledge Base (PharmGKB) are directly amenable to clinical significance translation as 'actionable', 'informative', and 'uncertain clinical significance' are defined as PharmGKB Level 1, 2 and 3/4 variants, respectively. Notably, interrogation of the star (*) alleles and/or known variants of seven genes (*CYP2C9*, *CYP2C19*, *CYP3A5*, *DPYD*, *HLA-B*, *SLCO1B1*, and *TPMT*) with Clinical Pharmacogenetics Implementation Consortium (CPIC) practice guidelines identified 49 actionable, 59 informative, and 107 variants of uncertain clinical significance. Presentation of pharmacogenomic variants in ClinVar will also be accompanied by allele function terms (when known) as recently defined by the CPIC Term Standardization Project. In addition to classifying known pharmacogenomic variants, utilization of these clinical significance terms in ClinVar will allow for more consistent pharmacogenomic variant interpretation by clinical laboratories and could facilitate more rapid agreement on the clinical significance of rare or novel pharmacogenomic variants as they are identified by ClinVar submitters.

1649T

The Million Veteran Program: A mega-biobank for precision medicine. *S. Muralidhar, MVP Executive Committee and Operations Teams.* Res & Development (10P9B), Dept Veterans Affairs, Washington, DC.

The Million Veteran Program (MVP), sponsored by the Department of Veterans Affairs, is partnering with Veterans to establish one of the world's largest databases of genomic, clinical, lifestyle and military exposure information. Leveraging VA's rich electronic health record (EHR), the ultimate goal of MVP is to create a comprehensive database for genomic discoveries that will lay the foundation for precision medicine for Veterans, and the population at large. In addition, and under the President's Precision Medicine Initiative, MVP is collaborating with the Department of Defense and expanding enrollment to active duty service members. Launched in 2011, consented participants: a) provide a blood sample for future, to-be-defined research; b) complete health and health-related questionnaires; c) provide access to their electronic health records; and d) agree to be contacted in the future. Recruitment is largely through mail invitations and publicity within 50+ VA medical centers; a web-enrollment portal is under development. Blood samples are processed at the MVP central biorepository promptly after receipt, and DNA is sent out to contracted vendors for characterization on a customized Affymetrix Axiom Array "MVP chip." Remaining DNA, buffy coat, and plasma are stored for future use. Quality control of genomic data is performed centrally by a core team. Self-reported data from questionnaires and EHR data are curated by phenomic teams; selected core variables are curated for general applications, and disease-specific variables are curated initially on a project-by-project basis. All data are stored in a central, secure data repository, and approved researchers access the data and conduct their analysis within a secure computing environment. Results of research analyses and the associated metadata, are retained and repurposed for future research. With the goal of enrolling one million participants, MVP has enrolled over 483,000 Veterans as of early June 2016. Over 400,000 samples have been genotyped, and quality control on the first 200,000 genotypes completed. Subsets of samples have been characterized by whole-exome or whole-genome sequencing. Curation of clinical data for core variables, such as demographic factors and laboratory values, has also been completed. Eight scientific projects utilizing MVP data are currently underway. An overview will be presented of the MVP infrastructure, data access policy, scientific projects, and lessons learned.

1650F

A software platform facilitating community analyses of genetic datasets for complex disease. *B. Alexander, M. Duby, M. Sanders, K. Yuksel, M. von Grotthuss, R. Koesterer, O. Ruebenacker, C. Gilbert, T. Green, J. Flannick, Accelerating Medicines Partnership for Type 2 Diabetes Knowledge Portal (AMP T2DKP). Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA.*

   While genome wide association studies (GWAS) have identified numerous variants of modest effect on complex disease risk, converting these findings into biological insights has proved challenging. The large quantities of data produced by different GWAS and sequencing studies too often remain disparate, and are inaccessible to the broader biomedical community.   Here we present a new platform to facilitate analyses of genetic data for complex disease. Our platform, which powers the Type 2 Diabetes Knowledge Portal (T2DKP), includes software for aggregating and harmonizing genetic association study results, along with highly customizable genotype-phenotype association analyses across data sets, all while protecting sensitive patient data. The platform allows users to perform: 1) *Forward genetic analyses:* visualize the top association statistics from any collection of GWAS or sequencing studies in either tabular or graphical form, and customize searches across datasets based on arbitrary association criteria. As an example, we demonstrate the identification of candidate association signals proposed to affect T2D through effects on cholesterol efflux from the liver based on searches across glycemic and lipid datasets. 2) *Reverse genetic analyses:* return a list of association statistics for any variant, gene, or region across all datasets, with harmonized directions of effect, optionally drilling down to explore ethnic- or cohort-specific effects. As an example, we demonstrate the identification of secondary protective effects on T2D for a variant implicated in coronary artery disease (CAD) risk based on a simple search for the variant. 3) *Custom association analysis:* specify a phenotype, and optionally a property for stratifying samples, along with sample filters, and phenotype or property covariates, then dynamically perform association tests on one or more variants. As an example, we demonstrate the prediction of a potential method of action for a T2D association based on interactive analysis of patients with a specific disease subtype. The T2DKP is built upon this platform, and is freely and publicly available at www.type2diabetesgenetics.org. The T2DKP draws over 100 users a day, each of whom interacts with the portal for over six minutes on average. Portals for other complex genetic diseases, such as stroke and heart attack, have been demonstrated using the platform. This open source framework may be useful to the larger community. .

1651W

Association rule learning on electronic health records to identify groups of frequently co-occurring diseases among Geisinger Health System patients. *H.L. Kirchner[1], M. Byrska-Bishop[1], J.B. Leader[1], A. Verma[1], M.D. Ritchie[1,2]. 1) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 2) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.*

   Diseases that a given person develops throughout his/her lifetime may be independent from one another or may be linked by shared genetic or environmental background. We used longitudinal Electronic Health Record (EHR) data spanning nearly 20 years from Geisinger Health System (GHS) to identify clusters of diseases and conditions that co-occur in individuals throughout their lifetime more often than expected by random chance. Specifically, we performed association rule learning using the Apriori algorithm on ICD-9 diagnosis codes from nearly 100,000 Geisinger primary care patients, who participate in the MyCode® Community Health Initiative project, as input. The Apriori algorithm outputs association rules, which provide information about relationships between items (in this case, ICD-9 codes) in large databases. Using this approach, we identified hundreds of co-morbid disease groups, many of them representing known associations, such as an association between type II diabetes mellitus, hyperlipidemia, and hypertension, or an association between hypothyroidism and hypertension, suggesting that association rule learning can be used successfully to identify disease relationships and patterns in EHR data. In the future, we plan to expand this analysis to all 1.5 million GHS patients to get a more comprehensive overview of disease co-occurrence during our patients' lifetime. These relationships can then be used to predict disease likelihoods based solely on a medical history of a patient. The final goal of this analysis will be to identify potential common genetic associations of co-occurring diseases – a phenomenon often referred to as pleiotropy. To accomplish this, we plan to use genome-wide genotyping data, as well as whole exome sequencing data generated as part of the Geisinger-Regeneron DiscovEHR Collaboration. DiscovEHR currently contains genotype and exome sequencing data from more than 50,000 participants of MyCode, with a goal to reach 250,000. We hope that combining frequently occurring patterns within the EHR data with genetic information will facilitate building better predictive models for diseases and subsequently improve preventative care of our patients.

1652T

Training a type 2 diabetes specific functional sequence predictor. *K.M. Lorenz[1,2,3], B.F. Voight[1,2,3]*. 1) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA; 2) Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA; 3) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA.

Type 2 diabetes (T2D) is a complex polygenic disorder characterized by an inability to maintain glucose homeostasis. Genome wide association studies have identified >100 independent loci contributing to the disease, with evidence for hundreds more. However, much of this variation lies in noncoding regions of the genome, making causal variants and sequences difficult to identify and study. We hypothesize that the regulatory features underlying causal variants are disease specific and identifiable from data—the regulatory architecture that influences T2D susceptibility is distinct compared to other diseases. However, models that underlie contemporary functional variant predictors, such as CADD, GWAVA, and FATHMM-MKL, are disease-agnostic to maintain generalizability. To better characterize and identify T2D variants, we have trained a disease specific functional variant predictor using known disease loci and publicly available genomic and epigenomic annotations from the ENCODE and Epigenome Roadmap consortia. Our current model uses a binomial based penalized logistic regression (elastic net) for feature selection. We are better able to classify T2D loci compared to current predictors, with an area under the curve (AUC) of 0.84 compared to their AUCs of 0.76-0.79. Our model retains biologic intuition, as we observe that selected features skew towards T2D-relevant tissues. When we include additional features from T2D-specific tissue types outside of ENCODE or Roadmap, our AUC increases to 0.87. We intend to improve upon this foundation by identifying and including further T2D specific annotations, and will test additional methods of feature selection.

1653F

Multidimensional *in-silico* **analyses to elucidate structure-function relationship of** *SMPD1* **variants causal to Niemann–Pick disease.** *H. El Sokkary[1,2], H. Al-Zahrani[1,2], S. Al-Allasi[2], J. Al-Aama[1,2], R. Elango[1,2], N. Shaik[1,2], B. Banaganapalli[1,2]*. 1) Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, King Abdullaziz University, Jeddah, Saudi Arabia; 2) Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

Niemann-Pick disease (NPD) is an autosomal recessive neurovisceral lipid storage disease caused due to SMPD1 gene mutations. The excessive lipids accumulation in multiple organs in NPD leads to the onset of spectrum of diseases ranging from a neonatal fatal disorder to an adult-onset chronic neurodegeneration. On the basis of signs and symptoms, NPD is divided into A, B, C1 and C2 types. Both type A and B NPD are known to be caused by mutations in SMPD1 gene, which encodes for a lysosomal enzyme called acid sphingomyelinase. The SMPD1 shows a broad mutational range, with the majority being missense mutations. But, the specific relationship between SMPD1 genotype-protein phenotype-clinical phenotype still remains unclear. In this context, we showed the efficacy of multidimensional computational investigations in exploring and developing a better understanding of the structural and functional impacts of SMPD1 mutations on Niemann-Pick disease. Firstly, we discovered 14 (12.61%; against total 111) most deleterious NPD causal mutations of SMPD1 gene by integrating the prediction outcomes of 9 different computational algorithms. In second stage, through 3 dimensional structures analysis we identified that amino acid substitutions induces biochemically severe changes in SMPD1 structure and functions. In the final stage, through molecular drug docking analysis, we recognized that mutant SMPD1 shows high affinity against DESPIPRAMINE drug. Our findings are expected to assist in selecting best mutations for molecular diagnosis out of a crowd of neutral ones and also to design a better competitive inhibitors against SMPD1 mutations. .

1654W

Comparison of HLA allelic imputation programs. *L. Bastarache[1], J.H. Karnes[2,3], C.M. Shaffer[3], S. Gaudieri[4,5], A.M. Glazer[3], H. Steiner[2], J.D. Mosley[3], S. Mallal[5], J.C. Denny[1], E.J. Phillips[3,5], D.M. Roden[1,3,6].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, Tucson, AZ; 3) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 4) School of Anatomy, Physiology and Human Biology, University of Western Australia, Nedlands, Western Australia, Australia; 5) Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN; 6) Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN.

   Multiple programs are now available to impute human leukocyte antigen (HLA) alleles from SNP-level data. Such imputation programs are attractive due to the importance of HLA alleles in human disease and pharmacogenomics, the widespread availability of genome-wide association study (GWAS) data, and the expertise required for HLA allele sequencing. However, a comprehensive comparison and evaluation of these programs is not available. We compared HLA imputation results of HIBAG, SNP2HLA, and HLA*IMP:02 to sequenced HLA alleles in 3,265 samples from BioVU, a de-identified electronic health record database coupled to a DNA biorepository. We performed four-digit HLA sequencing for *HLA-A, -B, -C, -DRB1, -DPB1,* and *-DQB1* using long-read 454 FLX sequencing. All samples were genotyped using the Illumina HumanExome BeadChip platform and a GWAS platform (either Illumina HumanOmni1-QUAD or HumanOmni5-QUAD). For each program and locus, we calculated number of imputed alleles, concordance with sequencing results, and call rate. Call rates and concordance rates were compared by platform, frequency of allele, and race/ethnicity. Overall concordance rates were similar between programs in European Americans (EA) with SNP2HLA being the highest (0.975 [SNP2HLA]; 0.939 [HLA*IMP:02]; 0.976 [HIBAG]). SNP2HLA also predicted the largest number of alleles overall (n=210) with the highest call rate (1.00). Concordance rates were lower overall for African Americans (AAs) (0.919 [SNP2HLA]; 0.619 [HLA*IMP:02]; 0.929 [HIBAG]). These observations were consistent when accuracy was compared across HLA loci, with the highest concordance being 0.988 for SNP2HLA and HIBAG in *HLA-DQB1* in EAs and the lowest concordance being 0.414 for HLA*IMP:02 in *HLA-DRB1* in AAs. All imputation programs performed similarly for low frequency HLA alleles. Higher concordance rates were observed when HLA alleles were imputed from GWAS platforms versus the Exome Chip. Although concordance was similar for all imputation platforms, SNP2HLA had the highest overall concordance and call rates and provided a significant advantage in the number of alleles imputed. All programs demonstrated better concordance in EAs versus AAs and performed well for low frequency alleles. Our results suggest that high genomic coverage is preferred as input for HLA allelic imputation. These findings provide guidance on the best use of HLA imputation methods and elucidate their limitations.

1655T

PyDAIR: A tool for precise determination of the diversity of immunoglobulin sequences. *X. Fu[1], J.Q Sun[2], J. Geng[1], S. Asakawa[3].* 1) State Key Laboratory of Biotherapy, Sichuan University, Chengdu, Sichuan, China; 2) Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo; 3) Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo.

   The immune system's antibody repertoire is highly plastic and can be directed to create antibodies with broad diversity and high selectivity. Large-scale repertoire analysis of antibody can provide powerful results including insight into a better understanding or a snapshot of the adaptive immune repertoire, and an improved understanding of the way by which the immune system eliminates unwanted infections. The enormous antibody diversity is generated by somatic recombination of a large number of immunoglobulin (Ig) gene segments, which involves a random recombination of tandemly arranged variable (V), diversity (D) (only in Ig heavy (H) chain) and joining (J) germline gene segments. The vast majority of the diversity in the naïve antibody repertoire is within the H chain complementarity-determining region 3 (CDR-H3), which encompasses the VDJ recombination junctions. Advances in sequencing technologies (e.g., next-generation sequencing (NGS)) enable the current data sets of the immunological repertoire. Information gained from NGS of Ig genes (Ig-seq) thus requires efficient informatics analysis tools to assist in categorizing and transforming these data into valuable biological insight. To this end, we develop a Python package, named PyDAIR (Python library for Diversity Analysis of Immune Repertoire), to study the antibody diversity based on NGS data. In detail, this package contains three main steps: (i) Ig sequences are aligned to all V and all J segments (species-specific) to determine the optimal alignment of V and J using BLAST; (ii) The CDR-H3 region is determined based on the two conserved motifs in V and J genes; (iii) Strategies used to define the un-aligned region for full D segment determination. We then apply PyDAIR to characterize the antibody repertoire of torafugu (*Takifugu rubripes*), species that possess an recognizable adaptive immune system and one of the smallest genomes (~400 Mb) among vertebrates, makes it an ideal model system for comparative adaptive immune system studies. As a result, PyDAIR efficiently defined torafugu-specific V, D, J, and CDR-H3 sequences. The analysis also highlighted the strong VDJ usage bias as well as the diverse size of the available CDR-H3 repertoire. Analysis of the published sequences data set of human is underway. The PyDAIR package will be a valuable tool for further in depth analysis of the antibody repertoire.

1656F

HLAminer: Prediction of HLA types from NGS data. *R.L. Warren, S. Yeo, J. Chu, R. Holt, I. Birol.* BC Cancer Agency, Genome Sciences Centre, Vancouver, BC, Canada.

The Human Leukocyte Antigen (HLA) is a gene locus that encodes cell-surface receptors that present epitopes to T cells, and is the foundation of immune system regulation in humans. The HLA complex enables the immune system to distinguish between a person's normal healthy cells, and cells that are infected, mutated or are derived from some other individual. Due to the necessity of recognizing diverse molecular signatures, there is extreme diversity in the genes that encode HLA proteins, and determining the specific gene variants of an individual is of considerable importance in clinical medicine and in research. For instance, knowing HLA genes and matching donor and recipients with the same HLA is key to successful organ grafts. Traditionally, HLA typing was done by PCR amplification, but often offers low allele-level resolution. The drastic decrease in sequencing costs we witnessed over the past decade is facilitating the uptake of DNA sequencing technologies in the clinic, but the bioinformatics analysis of whole genome and transcriptome shotgun data still represents a significant barrier to entry and is out of reach to many researchers and clinicians. As more shotgun datasets become available, targeted interrogation of HLA is value-added since those datasets can be mined retrospectively to predict types. We present HLAminer, the first reported and highly cited HLA prediction software for next-generation sequence data. The flexible pipeline enables HLA predictions from direct read alignments to a set of known, references and, for best results, uses targeted de novo assembly of sequence reads. The software was used to robustly predict the HLA genes of the colorectal cancer patient cohort and in over 7,000 TCGA RNA-seq samples (of which 515 from six tumor sites were published) [1-3]. We present changes to the code since publication, including the utilization of BioBloom Tools [4], an efficient sequence read partitioning software. Since January 2016, the HLAminer software had over 1,500 downloads world-wide. 1. Warren RL, et al. 2012. Derivation of HLA types from shotgun sequence datasets. Genome Med. 4:95. 2. http://www.bcgsc.ca/platform/bioinfo/software/hlaminer 3. Brown SD, et al. 2014. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 24:743-50. 4. Chu, et al. 2014. BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters. Bioinformatics. 30:3402-4.

1657W

Rare hematopoietic clone detection using error-corrected sequencing. *A.L. Young[1,2], G.A. Challen[3], B.M. Birmann[4], T.E. Druley[1,2].* 1) Department of Pediatrics, Washington University School of Medicine, Saint Louis, MO; 2) Center for Genome Sciences and Systems Biology, Washington University School of Medicine, Saint Louis, MO; 3) Department of Internal Medicine, Division of Oncology, Washington University School of Medicine, Saint Louis, MO; 4) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Acute myeloid leukemia (AML) is the most common hematopoietic malignancy, diagnosed in approximately 20,000 individuals each year in the United States. While the goal of treatment is to induce a stable and durable remission, the five year survival rate is only 25%. Accurately assessing response to treatment is essential for therapeutic selection and long term survival. Currently, the gold standard for assessing remission status is multiparameter flow cytometry (MPFC) targeting leukemia-specific cell surface markers, which is prognostic to the detection limit of 1:10,000 cells. However, only half of AML cases have a suitable immunophenotype for MPFC. Conversely, virtually every AML is marked by leukemia-specific somatic mutations, presenting an opportunity for the sequencing-based detection of persistent leukemia cells following treatment. However, previous approaches have been limited by the error rate of the sequencing technology (approximately 1%) and only interrogated specific exons of recurrently mutated genes. To circumvent these limitations, we combined targeted capture of AML-associated genes with error-corrected sequencing to create a broad platform for residual disease quantification with a limit of detection similar to MPFC, but applicable to virtually all cases of AML. Here, we present the novel wet-lab techniques and bioinformatics workflow necessary to identify rare, leukemia-associated mutations. Briefly, we use unique molecular identifiers (UMIs) to tag individual molecules captured from genomic DNA using a 54-gene panel. Sequencing errors are corrected by collapsing reads with the same UMI. Further position-specific error modelling allows us to mitigate the effect of artifacts introduced by DNA degradation (e.g. cytosine deamination to uracil and guanine oxidation to 8-oxoguanine). As a proof of principle we used these techniques to identify rare hematopoietic subclones in the blood of healthy individuals. Variants identified by this technique were validated by droplet digital PCR—an orthogonal non-sequencing-based technology. Ongoing work with the Children's Oncology Group is focused on implementing these methods for residual disease assessment in children with de novo AML.

1658T

Meta-analysis of biopsy expression profiles of kidney transplant patients identified key drivers for acute rejection targeted by minocycline. *W. Zhang[1], Z. Yi[1], K. Keung[1,2], B. Zhang[1], Z. Li[1], C. Wei[1], L. Li[1], M. Menon[1], C. Xi[1], P. O'Connell[2], B. Murphy[1].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Medicine, Westmead Clinical School, The University of Sydney, Sydney, NSW, Australia.

Acute rejection (AR) is a major cause of organ transplant failure and has a variable negative impact on allograft survival even among patients who recover after immunotherapy. Molecular signatures in kidney allografts with AR have been identified by several independent studies, but results have varied. A meta-analysis of gene expression data from kidney transplant biopsies in the GoCAR (Genomics of Chronic Allograft Rejection) study along with five publicly available datasets was performed to identify a meta- signature associated with AR. From 735 biopsy sampleswe identified 982 (565 up and 417 down in AR vs non-AR patients) meta genes at FDR<0.05. Genes involved in immune response, T/B cell activation and proliferation, antigen processing and presentation, protein kinase cascade and NFKB signaling pathway were upregulated, whilst genes involved in metabolism were downregulated. 14 key driver genes were further derived from the differential submodules of meta- coexpression network of metagenes in AR patients compared to non-AR patients, withCASP1 at the center of the network. The key drivers were validated in independent expression datasets of renal transplants with AR as well as native kidney diseasesincluding IgA nephropathy, diabetic nephropathy and Systemic Lupus Erythematosis (SLE) suggesting the pivotal role of these key genes in driving the inflammatory response in acute allograft rejection and immune-driven native kidney diseases. Finally, application of a drug repurposing strategy identified minocycline as a potential therapeutic agent targeting these key drivers. The administration of minocycline in a murine cardiac allograft model of AR demonstrated a significant reduction in transcript expression of a number of these key genes, as well as multiple proinflammatory cytokines and chemokines. Our study identifies potential new targets for therapeutic intervention, and suggests minocycline as potential novel candidate for the prevention of AR. .

1659F

Computational estimation of leukocyte distribution in cord blood methylation arrays using a transfer learning framework. *L. Dinh[1], M. Jones[2], O.M. de Goede[2], H.R. Razzaghian[2], J.L. MacIsaac[2], W.P. Robinson[2], P. Lavoie[2], M. Kobor[2], R. Ng*[1], S. Mostafavi*[3].* 1) Computer Science, University Of British Columbia, Vancouver, BC, Canada; 2) Child and Family Research Institute, University of British Columbia; 3) Statistics, University Of British Columbia, Vancouver, BC, Canada.

Background: Computationally quantifying cell type composition in blood, a process called cell type decomposition, is useful in disentangling the effect of cell type composition variation from differential methylation or expression in disease studies. For example, diseases like cancer or SLE leave a characteristic imprint on the leukocyte population that can be detected by quantifying cell type composition. In differential expression or methylation studies, cell type heterogeneity is a major confounding factor, resulting in false positive ('spurious') associations, if not quantified and accounted for. Because the majority of large-scale human studies use whole tissue data, developing computational algorithms that can predict and adjust for cell type composition is a key step in reducing false positive rates. **Results:** Accurate computational algorithms for quantifying cell type proportions in adult whole blood exist, however, these algorithms do not generalize well to the cord blood setting. In this study, we first comprehensively assess the accuracy of existing adult deconvolution algorithms when applied to cord blood. To understand the impacted accuracy, we examined potential culprits like increased variability of methylation within cell types, the unaccounted appearance of nucleated red blood cells, and the shift in cell type profiles attributed to adult leukocyte lineages. To resolve these difficulties, we introduce an approach, based on joint analysis of methylation profiles in purified leukocytes in adult and cord blood, to select cell type signatures that more accurately estimate cell type proportions in cord blood samples. **Summary:** In conclusion, we show that current whole-blood-based algorithms for computing cell type proportions have low accuracy when applied to cord blood samples, and that our new approach enables us to improve the accuracy of estimated cell type proportions in this setting.

1660W

Whole genome sequencing study on bronchodilator drug response in ethnically diverse children with asthma. *A.C.Y. Mak[1], M.J. White[1], Z. Szpiech[1], M. Pino-Yanes[2], S. Oh[1], D. Hu[1], S. Huntsman[1], C. Eng[1], P.Y. Kowk[1], D. Torgerson[1], J. Galanter[1], E. Ziv[1], M.A. Seibold[3], N. Zaitlen[1], R. Hernandez[1], E.G. Burchard[1], New York Genome Center.* 1) University of California, San Francisco, San Francisco, California, United States; 2) Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain; 3) National Jewish Health, Denver, Colorado, United States.

   Introduction Albuterol, a bronchodilator medication, has been the standard of care for asthma treatment worldwide for the past three decades. Despite significant variation in bronchodilator response (BDR), albuterol is often the only medication prescribed for asthma, regardless of disease severity and the striking racial/ethnic disparity in BDR. Specifically, Puerto Rican and African American children with asthma have significantly lower BDR compared to children of European origin. Previous studies have identified a genetic contribution to variation in BDR, yet much of the heritability of BDR remains to be explained. We reason that rare, population-specific variants contribute to BDR in minority children with asthma. **Methods** We performed whole genome sequencing on 1,500 minority children with asthma from the tails of the BDR distribution. Our study included 250 high and 250 low drug responders from three racial/ethnic groups, African Americans (n=500), Puerto Ricans (n=500), and Mexican Americans (n=500). We performed association tests on pooled rare variants in 5 kb sliding windows using SKAT-O. Principal component analysis was performed using EIGENSTRAT to identify structure in our study population. All association models were adjusted for age, sex, BMI, and the top ten principal components. **Results and Conclusions** Our results identified ethnic-specific rare variants associated with BDR. This includes loci near genes previously associated with corticosteroid response (*FOXK1*), T-cell stimulation (*CD5*) and airway hyper-responsiveness (*FHL2*). Our study furthers our understanding of genetic analysis in admixed populations and may play an important role in advancing the foundation for precision medicine in "at risk" minority populations. *Authors contributed equally to this work.

1661T

The Longitudinal Pediatric Data Resource: Accelerating new knowledge discovery and translation of findings in genetic disorders identified through newborn screening. *A.M. Brower[1], M. Kuhlmann[2], J.G. Loutrel[1], G. Labilloy[3], N. Felicelli[3], B. Bowdish[1], C. Day[1], A. Rupert[3], M.S. Watson[1], M. Wagner[3], P.S. White[3].* 1) Newborn Screening Translational Research Network, American College of Medical Genetics and Genomics, Bethesda, MD 20814; 2) Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH 45229; 3) Division of Biomedical Informatics, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH 45229.

   Since its inception over half a century ago, state-based newborn screening (NBS) programs annually screen 4 million newborns for genetic conditions that require immediate treatment to prevent morbidity and mortality. Research that capitalizes on the NBS system offers an opportunity to improve understanding of genetic disorders in newborns, develop novel technologies, and study the effectiveness of new treatments. Researchers have begun to explore the use of genomics in NBS, and state-based NBS programs have begun screening for disorders that have a large amount of genetic heterogeneity, including onset of symptoms in childhood or beyond. Natural history studies that capture longitudinal information on cases resulting from this expanded NBS will establish the clinical validity and clinical utility of identified disorders, and pilots of novel NBS technologies will provide analytical validation. To support these efforts, the NICHD created the Newborn Screening Translational Research Network (NBSTRN). The NBSTRN is developing an infrastructure for researchers to access to robust resources and focused expertise in NBS. As part of this effort, the NBSTRN and Cincinnati Children's are developing infrastructure and software tools to securely house long-term follow-up data on children diagnosed with NBS conditions. This tool, the Longitudinal Pediatric Data Resource (LPDR), offers a common data model and a secure environment for researchers to aggregate data on these rare conditions and incorporates a set of common data elements (CDEs) established by a national panel of clinical experts. Elements are developed and managed through a custom Data Almanac application that provides opportunity for virtual co-authoring of element definitions and values. A REDCap™ based system that includes over 60 conditions and over 10,000 CDEs is available to NBS researchers. A data discovery and query application, NBSMart, enables interaction with long-term outcomes as well as genomic information across different studies. Over 100 researchers across 8 studies are utilizing the LPDR. To date, data from 1900 participants with over 1 million data points have been collected. The LPDR will also house clinical and whole-exome and genome sequencing data from the Newborn Sequencing in Genomic Medicine and Public Health (NSIGHT) project. The use of the LPDR by researchers will create unique cohorts of individuals with genetic disease and accelerate understanding of these conditions.

1662F

iReceptor: Bioinformatic platform for storing and sharing next generation sequencing (NGS) data from immune repertoires. *F. Breden[1], N. Marthandan[1,2], B. Zimonja[1], J. Jagele[1], J. Scott[1,2,3], B. Corrie[4].* 1) The IRMACS Centre, Simon Fraser University, Burnaby, British Columbia, Canada; 2) Department of Molecular Biology, Simon Fraser University, Burnaby, British Columbia, Canada; 3) Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada; 4) Centre for eResearch, Auckland, New Zealand.

   The diversity of the human Adaptive Immune Receptor Repertoire (**AIRR**) (antibody/B-cells or T-cells) is huge; there are as many as 10^{13} possible B-cell receptor sequences, and the human body can contain up to 10^{11} B- or T-cell receptor producing cells. AIRR data are critical for vaccine and therapeutic antibody development and anti-cancer immunotherapy. Only by analyzing data across studies, with associated data on patient demography, treatment and clinical outcome, will the true potential of this NGS revolution in immunotherapy be realized. The iReceptor Gateway will link AIRR databases, thus allowing the query of these databases for sequences satisfying specific criteria, between studies and institutions. iReceptor implements a distributed data set model, such that researchers can store their data under their own IRB constraints, and release these data to other nodes in the iReceptor environment based on shared consent, reciprocal data transfer and IP agreements. In order to address all of the challenges involved in storing and sharing these data, we have formed a consortium to develop protocols and best practices for producing, reporting and analyzing AIRR data. The most recent Adaptive Immune Receptor Repertoire Community Meeting was held June 28-30 2016 at NIH Fishers Lane Facility in Rockville MD (contact breden@sfu.ca).

1663W

Novel, allele sequence signature approach enables HLA-typing for biomarker identification using gene expression data. *M.L. Buchkovich, C.C Brown, K. Robasky.* Bioinformatics and Clinical Systems, Q2 Solutions - EA Genomics, Morrisville, NC.

   The human leukocyte antigen (HLA) system encodes proteins that identify self from non-self cells, and play critical roles in immunotherapy, regenerative medicine, patient safety monitoring and autoimmunity. Studies, already generating data to quantify gene expression, can greatly benefit from HLA-typing in RNA-seq data. Genotyping efforts are limited by significant sequence similarity coupled with high polymorphic rates in the HLA region, and specialized assays are not designed to handle analysis of FFPE-preserved tumor samples. To help overcome these limitations, we developed a novel sequence signature-based approach for identifying HLA alleles in RNA-seq data. Our approach first determines the HLA gene for each sequencing read. Once all reads have been assigned, the sequence signature for each gene is used to determine the HLA type. The sequence signature reference set, generated using the full IPD-IMGT/HLA Database reference, allows for the identification of uncommon HLA alleles. We characterized the performance of this approach using two RNA-seq datasets, 1) publicly available 75-bp PE data from 342 samples with HLA types available for five HLA genes (A, B, C, DRB1 and DQB1), and 2) 50-bp PE data generated in-house for 15 FFPE hepatic tumor samples sequenced in replicate. We assigned 4-digit precision HLA types to the publicly available samples with >90% accuracy. In particular, we achieved a 29.5% gain in accuracy for 4-digit precision of DQB1 over the accuracy obtained with another leading methodology. Additionally, we sequenced 15 FFPE hepatic tumor samples in replicate and obtained 91% overall replicate concordance (93% Class I and 87% Class II) with 2-digit precision, compared to the 88% we previously presented. Reproducible and accurate HLA typing, especially in FFPE-preserved oncology samples, enables biomarker identification using HLA types in the wealth of whole transcriptome data that has already been and is being generated. Overall, our approach enables quick, accurate and reproducible HLA typing in support of those large pre-clinical research programs that are already exploring these data for differential expression and antigen presentation, with no additional lab work and with minimal additional computational resource requirements.

1664T

A clinical catalogue of phenotypes associated with variation in the MHC locus. *J.H. Karnes[1,2], L. Bastarache[3], C.M. Shaffer[2], S. Gaudieri[4,5,6], Y. Xu[7], A. Glazer[2], J.D. Mosley[2], S. Raychaudhuri[8,9,10,11,12,13], S. Mallal[5,14], Z. Ye[15], J.G. Mayer[15], M.H. Brilliant[16], S.J. Hebbring[16], D.M. Roden[2,3,17], E.J. Phillips[2,4,6,14], J.C. Denny[3].* 1) Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, Tucson, AZ; 2) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) School of Anatomy, Physiology and Human Biology, University of Western Australia, Nedlands, Western Australia, AustraliaSchool, Boston, USA; 6) Institute of Immunology & Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia; 7) Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN; 8) Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA; 9) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA; 10) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, USA; 11) Partners Center for Personalized Genetic Medicine, Boston, USA; 12) Institute of Inflammation and Repair, University of Manchester, Manchester, UK; 13) Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 14) Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN; 15) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA; 16) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA; 17) Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN.

   Human leukocyte antigen (*HLA*) genes have more phenotypic associations than any other locus. However, a systematic approach has not yet been undertaken to investigate the influence of *HLA* variation across the phenome. We set out to replicate known associations, determine novel associations, and provide a freely available reference of *HLA*-phenotype associations. We performed phenome-wide association studies (PheWAS) for *HLA* genomic variation using de-identified electronic health records (EHR) databases coupled to DNA biorepositories. Using SNP2HLA, we imputed *HLA* variation, including two and four digit *HLA* alleles, *HLA* allele haplotypes, and non-synonymous (NS) *HLA* variants, from two populations of 29,712 and 11,368 European ancestry individuals genotyped on the Illumina® HumanExome BeadChip. We tested association of *HLA* variation with 1,545 phenotypes derived from EHRs and conducted a Trait-based Association Test to determine cross-phenotype associations driven by pleiotropic effects. Significance was considered at FDR q-value<0.10 in the discovery cohort as well as p-value<0.05 in the replication cohort. A total of 115 four-digit and 96 two-digit *HLA* allele-phenotype associations were significant in both discovery and replication cohorts. Most associations were seen with Class II *HLA* loci, with the strongest being *HLA-DQB1*03:02* and type 1 diabetes (odds ratio 4.31 [3.57-5.19], meta-analysis p=1.40x10[-53]). Previously known associations were strongly represented in these results, particularly for autoimmune diseases. These included rheumatoid arthritis, ankylosing spondylitis, multiple sclerosis, psoriasis and psoriasis vulgaris, celiac disease, Sicca syndrome, and primary biliary cirrhosis, with some having strong cross-phenotype associations potentially driven by pleiotropic effects. We observed several novel associations among the phenotypes analyzed. A total of 2,044 significant associations were observed with NS variants. Novel associations were observed and replicated between NS variants and nodular lymphoma, partial epilepsy, mild cognitive impairment, and atherosclerosis of the extremities. We provide strong evidence of the validity of the PheWAS approach and further emphasize the importance of the major histocompatibility complex (MHC) region in human health and disease. We will provide a comprehensive, publicly available database of *HLA* loci to allow for in-depth evaluations of underlying genetic architectures for a range of diseases.

1665F

Large-scale gene regulatory network inference to identify novel regulators of innate inflammatory response. *T. Yang, R. Xue, M. Beste, B. Hayete, I. Khalil.* GNS Healthcare, Cambridge, MA.

   Systematic mapping of expression Quantitative Trait Loci (eQTL) has increasingly enabled the discovery of tissue- and context-specific cis-regulatory domains governing the intermediate transcriptional effects of genetic variation. Subsequent inference of gene regulatory networks provides a complementary framework for evaluating and validating causal relationships captured during initial eQTL mapping. Here we leverage context-specific eQTL maps to infer genome-scale Bayesian networks for the human monocytic response to LPS and IFNg using Reverse Engineering and Forward Simulation (REFS), a distributed high-performance computing engine for network learning and predictive simulation. REFS efficiently implements a parallelized score-based search to sample statistical ensembles of network structures from the Bayesian posterior of model structures given the data. To improve the accuracy of the learned networks, we encode a combination of structural priors encompassing (i) the adjacency matrix of the eQTL map, automating the identification of causal, reactive, and independence relations among cis- and trans- gene pairs and (ii) transcriptional regulatory annotations from high-resolution FANTOM5 profiles. The resulting networks for >30 thousand gene isoforms recapitulate a range of canonical transcriptional mechanisms, including distinct MYD88/IRAK, p38/JNK, and JAK1/STAT1 sub-networks. Prioritization of inferred regulatory modules identified 28 candidate regulators of acute transcriptional response for further validation. Finally, we demonstrate an application of REFS forward simulations to identify an optimal set of targeting perturbations to maximally rescue a quiescent phenotype from the activated LPS transcriptional response.

1666W

Imputation of gene expression implicates expected and unexpected genes associated to Crohn's disease. *K. Gettler[1], L. Chuang[2], N. Hsu[2], G. Hoffman[2], S. Kugathasan[3], J. Cho[2].* 1) Yale University, New Haven, CT; 2) Mount Sinai, New York City, NY; 3) Emory University, Atlanta, GA.

   A major limitation of GWAS has been the imprecise definition of pathogenic genes within loci. The vast majority of loci are driven by non-coding variation. Mapping specific alleles most implicated in trait association to altered gene expression has been challenging as regulation of gene expression is likely highly context-dependent. Promising approaches have been recently developed, where reference eQTL datasets are used to impute gene expression of GWAS cohorts to perform transcriptome-wide association studies (TWAS). We hypothesized that imputation of gene expression in the most relevant tissue type, here the terminal ileum in newly-diagnosed, untreated Crohn's disease (CD), will define CD-associated genes within GWAS loci. RNA-seq data were collected from intestinal biopsies taken from 302 untreated CD patients genotyped with the Illumina Omni 2.5 chip. We imputed expression of 1280 candidate genes from CD loci in 5,956 CD cases and 14,927 controls. TWAS identified 39 genes from 25 loci demonstrating Bonferroni-corrected significant association of altered gene expression to CD (P = 4 x 10-5). We confirmed the well-established role of decreased expression/function of NOD2 (chr16, 50 Mb, TWAS P = 8.2 x 10-12) and ATG16L1 (chr2, 234 Mb, TWAS P = 3.2 x 10-9) to CD. NOD2 and ATG16L1 TWAS P-values were 3- and 5- orders of magnitude more significant than the next most significant genes in their loci. Conversely, increased PTGER4 (prostaglandin receptor 4) gene expression was uniquely associated with CD (P = 6.0 x 10-8). For the loci at chr5 130Mb and chr3 at 49 Mb, TWAS patterns equally implicate multiple genes. However, for many loci, TWAS uniquely implicates one gene over others, in many cases, highlighting unexpected genes. For example, at chr6 at 167Mb, we observe much more significant association to the ribonuclease RNASET2 (TWAS P = 2.6 x 10-12) compared to other candidates (e.g. CCR6, TWAS P = 1.6 x 10-3). Similarly, TWAS implicates SEC16A (P = 1.8 x 10-10) over CARD9 (P = 2.2 x 10-5), CUL2 (P = 2.0 x 10-7) over CREM (P = 0.01), PDGFB (P = 3.7 x 10-7) over SMCR7L (P = 3.2 x 10-5) and GSDMB (P = 5.6 x 10-7) over ORMDL3 (P = 0.03). We will present the validation of the TWAS gene and gene expression predictions by testing in an exome sequencing cohort whether an excess of predicted rare deleterious alleles are observed in CD cases or controls in genes where risk alleles are predicted to decrease and increase gene expression, respectively.

1667T

Multidimensional computational structure-function relationship analysis of *FRGR3* **genetic mutations causal to achondroplasia.** *R.H. Baaqeel[1,2], R.S AlHarbi[1,2], J.Y Al-aama[1,2], R. Elango[1,2], N.A Shaik[1,2], B. Banaganapalli[1,2].* 1) Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, King Abdullaziz University, Jeddah, Saudi Arabia; 2) Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

   Achondroplasia (ACH), a common form of short-limbed dwarfism, occurs due to FGFR3 gene mutations in 80% of cases, either sporadically or by defective allele inheritance in an autosomal dominant pattern. The two most common amino acid substitution mutations of FGFR3 gene known to cause the development of ACH are G369C and G380R. However, the specific impact of these amino acid substitutions on protein-structure functional relationship is not clear. The molecular validation of these mutations through experimental evidences is difficult owing to the requirement of time, technical expertise and huge costs. Hence, our objective was to examine the efficacy of advanced multidimensional computational strategies in understanding the structural plasticity of FRGR3 pathogenic mutations and its potential implications for protein-drug interactions. By exhaustive literature survey and database base searching, we identified 17 lethal variants of FGFR3 gene causal to ACH development. By integrating the predictions of nine different computational algorithms with different operating principles, we discovered that Y278C, S279C, L379R,N542K, T548K, K652E and T653P mutations are lethal to structure and function of FRGR3 gene. Moreover, to our surprise, we find that the deleterious impact of G369Cand G380R mutations (most common to ACH) is 2 fold less than the above mentioned 7 mutations. We identified that Y278C, S279C, L379R,N542K, T548K, K652E and T653P mutations induces structural deformities on FGFR3 protein and can also disturb its affinity towards common ACH drugs like dovitinib dilactic acid and pazopanib. Our computational findings are expected to help in recognizing lethal FGFR3 mutations causal to ACH and also to select better competitive inhibitors for mutated forms FGFR3 protein.

1668F

Skeletal phenomics in zebrafish via microCT-based barcoding. *R.Y. Kwon[1], M. Hur[1], P. Huber[1], J. Lee[1], M.H. Thompson[1], C.J. Watson[1], S.K. Mc-Menamin[3], D. Parichy[2].* 1) Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA; 2) Department of Biology, University of Washington, Seattle, WA; 3) Department of Biological Sciences, University of Massachusetts, Lowell, MA.

Bones comprise different morphologies, compositions, and modes of developmental origin. Skeletal phenotyping has traditionally focused on intensive analysis within a single skeletal compartment, yet single genetic variants are often associated with changes in multiple traits across multiple bones. These coordinated changes may reveal gene pleiotropy, expose shared genetic mechanisms, and provide heightened sensitivity in detecting subtle changes in bone mass that would not be perceptible by studying a single bone in isolation. Motivated by the emergence of zebrafish as a rapid testbed for human gene analysis, as well as the potential to merge large-scale gene testing with phenomic discovery, we developed FishCuT, a novel algorithmic approach for microCT analysis of the adult zebrafish skeleton. We coupled supervised region growing with connectivity analysis to enable rapid (<5min per fish) computation of 288 different measures (24 vertebral bodies x 3 vertebral elements x 4 quantities per fish) of axial bone morphology and mineralization in a process we call "skeletal barcoding". As a proof of principle, we barcoded zebrafish mutants previously isolated from genetic screens, and identified two mutants with abnormal skeletal phenotypes: *puma* and *opallus*. *opallus* (which harbors a mutation in thyroid stimulating hormone receptor (*tshr*) identical to a human mutation causing constitutive TSHR activity and hyperthyroidism [1]) exhibited significantly elevated TMD across all centra. *puma* (which harbors a mutation in the alpha tubulin gene *tuba8l3a* [2]) exhibited increased volume in anterior haemal arch/rib elements only, suggesting the potential to detect both global and regional shifts in mass/mineral accrual. In humans, hyperthyroidism is associated with low BMD [3], whereas TSHR gain-of-function has been associated with high BMD [4]. Thus, this study suggests the potential for the zebrafish skeleton to predict a complex phenotype integrating competing influences at both systemic and local levels. Together, this study advances new approaches to skeletal phenomic analysis, and supports the translational value of zebrafish as a model of human genetic bone disorders.   [1] Mc-Menamin SK, et al., Parichy DM. Science, 2010. [2] Larson TA, et al., Parichy DM. Zebrafish, 2010. [3] Greenspan SL, Greenspan FS. Ann Int Med, 1999 [4] de Lloyd A, et al., Ludgate M. J Endocrinol, 2010.

1669W

A novel method for gene set enrichment analysis of ExomeChip data, EC-DEPICT, identifies relevant biological pathways. *R.S. Fine[1,2,3], T.H. Pers[4,5], J.N. Hirschhorn[1,2,3], the deCODE Genetics, BBMRINL, GOT2D, CHARGE, and GIANT Consortia.* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 5) Department of Epidemiology Research, Statens Serum Institut,Copenhagen, Denmark.

Pathway analysis is a powerful tool for biological interpretation of genetic association data. We developed EC-DEPICT for pathway analysis of ExomeChip (EC) data by adapting DEPICT, our genome-wide association study (GWAS) pathway analysis method. DEPICT extends and reconstitutes existing gene sets by using coexpression to assign gene set membership probabilities to every gene in the genome. These quantitative annotations are less biased and leverage signal from uncharacterized genes. In EC-DEPICT, we use these membership probabilities to generate test statistics that assess, for each gene set, the enrichment of genes with significantly associated coding variants. To control for inherent biases in the data, we convert these test statistics to empirical P values and false discovery rates (FDR) by generating test statistics from 2000 sets of null ExomeChip genotype data. We applied EC-DEPICT to EC data for height and BMI in up to ~700,000 individuals, using conditional analysis to select associations independent of known GWAS signals. For height, 496 gene sets were significantly enriched (FDR<0.01) in genes with associated coding variants. Many gene sets identified from EC data overlapped with those independently identified from noncoding GWAS variants, but EC-derived gene sets more strongly and specifically implicated growth plate-related biology and uniquely identified proteoglycan binding. Unsupervised clustering and visualization of implicated genes and gene sets revealed a cluster of genes that are known to underlie monogenic disorders of skeletal growth, contain height-associated coding variants, and have high membership probabilities for a shared group of EC-enriched gene sets. Several other genes with height-associated variants but no prior connections to growth are interspersed in the same cluster and are new, promising candidates for monogenic growth disorders. For BMI, EC-DEPICT identified neurotransmission- and synapse-related gene sets, strongly and independently replicating recent pathway analysis from BMI GWAS results. EC-DEPICT can be used for analysis of EC data for any phenotype, and can be adapted for exome sequencing. Our results indicate that pathways implicated by coding and noncoding variation often overlap but, at least for height, can also implicate specific aspects of biology. EC-DEPICT provides a useful framework for biological interpretation of exome array results and points to specific genes for follow-up studies.

1670T

Birth prevalence estimation of severe Hypophosphatasia in European populations using large-scale protein-coding genetic variation databases. *G. del Angel[1], E. Mornet[2], J. Reynders[1].* 1) Genomics and Data science, Alexion Pharmaceuticals Inc., Lexington,MA, MA; 2) Laboratoire de génétique constitutionnelle prénatale et postnatale. Centre Hospitalier de Versailles. Le Chesnay, France.

Hypophosphatasia (HPP) is a rare, inherited metabolic disorder caused by mutations in the alkaline phosphatase gene (ALPL, OMIM:171760). HPP is characterized by defective bone mineralization and reduced serum ALP activity. HPP is a clinically heterogeneous disease and is classified according to severity and age of onset: perinatal, infantile, childhood, adult and odonto HPP. We propose here a novel method to estimate severe HPP (i.e., perinatal and infantile) birth prevalence based on leveraging large-scale population genomic databases to obtain individual allele frequencies for disease mutations. This approach presents an unbiased methodology that avoids data sparsity and case ascertainment problems in previous incidence computation approaches based on case reports. We analyzed cohort and mutation data from 345 patients, out of which 182 presented perinatal or infantile forms. We classified ALPL mutations onto three tiers: Tier 1: Severe mutations: Homozygous mutations in severe HPP cases; mutations known to have dominant negative effect or low residual ALP enzymatic activity (<10% as discussed in Zurutuza (1999); loss of function mutations, and severe mutations in compound heterozygote form with high residual activity. Tier 2: Mutations with high residual activity. Tier 3: Missense mutations in public population databases not seen yet in patients with HPP but with strong computational evidence for deleteriousness. Allele frequency estimates for each mutation above were obtained using the Exome Aggregation Consortium (Lek 2016), which aggregates protein-coding mutations for over 30,000 samples from individuals with European ancestry. HPP birth prevalence was estimated using Hardy-Weinberg equilibrium. Based on this tier classification, upper and lower limits on severe HPP birth prevalence were determined. The lower bound considers Tier 1 and Tier 2 mutations, yielding an estimate of 2.6 per million births (95% CI: [1.7, 3.7]). An upper bound was obtained considering all 3 Tiers, yielding an estimate of 3.8 per million births (95% CI: [2.7, 5.1]). These bounds fall close to previously published estimates, and provide new methods to estimate birth prevalence of other rare genetic diseases. Future work will be aimed at better understanding of ALPL genotype/HPP phenotype relationships and mutation severity.

1671F

VICTOR: A pipeline for Variant Interpretation in Clinical Testing Or Research. *B.-J. Feng[1,2], K. Callis Duffin[1], G.G. Krueger[1], W. Kohlmann[2], J.D. Schiffman[2], M. Schmidt[3], A. Meindl[4], R. Berruti[5], R. Schmutzler[6], E. Hahnen[6], M. Vallée[7], A. Droit[7], D.F. Easton[8], S.V. Tavtigian[2], J. Simard[7], D.E. Goldgar[1,2].* 1) Dermatology, University of Utah, Salt Lake City, UT, USA; 2) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 3) Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; 4) Department for Gynecology and Obstetrics, Technical University of Munich, Munich, Germany; 5) Institute of Human Genetics, Helmholtz Zentrum München, Munich, Germany; 6) Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO) and Center for Molecular Medicine Cologne (CMMC), Medical Faculty, University of Cologne and University Hospital Cologne, Cologne, Germany; 7) Université Laval Research Centre, Québec, Canada; 8) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care and Department of Oncology, University of Cambridge, Cambridge, UK.

We have developed a variant interpretation pipeline that starts from a raw genotype file in VCF format. It conducts genotype-, variant-, and sample-wise quality control of data. This pipeline implements a novel functional consequence annotation program that annotates against the predominant transcripts whenever such information is available, chooses the most biologically relevant 5' or 3' representation for short insertions or deletions (InDel), merges multi-nucleotide polymorphisms (MNP), labels loss-of-function (LoF) variants, supports non-coding regions, and is robust to reference sequence errors. For clinical testing, this pipeline quantitatively integrates multiple deleteriousness scores, allele frequencies in different populations, co-segregation within pedigrees, and association among case-control samples to calculate a posterior probability of pathogenicity for each variant. For gene discovery research, it performs gene prioritization by integrating a region-based linkage analysis, a novel rare-variant association analysis where variants are weighted by deleteriousness and call quality, the relatedness of each gene to known disease genes within a gene-gene association network, and the differential gene expression between lesional and non-lesional tissues. In both scenarios, all components are combined in a quantitative fashion. Being light-weighted and fast with low demands on memory, this pipeline is scalable to whole genome sequencing (WGS) of a large sample of individuals that is typical of a complex disease research. Components of this framework can be assembled in various ways to accommodate different study designs and analysis goals. Using this pipeline, we have re-classified a TP53 variant of unknown significance (VUS) for Li-Fraumeni Syndrome (LFS), analyzed the whole exome sequencing (WES) of 1368 breast cancer cases and 3725 healthy controls from the PERSPECTIVE project (PErsonalised Risk Stratification for Prevention and Early deteCTIon of breast cancer), and analyzed the WES of 42 cases from 16 high-risk psoriasis pedigrees in the Utah Psoriasis Initiative project (UPI). The results demonstrated the value of the VICTOR pipeline in variant classification and gene discovery applications.

1672W

Replacing MLPA by XHMM for CNV detection on targeted panel NGS data in a diagnostic setting. *E. Voorhoeve, I. Clemens, D. van Beek, I. Bakker, Q. Waisfisz, E. Sistermans, M. Weiss, A. Maugeri.* Department of Clinical Genetics, VU medical center, Amsterdam, Netherlands.

Purpose MLPA and arrays are routinely used for the detection of clinical relevant copy number variations (CNVs). Recently several tools for the detection of CNVs on NGS read-depth data have been developed. CNV detection by read-depth methods - such as XHMM (eXome-Hidden Markov Model) - is efficient, because this type of analysis uses the NGS data already available from the targeted panels. In this study we validated the use of XHMM for the detection of CNVs in a NGS based targeted connective tissue disorder (CTD) gene panel to replace the diagnostic MLPA kits (7 genes) and extent the CNV analysis to all 78 genes in the panel. We summarize the diagnostic yield after analyzing 502 CTD samples using XHMM as the first tier test for the detection of CNVs. **Method** XHMM was optimized on 1210 samples sequenced with the CTD gene panel version 1: 1176 exons were captured in solution. Default parameter settings proved to be a good choice. For single exon CNVs a separate parameter setting was validated and implemented next to the run with default settings. The performance of XHMM was tested on 318 samples sequenced with CTD panel version 2: 1977 exons were captured in solution. This set contained 37 known CNVs detected with MLPA or array: 12 duplications, 24 deletions – all heterozygous – and 1 homozygous deletion. Both the duplications and deletions range in size from single exon to over 60 exons. Based on these analyses a reference set was built and the weak performing exons (due to pseudo genic regions and low coverage) were excluded. **Results** XHMM correctly called all 37 known CNVs, including 8 single exon deletions and 4 single exon duplications. XHMM called an additional 68 CNVs. Roughly half of these CNVs were frequently detected and thus considered to be benign or artifacts. The other half were detected in genes that didn't fit the diagnosis and were therefore not confirmed by another method. **Conclusion** CNV detection by XHMM proved to be as sensitive as the current methods. Our department has implemented XHMM analysis replacing MLPA for the clinical diagnosis of CNVs in CTD samples. Since the validation of XHMM 502 samples have been tested and revealed 15 clinical relevant CNVs, 3 of which are located in CTD genes previously not tested with MLPA: 1 whole gene and 2 single exon deletions. All 15 CNVs are confirmed with MLPA or array. Implementation of XHMM has increased the diagnostic yield and saves the costs of running diagnostic MLPA (~20.000 per year in our setting).

1673T

Cross-phenotype analyses of three clinical measures of multiple sclerosis disability suggests involvement of new biological processes. *L.J. Leung[1], F.B.S. Briggs[2,3].* 1) College of Medicine, University of Vermont, Burlington, VT; 2) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH; 3) Institute of Computational Biology, Case Western Reserve University, Cleveland, OH.

Background: Multiple sclerosis (MS) is the leading cause of acquired neurological disability in young adults; patients suffer a spectrum of impairments throughout their lifetimes. Genetic studies of MS progression are few and have not yielded definitive results. These studies primarily used the Expanded Disability Severity Scale (EDSS), and the Multiple Sclerosis Severity Scale (MSSS), a derivative. No genetic studies of 9-Hole Peg Test (9HPT) and Timed 25-Foot Walk (T25FW) impairments have been conducted. We conducted a unique cross-phenotype analysis of MSSS, 9HPT, and T25FW to identify shared and distinct pathological mechanisms. **Methods:** 1,088 white participants in the Accelerated Cure Project for MS (ACP) with epidemiologic data were available. Clinical phenotypes and date of measurement were extracted from medical records for 648 subjects: N_{MSSS}=504; N_{9PHT}=313; and N_{T25FW}=471. 9PHT and T25FW were natural log transformed. The pairwise correlation between phenotypes was 20-34%. Samples were genotyped on a Human Illumina Custom BeadChip. Analyses were restricted to genic SNPs with MAF≥0.01 and genotype rate>99% in the total sample. 55,931 SNPs in 12,887 genes were investigated; models were adjusted for ancestry, gender, birth year, onset age, and disease course (6% progressive at onset, 9PHT and T25FW were adjusted for time from onset to measurement). **Results:** The top individual associations were: MSSS: *NQO1* missense, p=3x10^{-5}; T25FW: *AOC2* missense, p=7x10^{-6}; 9PHT: *NIN* missense, p=7x10^{-6}. Surprisingly, only the *NIN* missense was associated in another phenotype (p_{T25FW}=0.002). *HLA-DRB1*15:01* (rs3135388A) was not associated with any phenotype. SNPs associated (p<0.05) with all phenotypes were within 21 genes, with multiple SNPs in *CACNA1C*, *CEP104, COL2A1, HLA-DRP2* (pseudogene) and *STAT3*. These genes suggest plausible biological rationale for associated phenotypes, e.g. *CACNA1C* associates with bipolar disorder, schizophrenia and mediates hippocampal neuron survival; *COL2A1* encodes fibrillar collagen and mutations cause 16 type II collagenopathies including osteoarthritis. Pathway analyses also illustrate distinct, and shared mechanisms. **Conclusions:** These results illustrate challenges elucidating biological process mediating progression of a complex trait, and emphasizes the need for robust/diverse phenotypes. By comparing associations across clinical measurements, we have identified novel and specific MS targets for replication and functional analyses.

1674F

INFERNO – INFERring the molecular mechanisms of NOncoding genetic variants. *A. Amlie-Wolf[1,2,3], M. Tang[2,3], P.P. Kuksa[2,3], Y.Y. Leung[2,3], B. Slaff[4], J. King[3], B. Dombroski[3], G.D. Schellenberg[1,3], L.S. Wang[1,2,3,4].* 1) Genomics and Computational Biology Graduate Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA.

   Genome-wide association studies (GWAS) have identified genetic variants associated with many diseases, but limitations remain. First, GWAS-identified variants tag linkage disequilibrium (LD) blocks of potentially functional variants, many of which are not causal. Second, most GWAS variants are in the noncoding genome and may affect transcriptional enhancers. Enhancers are context-specific and annotations are incomplete, so information must be integrated across tissue contexts and data sources to identify enhancer variants. Finally, to translate GWAS findings into therapeutics, the affected genes and expression changes underlying disease risk must be identified. We propose INFERNO, a new tool that addresses these limitations by integrating hundreds of diverse data sets to identify the causal variants within LD blocks, the enhancers they affect, the tissue context of the enhancers, the affected target genes, and the direction of these effects. Given a list of tagging variants, INFERNO uses 1,000 Genomes Project data to define LD blocks. Each linked variant is compared with: enhancer RNA transcription sites, reflecting enhancer activity, across 112 tissue types from the FANTOM5 consortium; epigenetically-defined enhancers across 127 tissues and cell types from the Roadmap Epigenomics consortium; expression quantitative trait loci across 44 tissues from the GTEx consortium; and predicted transcription factor binding sites (TFBSs) for 332 transcription factors from HOMER. Tissues or cell types from each data source are grouped into 38 broad tissue categories for comparison, and empirical p-values for the enrichment of functional overlaps in each tissue category are obtained by bootstrapping. Application of INFERNO to 19 top Alzheimer's disease (AD) GWAS variants from the International Genomics of Alzheimer's Project (IGAP) 2013 study identified significant enrichment of functional variants affecting blood/immune enhancers in 18 of the tagged regions, bolstering the hypothesis of immune involvement in AD. Four regions contained variants supported by all three functional data sources as well as predicted TFBSs, and luciferase assays validated their effects on enhancer activity. Thus, INFERNO provides an easy and powerful approach for inferring the molecular mechanisms of noncoding genetic variants. We have implemented INFERNO in an efficient Python- and R-based pipeline with source code and access to a web server version available at lisanwanglab.org/INFERNO.

1675W

Cell specific, network-based pathway analysis of GWAS data provides further insights into pathogenesis of multiple sclerosis. *S.E. Baranzini, International Multiple Sclerosis Genetics Consortium (IMSGC).* Neurology, University of California San Francisco, San Francisco, CA.

   Hundreds of genetic studies have uncovered thousands of statistically significant associations. However, the mechanistic interpretation of those discoveries is often hampered by the difficulty in translating the statistical evidence of association into affected pathways or processes that can inform the underlying biology of the disease or phenotype in question. Here we use association data from the latest GWAS in multiple sclerosis (MS) covering up to 47,351 cases and 68,284 controls to conduct a cell-specific, network-based pathway analysis to further gain insights into the pathogenesis and heterogeneity of this disease. We integrated thousands of statistically associated variants in 200 genomic regions with regulatory information from ENCODE and Roadmap Epigenomics Project data to identify the most likely genes causative of the association signals in 10 different cytotypes (T and B lymphocytes, monocytes, brain cells, etc). Overall, more than 5 million data points were used to conduct this analysis, which enabled prioritization of likely regulated genes in each region. Further integration of these with a highly curated human protein interaction network, allowed for the identification of specific disease-associated interactomes. In all cell types studied (including those from brain), the emerging networks display higher connectivity than expected by chance, suggesting that associated regions influence transcriptional expression, and thus functional connectivity of biological pathways in multiple cell types in MS. These results lend support to the hypothesis that heterogeneity of genetic variation can result in phenotypic heterogeneity at the molecular and possibly organismal level, which has important implications for patient-specific risk evaluation and personalized disease management.

1676T

Mutational burden analysis of interaction networks in inherited neuropathies. *D. Bis[1,2], F. Tao[1,2], L. Abreu[1,2], A. Rebelo[1,2], S. Zuchner[1,2], Inherited Neuropathy Consortium.* 1) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL.

Inherited peripheral neuropathies are clinically and genetically heterogeneous diseases that can cause distal muscular atrophy and sensory loss. Alleles in over one hundred different genes have been shown to cause peripheral neuropathies; yet, greater than 50% of axonal neuropathy patients do not receive a genetic diagnosis. We have been involved in successful large scale exome studies to identify additional rare monogenic causes for neuropathies. For a portion of the remaining patients, the concept of oligogenic inheritance of weaker alleles is a distinct possibility. In a deviation from the classic linkage-based and heuristic approaches to gene identification, we are performing burden analyses in a large cohort of 439 families. In 117 known neuropathy genes, we saw that neuropathy cases carried on average 3.7 rare, non-synonymous variants, while 615 non-neuropathy controls harbored 2.4 variants (p=0.003, Mann-Whiney U-test). Another 400 neuropathy exomes are being added to this analysis in the coming weeks. To expand these studies into new candidates, we have created a high confidence, neuropathy-specific protein-protein interaction network from publicly available resources. The network is composed of 114 neuropathy genes and 7,514 interacting partners connected through 30,832 experimentally determined interactions. Genes within the network that interact significantly with known neuropathy genes are potential candidates. Any candidate gene with enriched alleles in cases will be studied for strong segregating alleles in neuropathy families. We designed this approach to uncover additional weak alleles, potentially important for phenotype modifying effects or oligogenic inheritance, as well as novel neuropathy genes.

1677F

In silico functional annotation of genomic variants: Application to the Alzheimer's Disease Sequencing Project. *W. Bush[1], M. Butkiewicz[1], E. Blue[2], P. Navas[2], M. Dorschner[3], C. Kang[4], S. White[5], A. Kuzma[6], A. Renton[7], X. Jian[8], D. Koboldt[4], J. Haines[1], Alzheimer's Disease Sequencing Project.* 1) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 2) Division of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Pathology, University of Washington, Seattle, WA; 4) The Genome Institute, Washington University, St. Louis, MO; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 7) Icahn Institute for Genomics and Multiscale Biology, Mt. Sinai, New York City, NY; 8) Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center, Houston, TX.

Large-scale genome sequencing efforts in disease oriented datasets routinely identify millions of variants. However, determining which variants warrant further attention is a significant bottleneck. The ADSP Annotation Group has tackled this problem by employing a strategy of integrating *in silico* functional information and applying it to the large scale whole-exome and whole-genome sequencing efforts within the Alzheimer's Disease Sequencing Project (ADSP). The ADSP has identified 28 million variants within its study participants. Of these, approximately 5 million are novel events not reported by the 1000 Genomes Project, Exome Aggregation Consortium (ExAC), or dbSNP releases, and both the statistical analysis and the interpretation of these variants is dependent on *in silico* functional annotation. The primary goals of the ADSP Annotation Working Group are to provide consistent, version-controlled genomic annotations for all ADSP projects and analyses, to integrate new relevant data and genomic resources that become available over the course of the ADSP, and to provide variant annotations across a range of biological contexts relevant to Alzheimer's Disease (AD) biology. Toward this end, we have developed a workflow to provide investigators with predicted functional impact (from the Ensembl Variant Effect Predictor), variant allele frequencies observed in other studies (from the Kaviar database and the Wellderly Cohort), predicted loss-of-function status (from SNPEff), and multiple scoring metrics for assessment deleteriousness (including CADD, CATO, and SPIDEX scores). Annotation of non-coding regions is especially challenging; due to the incredible depth of data available across a wide variety of cell and tissue types, we have developed approaches to collapse and combine regulatory annotations and (when possible) to assign them to downstream genes using expression QTL studies from congruent cell and tissue types. Leveraging these *in silico* annotation data in conjunction with brain and other tissue-based RNA-sequencing studies, we construct context-specific annotations of ADSP variants that enable hypothesis-driven analyses, and ultimately provide new insights into the pathogenesis of AD.

1678W

EHR data extraction shows utility for microphenotypes in genetic studies. *M.F. Davis[1], S. Sutton[1], L. Bastarache[2], R. Carroll[2], J.C. Denny[2].* 1) Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Multiple sclerosis (MS) is a neurodegenerative autoimmune disease of the central nervous system with over 100 known genetic loci associated with risk of development of the disease. Disease course is incredibly varied among individuals with MS and it is unknown if disease course is also associated with genetic variation. Few genetic studies are published on the microphenotypes of MS, largely due to the difficulty in collecting long-term, detailed information of a chronic disease. We have already demonstrated the ability to extract MS microphenotypes using natural language processing approaches from 6,000 electronic health records (EHR) at Vanderbilt University, using the Synthetic Derivative. (Davis, *JAMIA* 2013) We have significantly improved these algorithms and have also identified treatment response, disease progression, and adverse drug events in our patients. Previously, our recall to extract MS diagnosis date was 33% using only referral letters as data sources. We have improved recall to 62% by using the narrative text in clinic reports (97% precision, 98% specificity). We identified diagnosis dates in 1,849 of cases. 461 of these individuals were genotyped on the ImmunoChip and passed standard quality control procedures. We performed linear regression for age at diagnosis, using the first three principle components as covariates. While no SNPs reached genome-wide significance, identifying age at diagnosis is critical for future progression studies in MS. We have shown that EHR-data extraction can create large databases that can be used for microphenotype genetic studies in MS.

1679T

A novel statistical methods for inferring causal expression-phenotype network. *Z. Hu[1,2], P. Wang[2], Y. Zhu[3], J. Zhao[3], L. Jin[2], D. Bennett[4], M. Xiong[1].* 1) Houston Health Science Center, Houston, TX; 2) Fudan University, Shanghai, China; 3) Tulane University, New Orleans, LA; 4) Rush University Medical Center, Chicago, IL.

Gene expressions are high dimensional correlated traits and strongly affect variation of multiple correlated phenotypes. However, the current statistical methods for joint analysis of gene expression and phenotypes are (1) a single variate linear model that test association of gene expressions with a single trait and (2) multivariate linear models that test association of gene expressions with multiple traits. However, both gene expressions and multiple phenotypes are highly correlated. There are lack of statistical methods to organize both phenotypes and gene expressions into networks and link expression network with phenotype network. To overcome this limitation, we first use matrix completion methods to identify a few components for representing the gene expressions in a pathway. Then, we propose to use causal graphs as a major concept and a general framework for causal expression-phenotype network analysis and develop "score and search"-based methods for exact learning causal graphs of expression-phenotype networks to find the best-scoring structures for expression-phenotype networks. As a consequences, we can infer expression-phenotype networks for whole genome gene expression and dozens of phenotypes. The proposed methods were applied to AD data with 20 phenotypes and RNA-seq data of 4,937 genes from 181 pathways measured in 448 samples. At the first stage, the matrix completion methods identifies 792 components to represent 4,937 genes. Then, 792 components and 20 traits were used to infer expression-phenotype networks. The resulting networks consists of (1) phenotype networks with 20 nodes and 33 edges and (2) expression network with 90 nodes and 144 edges and (3) expression-phenotype connect network with 77 nodes and 166 edges with FDR less than 0.05. In the expression network, 41 (28%) connections were consistent with edges in KEGG Pathway database. The proposed methods provide a highly flexible general framework for causal expression-phenotype network analysis and provide more rich information than current gene expression and phenotype analysis. The exact learning algorithms will guarantee to find optimal solutions and hence provide accurate estimations of causal graphs of expression-phenotype networks. The causal expression-phenotype networks are able to uncover the mechanism of AD development.

1680F

Annotation regression for genomewide association studies. *S. Keles, S. Shin.* Biostatistics and Medical Informatics, University of Wisconsin, Madison, Madison, WI.

A key challenge for understanding GWAS results is that a large percentage of disease-associated genomic variants (GVs) are potential regulatory variants located in noncoding regions, making them difficult to interpret. Recent research efforts focus on going beyond annotating GVs by integrating functional annotation data with GWAS to prioritize GVs. However, applicability of these approaches are challenged by high dimensionality and heterogeneity of functional annotation data. Furthermore, existing methods often assume global associations of GVs with annotation data. This strong assumption is susceptible to violations for GVs involved in many complex diseases. To address these issues, we developed a general regression framework, named **A**nnotation **R**egression f**o**r **G**WAS (ARoG). ARoG is based on finite mixture of linear regression models where GWAS association measures are viewed as responses and functional annotations as predictors. This mixture framework addresses heterogeneity of effects of GVs by grouping them into clusters and high dimensionality of the functional annotations by enabling annotation selection within each cluster. Application of ARoG to autism and schizophrenia data from Psychiatric Genomics Consortium led to identification of GVs that significantly affect interactions of several transcription factors with DNA as potential mechanisms contributing to these disorders. Most notably, these analyses identified autism variants creating FOXL1 and Nkx2-5 binding sites as potential regulatory mechanisms. Computational and external data validations indicated that ARoG enhances identification of disease associated variants that would otherwise be missed due to low power or effect sizes and generates hypotheses regarding regulatory roles of the variants.

1681W

Synaptic protein-protein interaction clustering analyses identify distinct human brain regions. *G. Labilloy[1], A. Funk[2], M.R. Heaven[3], K.D. Greis[4], J. Meller[1,5,6,7,8], R.E. McCullumsmith[2].* 1) Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH; 3) Vulcan Analytical, LLC, Birmingham, AL, USA; 4) Department of Cancer Biology, University of Cincinnati College of Medicine, Cincinnati, OH; 5) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH; 6) Department of Electrical Engineering and Computing Systems, University of Cincinnati College of Medicine, Cincinnati, OH; 7) Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH; 8) Department of Informatics, Nicholas Copernicus University, Torun, Poland.

Human brain function depends on crosstalk between layers and regions. Postsynaptic density 95 (PSD95) is a multipotent scaffolding, trafficking, and clustering protein which plays a key role in signal transduction of macromolecular and intracellular complexes within specific brain regions. We analyzed mass-spectroscopy data of 424 peptides interacting with PSD95 in four brain-regions (anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), hippocampus (HIPP), and superior temporal gyrus (STG)) of three well-matched control subjects from the Alabama Brain Collection in order to detect region-specific protein-protein interactions. We processed the raw area under the curve from mass-spectroscopy data of sample triplicates using R version 3.1.3. Each measurement vector was normalized using PSD95 intensity to account for technical variations within measurements. We used the most abundant peptide intensity (IIPGGAAAQDGR) as normalization and then Log2-transformed the data. Only peptides with at least two out of three data points available for each sample were considered. Each sample was quantile-normalized in order to account for signal amplitude variation across measurements. Remaining missing values were taken care of by imputing the mean of the two other data. Standard deviation of each peptide was used to rank and select the 200 most variable peptides. Each peptide row was then z-transformed and a cluster was created with the Heatplot function from the made4 R-package using Pearson correlation as distance function. Each triplet of technical repeats robustly clustered together. ACC and DLPFC clustered in a similar fashion across patients, whereas HIPP and STG had specific clusters. This is consistent with the fact that ACC and DLPFC are more biologically similar compared to HIPP and STG which are more unique in their biological functions and gene expressions. This clustering analysis already produced biologically coherent results. More works needs to be done in order to improve the discriminating power of the clustering, for instance by selecting the most variable peptides by brain region in order to correct for unbalanced region representation within the peptide selection.

1682T

Expanding the ENCODE Encyclopedia: Applications for annotating non-coding variants. *J. Moore[1], M. Purcaro[1], A. van der Velde[2], T. Borrman[1], Z. Weng[1].* 1) Bioinformatics & Integrative Biology, UMass Medical School, Worcester, MA; 2) Bioinformatics Program, Boston University, Boston, Massachusetts, MA.

The Encyclopedia of DNA Elements (ENCODE) Consortium has generated hundreds of high throughput genomic datasets with the goal of cataloging functional elements in the human genome. Our goal was to integrate these complex data types to annotate regulatory elements and their target genes. We then used these annotations to functionally characterize non-coding variants associated with human disease. We began by developing an unsupervised method for predicting enhancer-like regions, which we evaluated using experimentally validated data from transgenic mouse assays. Using only H3K27ac ChIP-seq and DNase-seq data, our method shows high concordance with more complex genome segmentation methods with the advantage that it can be applied across many more cell and tissue types. Our next step was to then connect these predicted enhancers to their target genes. In order to compare different methods, we created a novel benchmark dataset by integrating eQTLs, ChIA-PET clusters and Hi-C links. We developed a Random Forest (RF) based approach using features such as the distance between the enhancer and gene, DNase and H3K27ac signals, and gene expression. With this model, we saw a dramatic increase in performance compared to methods based only on signal correlation. We then applied both methods to over 100 cell and tissues types across human and mouse and created a visualization tool (http://zlab-annotations.umassmed.edu/) where users can query and download regions of interest. Using this resource, we then sought to annotate non-coding variants in the NHGRI-EBI GWAS catalog. While there were many cases of variants residing in enhancers that targeted only the nearest gene, there were some interesting exceptions. In particular, rs1250564, which is in high LD with SNPs associated with Multiple Sclerosis (MS), resides in an enhancer-like region within the *ZMIZ1* gene. In lymphoblastoid cells, whose regulatory regions were previously determined to be enriched in MS variants, this enhancer targets both *ZMIZ1* and *PPIF*. *PPIF* encodes a protein that is part of the mitochondrial permeability transition pore which supports the hypothesis that mitochondrial dysfunction may play a role in the the onset of MS. This finding demonstrates that by using the ENCODE Encyclopedia, researchers can determine target genes of non-coding variants and gain a better understanding of the underlying disease.

1683F

Elucidating the structural and molecular etiology of hereditary cystatin C amyloid angiopathy. *K. Nguyen[1,4], M. March[1], A. Gutierrez Uzquiza[1], Y. Liu[1], L. Tian[1], P. Sleiman[1,2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Center for Dynamic Data Analytics, Rutgers University, Piscataway, NJ.

Hereditary cystatin C (CST3) amyloid angiopathy (HCCAA), also known as hereditary cerebral hemorrhage with amyloidosis – Icelandic type (HCHWA-I), is a rare, orphan disease that can cause a progressive loss of intellectual function (dementia), stroke, and other neurological disorders. It has a well-documented pathophysiology. However, its pathogenesis involving amyloid deposition is still unknown. A tandem effort involving genomic/genetic analyses and a therapeutic-driven approach were used to elucidate interactions at molecular resolution while providing viable ligands as drug candidates in a translational and precision medicine pipeline. A series of monomeric to oligomeric systems were modeled using classical simulation and enhanced sampling methods. The latter was done to quantitate the binding affinities of small low-weight compounds to large peptides satisfying the pharmacogenomic/-genetic approaches in this pipeline. Pathogenicities through mutation prediction methods were also used to initially determine the overall effect of conformational (in) stabilities. The dynamics and energetics – as well as the investigation of the partial unfolding mechanism and domain swapping for oligomerization – were observed using these methods. The L68Q mutant in CST3, as well as the V57[D,N,P], L47C/G69C, and F29C/M110C variants – were analyzed for overall comparison. As a result, a significant amount of conformational sampling was required to render the electrostatic behavior for ligand-protein and protein-protein interactions, steric interactions from the ligand and protein side chains, thermodynamically-driven processes leading to high-order and disordered systems, and long-range hydrophobic/-philic interactions from the bulk water. While these subjects are fundamentally controversial, they are required to observe the biological molecular recognition of any ligand and its respective macromolecule. Results of these *in silico* methods were also used to generate vibrational and electronic spectra, which will then be compared experimentally for validation.

1684W

NIA Genetics of Alzheimer's Disease Data Storage site (NIAGADS): 2016 update. *L. Wang[1,2,3], P. Gangadharan[1,2], A. Kuzma[1,2], O. Valladares[1,2], E. Greenfest-Allen[2,4], D.M. Childress[1,2], R. Cweibel[1,2], J. Malamon[1,2], H. Lin[1,2], Y. Zhao[1,2], L. Qu[1,2], Y.Y. Leung[1,2], A. Naj[5], C.J. Stoeckert Jr.[2,4], G. Schellenberg[1,3].* 1) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine; 2) Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine; 3) Institute on Aging, University of Pennsylvania Perelman School of Medicine; 4) Department of Genetics, University of Pennsylvania Perelman School of Medicine; 5) Department of Epidemiology, University of Pennsylvania Perelman School of Medicine.

Statement of Purpose: NIAGADS is a national genetics data repository that facilitates access of genotypic data to qualified investigators for the study of the genetics of late-onset Alzheimer's disease (AD) as well as other neurological diseases. Collaborations with large consortia such as the Alzheimer's Disease Genetics Consortium (ADGC) and the AD Sequencing Project (ADSP) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed by the research. NIAGADS' position as the centralized location for the storage of AD genetic data and its collaboration with the National Cell Repository for AD (NCRAD) to make available data and samples makes NIAGADS essential to the scientific community's achievement of the National Alzheimer's Project Act's (NAPA) research goal to discover therapeutic targets for the treatment and prevention of AD by the year 2025. **Methods:** Since 2012, NIAGADS has been supported by National Institute on Aging (NIA) under a cooperative agreement (U24 AG041689). This agreement expanded the NIAGADS storage and data sharing capacities for large-scale sequencing projects. All data derived from NIA funded AD genetics studies are expected to be deposited at NIAGADS or another NIA approved site. NIAGADS has partnered with the database of Genotypes and Phenotypes (dbGaP) and the Sequencing Read Archive (SRA) in this effort. In order to facilitate research, secondary data, including GWAS summary statistics and imputation data, as well as deep phenotypes, have expanded the existing ADGC collection. **Summary of Results:** As of June 2016, NIAGADS houses 37 datasets with >54k subjects and over 30 billion genotypes. With the completion of the Discovery Phase of ADSP, qualified investigators can retrieve sequencing data with ease and flexibility using the ADSP website and data portal (collaboration with dbGaP/SRA). ADSP generated sequencing data of >12,000 subjects and is currently selecting ~3000 additional subjects for WGS follow-up; BAMs and quality controlled VCF files are available through the ADSP portal and dbGaP (https://www.niagads.org/adsp/content/acknowledgement-statement). The redesigned NIAGADS Genomics Database provides a searchable annotation resource that links public NIAGADS resources to AD-relevant sequence features and genome-wide annotations. NIAGADS is a rich resource for AD researchers to utilize; datasets, guidelines, and new features are available on our website at https://www.niagads.org.

1685T

Genomic signatures in bone-marrow mesenchymal stem cells differ by disease course in Multiple Sclerosis. *P. Wilkinson[1], M. Cameron[1,3], S.M. Planchon[2], J.A. Cohen[2], F.B.S. Briggs[1,3].* 1) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH; 2) The Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic, Cleveland, OH; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

Background: It is hypothesized that failures in endogenous repair mechanisms may mediate the transition from relapsing remitting (RR) to secondary progressive (SP) disease course in multiple sclerosis (MS) patients. This has precipitated efforts to explore several types of stem cells as experimental therapeutics. Adult mesenchymal stem cells (MSCs), including bone marrow (BM) MSCs, are immunomodulatory and neuroprotective. We hypothesize that these process are dysregulated/diminished in BM-MSCs of SP vs RRMS patients. **Methods:** BM-MSCs were collected, isolated and cultured using standard procedures. Cells collected at passage 1 from 3 non-Hispanic white RR cases and 3 race, age (+/- 5 yrs) and gender frequency-matched SP cases. RNA was isolated using the Qiagen AllPrep procedure. RNA-seq data was generated via an Illumina HiSeq 2500 (paired-end, 100 cycle, 50×10^6 mapped reads/sample); and were aligned using *STAR aligner* and counted with *HTSeq*. Genes with at least 1 count/10^6 for 3 subjects were analyzed using *EdgeR*. Pathway enrichment was evaluated using DAVID v6.7 of top 2% ranking (nominal p) under- (UE) and over-expressed (OE) genes in SP compared to RR patients. Global pathway activity of all genes was evaluated using gene set variation analysis (GSVA). **Results:** In BM-MSCs, multi-dimensional reduction differentiated SP from RR patients. Genes involved in axon guidance (fold enrichment (FE)=7.6; p=0.0004) and dopamine signaling (FE=5.1; p=0.007) were UE in SP vs RR patients. OE genes were involved in arachidonic acid metabolism (FE=9.3, p=0.0002). Globally, the activity in several pathways and processes differed between SP and RRMS, including: neurotransmitter activity (p=0.001), interleukin receptor activity (p=0.003), arachidonic acid metabolism (p=0.009), synaptic regulation (p=0.009) and antigen processing and presentation (p=0.02). **Conclusions and Ongoing analyses:** These results demonstrate that both immunomodulatory and neuroprotective mechanisms have differential activity across RR and SPMS; particularly under-expression of genes involved in neuronal process in SPMS. We also hypothesize there may be shared genomic signatures between BM-MSCs and PBMCs of MS patients. Analyses of RR/SP PBMCs from the same subjects are underway, including a weighted gene co-expression network analysis aimed to identify overlapping gene co-expression modules across the two tissue types to reveal higher order transcriptomic organization.

1686F

Identification of novel genetic risk factors in sporadic ALS using discordant monozygotic twins. *G.H.P. Tazelaar[1], W. van Rheenen[1], M. Seelen[1], M. Moisse[2], P. Hop[1], R.L. McLaughlin[3], A. Vajda[3], C. Neuwirth[4], M. Weber[4], V. Silani[5], A. Chio[6], M. de Carvalho[7], J.S. Mora Pardina[8], J.M. Ravits[9], O. Hardiman[10], P.M. Andersen[11], P. van Damme[2], L.H. van den Berg[1], J.H. Veldink[1], M.A. van Es[1].* 1) Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 2) KU Leuven (University of Leuven), Department of Neurosciences, Experimental Neurology, Leuven, Belgium; 3) Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland; 4) Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, St. Gallen, Switzerland; 5) Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy; 6) Department of Neuroscience, University of Torino, Torino, Italy; 7) Department of Neurosciences, Hospital Santa Maria, Centro Hospitalar Lisboa Norte, Lisbon, Portugal; 8) ALS Unit, Department of Neurology, Hospital Carlos III, Madrid, Spain; 9) Department of Neurosciences, University of California, San Diego, USA; 10) Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland; 11) 11.Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden.

Amyotrophic lateral sclerosis (ALS) is a fast progressive and lethal neurodegenerative disease with a lifetime risk of 1 in 400. Next-generation sequencing of pedigrees and Genome-wide association studies (GWAS) have shown to be successful tools in identifying genetic risk factors in ALS. However, especially in the case of sporadic ALS, where there are no other affected family members, these known factors only partially explain the genetic contribution as estimated by ALS concordant and discordant twin studies. This discrepancy could possibly be attributed to rare genetic variants that might be caused by *de novo* mutations. In our study, we set out to identify novel (epi)genetic risk factors for sporadic ALS by analyzing the methylation patterns and genomic DNA from whole-blood of 21 monozygotic twin pairs discordant for ALS. In order to discriminate between early and late postzygotic mutations, we additionally analyzed DNA obtained from fibroblasts and saliva in a subset of twins. So far, we have investigated the single nucleotide variants (SNVs) using whole genome sequencing. To increase the chance of detecting *de novo* SNVs, we used three different variant calling algorithms and build a custom-filtering pipeline to obtain the most reliable discordant variants. In total we identified 149,388 SNVs of which 129,387 were eligible for further validation on a custom Axiom Genotyping Array. Preliminary results of this first step validation show a mean of 3.0 (SD 3.2) *de novo* SNVs per individual. This post-zygotic mutation rate however, was determined after exclusion of three outliers. One non-ALS twin had 690 probable *de novo* mutations, which might be attributed to its comorbidity, namely leukemia. Notably, the other two outliers were both ALS affected twins, with 42 and 77 probable *de novo* SNVs. After exclusion of the twin with leukemia, we saw no significant difference in *de novo* mutation rate between ALS and controls, but we did find a significant inequality of variance (p=0.02, Levene's test). Additionally, analysis of mosaicism in one of the two ALS outliers indicates that at least 5 of the 42 SNVs are shared between blood, saliva and fibroblasts, pointing towards early postzygotic mutations. Further validation and analysis is required to confirm these *de novo* mutations and determine their possible significance as a genetic risk factor for ALS.

1687W

Are *in silico* pathogenicity tools reliable indicators of variant pathogenicity? A retrospective look at data from 46 *de novo* pathogenic missense variants identified in a clinical diagnostic setting. *C. Stanley, H. Hornung, D. Pineda-Alvarez, A. Zare, T. Foss, J. Warner, D. Smith.* Chief Director, Clinical Genomics, Courtagen Life Sciences Inc, Woburn, MA.

Next generation sequencing identifies a large number of sequence variants, which are evaluated by clinical geneticists for likelihood to cause disease based on evidence-based guidelines issued by the American College of Medical Genetics and Genomics in 2015. *In silico* prediction tools are utilized, within these standards, as a supporting line of evidence for predicting variant pathogenicity, albeit a lower-weighted line of supporting evidence. We evaluated four prediction tools (Mutation Assessor, PolyPhen2, MutationTaster and SIFT) to test *de novo* pathogenic and likely pathogenic missense variants identified in patient samples submitted for evaluation with our large epilepsy gene panels. The variants identified were in genes associated with diseases that correlated to the patient's clinical phenotype. The variants were also absent from control populations and, therefore, were not used as training sets for the development of *in silico* tools. Our results indicated that the majority (96%) of pathogenic *de novo* variants identified were considered deleterious by *in silico* prediction tools and support their utilization as a line of evidence for variant classification.

1688T

Scalable map-free typing of clinical and structural variants across heterogeneous whole-genome sequence data. *W. Salerno[1], S.N. Shekar[2], L. Herta[2], W. Nassar[1], A. English[1], A. Mangubat[2], J. Bruestle[2], E. Boerwinkle[1], R.A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Spiral Genetics, Seattle, WA.

The complexity and variety of SVs pose specific challenges to clinical whole genome applications, which require high sensitivity, low false discovery rates, rapid turnaround time and the ability to distinguish de novo and shared variation. Here we present a compressed read data format (Biograph) that allows for rapid and scalable typing of whole-genome variation. In this format, read data can be directly queried by locally assembling full reads across breakpoints, fusion junctions and novel insertions. Agnostic to reference genomes, Biograph data avoid mapping-based artifacts and thus mitigates the effects of heterogenous insert-size distributions, read length and coverage. PCA of SV calls generated from from three distinct sequencing and mapping protocols across an aggregate 578 samples and 11 detection methods reveals clear grouping by protocol and even protocol-specific substructure. Three samples were each subject to all three protocols, and on average 32% of SV calls previously described are shared across protocol replicates. Analysis of the same read data in the Biograph format shows that >96% of 30-mers occurring in at least 4 reads were common to protocol replicates. Position-based SV typing of 7393 events returns consistent breakpoints across protocol replicates for all positions with at least 4x coverage. The Biograph typing of all 7393 events across 9 individuals (66,537 events in total) was performed in less than 21 minutes. To examine sequence-based variant typing, we populated the GRCh38 FASTA with 11,511 ClinVar SNVs and simulated a 30x WGS FASTQ with ART Illumina, which was then converted to Biograph format in less than 20 hours on a 32-core machine. The Biograph data were queried by assembling reads over the SNVs, typing all the events in less than 2 minutes. Efforts are underway to spike-in all ClinVar variants, including SVs, into a single "worst genome" to generate all possible variant characteristic sequences. Biograph data can be batched across samples, with per-sample footprint scales less than linearly with number of samples. The aggregate read data for 125 30x whole-genomes was ~300 GB: <2 GB per sample of sample-specific information and ~100 GB of aggregate variant information. This lightweight multi-sample "sequence set" provides implicit anonymization and rapid whole-genome variant typing for groups of samples, suggesting a natural way for clinical sites to compare patient variation profiles without raising privacy concerns.

1689F

Genome wide expression analysis indicates the role of molecular binding genes in tumor-induced epilepsy. *M.R. Rajeswari, H.N. Singh.* Biochemistry, All India institute Of Medical Sciences, New Delhi, Delhi, India.

Epilepsy is a common cause of morbidity affecting approximately one third of patients with primary brain tumors. However, the disease mechanisms of tumor induced epilepsy are barely explored. The alteration in peritumoral microenvironments might play a crucial role in inducing epilepsy by means of gene dysregulation. Therefore, we have performed genome-wide gene expression study to find altered genes in the tumor induced epileptic human brain cells. The microarray dataset GSE32534 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32534) was retrieved from the GEO database, which includes mRNA expression data of peritumoral cortex tissue from 5-seizure vs. 5-non-seizure low grade brain tumor patients. The bioconductor "affy" package was used to calculate mRNA expression and further pair-wise t-test was applied to find the significant differential expression (p<0.05). Several genes were found to be altered in the tumor induced epilepsy, however, only seventeen genes (LRRN4CL; CP; MIR143HG; FCGR2A; COL3A1; MGP; SP100; CYTH1; IL1R1; DKK1; PHLDA2; COL14A1; CFH; LYZ; SOAT1; CNKSR3; EXPH5) were observed with 1.5 fold expression change and showed a high concordance. Gene ontology analysis showed that significant portions of the differentially expressed genes functioned as molecular binding including enzyme binding and transcription factor binding and were found to be involved with the epilepsy disease mechanisms. In conclusion, our study showed that dysregulation of gene expression in the peritumoral tissues may be one of the major mechanisms of brain tumor induced-epilepsy. However, further validations are required and identified genes might be helpful to understand the disease etiology leading to management of tumor induced epilepsy.

1690W

SMRT sequencing of STR expansions in SCA31 brain disease. *W. Qu[1], H. Ishiura[2], K. Doi[1], D. Hsu[3], J. Yoshimura[1], J. Mitusi[2], M. Boitano[3], J. Korlach[3], S. Tsuji[2], S. Morishita[1].* 1) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, 277-8583, Japan; 2) Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-8655, Japan; 3) Pacific Biosciences, Menlo Park, CA 94025, USA.

Expansions of short tandem repeats (STRs), DNA repeats of 2-6 nt, have been reported to be associated with some brain diseases. We studied a well-characterized sampleknown example, spinocerebellar ataxia type 31 (SCA31) that has ~3 kbp expansions of three STR types (TAGAA, TGGAA, and TAAAATAGAA) in an intron of genes BEAN1 and TK2 in human chromosome 16. To determine the sequences of the focal STR in SCA31 samples, we amplified the STR and performed SMRT sequencing in 2014 and 2016, independently. As the performance of PacBio RS II has remarkably improved for the last couple of years, we could obtain a sufficient number of ~5 kbp subreads that covered the STR regions of length ~3 kbp in six SCA31 samples. We assembled these long subreads into a contig with the STR and determined the sequences of the STR regions in the six individuals. Most of the STR is heterozygous and is found in one of the two alleles. To confirm this property, we calculated the length of an STR occurrence in each raw subread and generated the length distribution of STR occurrences in raw subreads. If the STR is heterozygous, we should detect one peak in the distribution that represents long STR occurrences in one allele and another peak that indicates short, normal STR occurrences in the other allele. Indeed, we observed such bimodal peaks in five of the six individuals. In the remaining one sample, the STR was homozygous because a single peak was identified in the distribution. Overall, our analysis demonstrates that SMRT sequencing enables to sequence long STR occurrences and determine their heterozygosity, simultaneously.

1691T

A multikernel approach for associating statistical interactions between genomic layers to phenotypes. *B. Ng[1], C. White[2], H.U. Klein[2], J. Xu[2], E. Patrick[2], L. Chibnik[2], C. Gaiteri[3], D.A. Bennett[3], P.L. De Jager[2], S. Mostafavi[1].* 1) Statistics, University of British Columbia, Vancouver, BC; 2) Neurology, Harvard Medical School, Boston, MA; 3) Neurological Science, RUSH University, Chicago, IL.

Introduction: The availability of large-scale multi-omic data provides a unique opportunity to investigate how interactions (i.e. epistasis) between various regulatory mechanisms give rise to complex phenotypes. However, majority of existing analysis approaches are ill-suited for studying the statistical interactions between different types of genomic measurements in a genome-wide manner. The classical approach of applying an univariate test to assess the association between each possible pair of genomic variables in relation to a phenotype would result in an explosion in the number of tested hypotheses. Increasing statistical power via devising new dimension reduction strategies that retain biological interpretability is thus crucial. Methods: We present here a multikernel machine (MKM) approach for associating the interactions between genomic layers and phenotypes. Given a set of variables, e.g. a set of SNPs within a predefined genomic window, the idea behind KM is to first estimate the similarity between all subject pairs based on their SNP patterns (alleles) in the given window. Variable sets with similarity kernels that significantly explain the variance in the observed phenotype are declared as relevant. To model interactions between genomic layers, we draw on properties of the reproducing kernel Hilbert space. Specifically, an interaction kernel between e.g. SNP and methylation, can be built by taking the Hadamard product of their respective main effect kernels. We restrict each variable set to variables within a certain window from a gene to facilitate gene level interpretation. With this approach, the number of hypotheses effectively reduces to the number of genes, which substantially increases statistical power. Results: We applied MKM to data from the ROS/MAP studies, which comprise genotype, DNA methylation, and histone modification data derived from the dorsolateral prefrontal cortices of 400 Alzheimer's disease (AD) subjects. The subjects were also phenotyped for traits, such as cognitive decline, amploid beta, and tau. We focused on modeling the interactions between SNP and DNA methylation as well as SNP and histone modification. In addition to significant main effects of each data type, we identified a number of genes where only the interactions between SNP and a cis cellular trait were significantly associated with AD phenotypes. Our results thus demonstrate the importance of additionally examining the interactions between genomic layers.

1692F

Disease profiling with multi-omics data: An application in multiple sclerosis. *Y. Natanzon[1], S.L. Andersen[2], J.H. Winnike[3], S. Maichle[5], X. Zhang[6], K.J. Knagge[3], L.K. Newby[7], F.B. Briggs[1], S.G. Gregory[2,4].* 1) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Multiple Sclerosis Research Group, David H. Murdock Research Institute, Kannapolis, NC; 3) Analytical Sciences Laboratory, David H. Murdock Research Institute, Kannapolis, NC; 4) Duke Molecular Physiology Institute, Duke University, Durham, NC; 5) Duke Translational Research Institute, Duke University, Durham, NC; 6) Departments of Chemistry, Pharmacology & Toxicology, University of Louisville, Louisville, KY; 7) Division of Cardiovascular Medicine, Duke Clinical Research Institute, Duke University Medical Center, Durham, NC.

Multiple sclerosis (MS) is a complex neurodegenerative disease with variable pathoetiology. Understanding the genetic architecture underlying MS onset is critical to unraveling potential downstream drug targets. Genome-wide association studies have been successfully identified >120 MS risk variants, however, the functional mediating MS onset is not yet clear. Despite several studies investigating potential biological markers (biomarkers) for MS diagnosis, no biomarker or sets of biomarkers have been identified to have sufficient diagnostic accuracy. Among the available *omics* data for biomarker development, metabolites are biologically the closest to the disease state; however, metabolites are an inherently noisy data. The goals of our study are two-prong: 1) create a metabolomic profile that discriminates MS cases from healthy controls; 2) refine MS phenotype by effectively combining multi-*omics* data to reduce the genetic search space and generate a hypothesis of biological mechanisms contributing to MS onset. Data on 427 metabolites was generated on 109 individuals (54 drug-naïve MS cases with 55 age-, sex-, and race-matched controls) using GCxGC-MS, LC-MS and Lipidomic analysis of serum. Genome-wide expression data and genotyping of 25 GWAS-confirmed SNPs was also available in 13 males in the sample (6 cases and 7 controls). We first performed Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) to select a metabolite set that best discriminates MS case status. OPSL-DA method captures the variability in the metabolites related to MS cases status while eliminating the variability unrelated to case status. The OPSL-DA model identified 136 metabolites that statistically significantly discriminated between cases and controls based on VIP Score of ≥ 1. We selected the top 50 metabolites and performed principal component analysis (PCA) for further dimension reduction. We performed an association between the three PCs and a genetic risk score compiled from the 25 known MS SNPs. We also performed an association between the PCs and the 364 differentially expressed genes between MS cases and controls. Our analysis was able to generate a MS metabolic profile using PCA. Metabolites clustered into functional groupings, of which several are relevant to MS pathogenesis, including neurotoxicity, immunomodulation, and oxidative stress. We observed several associations between the metabolic profile, differentially expressed genes and known MS SNPs.

1693W

Post-variant calling quality control (QC) pipeline and multi-pipeline genotype consensus caller for large-scale whole genome and whole exome sequencing studies. *J. Malamon[1], A.C Naj[2], M. Schmidt[3], B.N. Vardarajan[4], Y. Ma[5], D. Lancour[5], A. Kuzma[1], O. Valladares[1], L. Qu[1], H. Lin[1], E.R. Martin[3], L.-S. Wang[1], J.L. Haines[6], R. Mayeux[4], L.A. Farrer[7], M.A. Pericak-Vance[3], G.D. Schellenberg[1], Coordinating Center for Genetics and Genomics of Alzheimer's Disease.* 1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA., Select a Country; 2) Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania Philadelphia, PA, USA; 3) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 4) Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 5) Biomedical Genetics Program, School of Medicine, Boston University, Boston, MA, USA; 6) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 7) Departments of Medicine, Neurology, Ophthalmology, Genetics & Genomics, Epidemiology, and Biostatistics, School of Medicine, Boston University, Boston, MA, USA.

While pipelines exist for quality control (QC) of upstream processing of whole genome sequencing (WGS) and whole exome sequencing (WES) data, little work has been done to develop and implement QC on the downstream genotype data. Also, while systematic integration of genotype calls from multiple pipelines ensures improved overall genotype quality, limited progress has been made in developing approaches to implement "consensus calling". To analyze WGS of 578 family samples and WES of 10,692 case/control samples from the Alzheimer's Disease Sequencing Project (ADSP), structured protocols were developed and implemented for post-variant calling QC and consensus calling, and will be implemented on multiple replication sequence datasets. To provide a tool for implementing QC and consensus calling on these datasets, the Coordinating Center for Genetics and Genomics of Alzheimer's Disease (CGAD) set out to construct a pipeline based on ADSP QC/consensus calling best practices extensible to other large-scale sequencing studies. The CGAD pipeline applies QC apropos of multiple calling pipelines accounting for center-specific effects. After initial sequencing center-recommended filtering is applied, additional variant-level metrics (e.g., call rate, average read depth) are computed across all subjects for each bi-allelic single nucleotide variant or insertion-deletion, and filters specified *a priori* are applied to exclude low-quality variants. Subsequent sample-level QC examines within-subject distributions of metrics (e.g., average read depth, Ti/Tv ratio), and outliers (by center or ethnic group), and based on genome-wide sample summaries, potential low quality samples are excluded. The resulting set of QCed calls from each pipeline are then combined via a consensus protocol to integrate (a) genotypes concordant between pipelines and (b) high quality genotypes unique to one caller. For consensus calling, quality metrics such as "GQ" from GATK are used to establish additional genotype-specific filters on pipeline-unique variants. The pipeline generates fully-QCed/consensus-called genotypes in multiple formats, as well as extensive annotation of variants and genotypes both passing and failing the QC/consensus calling steps. The pipeline is currently in beta-testing, and pipeline design and results of the implementation of the pipeline on both ADSP Discovery and preliminary ADSP replication datasets will be presented.

1694T

Prioritization of ASD-associated genes by variant annotation reveals trends in rare and common variant discovery. *E. Larsen[1], I. Menashe[2], W. Pereanu[1], S. Banerjee-Basu[1].* 1) MindSpec Inc., McLean, VA; 2) Ben-Gurion University of the Negev, Beer-Sheva, Israel.

The search for genetic factors underlying autism spectrum disorders (ASD) has led to the identification of hundreds of genes containing thousands of variants that differ in their mode of inheritance, effect size, frequency and function. These data are summarized in our Autism Database (AutDB; also known as SFARI Gene), an open-access database for genetic variations associated with ASD. However, a major challenge in the field of ASD biology involves assessing the collective genetic evidence in an unbiased, systematic manner. We previously described a novel scoring algorithm for the prioritization of candidate genes based on the cumulative strength of evidence from each ASD-associated variant in curated in AutDB. Under our algorithm, each individual variant is manually annotated with multiple attributes extracted from the original report, followed by score assignment using a set of standardized scoring parameters that were summed up to yield a single score for each gene in the database. Here, we present the results using an updated dataset of ASD-associated rare and common genetic variants. A total of 534 annotated research articles were analyzed to generate a dataset of 3283 ASD-associated rare variants and 840 ASD-associated common variants distributed across 654 candidate genes (March 30, 2016 data freeze). There were remarkable variations in gene scores resulting in a log-normal distribution of scores with a mean gene score of 16.93 ± 32.68. We were able to identify a collection of high confidence candidate genes with scores deviating more than two standard deviations (SDs) from the mean score of all genes, with very high scores for eight genes (*ADNP*, *CHD8*, *DYRK1A*, *POGZ*, *PTEN*, *SCN2A*, *SHANK3*, and *SYNGAP1*). The gene scores generated by our approach were once again significantly correlated with those in the SFARI Gene scoring module, indicating a strong agreement between gene prioritization using our approach and the expert-mediated SFARI Gene scoring initiative. Finally, we performed a time course analysis of ASD-associated variants and identified important trends in rare and common variant discovery. Altogether, our scoring algorithm provides a framework for assessment of diverse types of genetic variants associated with ASD that are likely to be important for defining the genetic risk architecture in ASD.

1695F

The contribution of short tandem repeats to *de novo* **variation.** *T. Willems[1,2], D. Zielinski[2], M. Gymrek[3], Y. Erlich[2,4,5].* 1) Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA; 2) New York Genome Center, New York, NY; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY; 5) Center for Computational Biology and Bioinformatics, Columbia University, New York, NY.

Recent studies have obtained genome-wide estimates of the number of de novo mutations for nearly every class of genetic variant, ranging from SNPs and indels to copy number variants. Despite these results, no such estimate exists for short tandem repeats (STRs), highly repetitive loci that mutate orders of magnitude more rapidly than most genetic elements. STRs have been implicated in over 40 Mendelian diseases and evidence of their involvement in complex traits continues to accumulate, but current bioinformatics tools are too error-prone to reliably identify de novo STR mutations. To address this issue, we developed HipSTR, a state-of-the-art STR variant caller. HipSTR learns an individual error model for each STR in the genome and uses a specialized hidden Markov model to determine the STR's sequence and size. Using gold standard datasets, we demonstrate that this approach generates substantially more accurate STR genotypes than all existing variant callers, include GATK Haplotype Caller, Platypus, freebayes and various STR-specific tools. Empowered by this improved accuracy, we used HipSTR and datasets from the Illumina Platinum Genomes and Genome in a Bottle projects to genotype STRs in a deeply sequenced trio. This analysis uncovered hundreds of putative de novo STR mutations, 85% of which replicated in an orthogonal dataset. We selected a handful of the replicated de novo mutations for Sanger sequencing and in all cases, the de novo mutation experimentally validated, illustrating the extreme precision and accuracy of HipSTR. Lastly, to characterize de novo STR variation on a population scale, we used HipSTR to analyze high-coverage whole-genome sequencing data for 500 quads in which one child has been diagnosed with autism. Collectively, our work highlights the substantial contribution of STRs to de novo variation in healthy individuals and in individuals with autism spectrum disorder.

1696W

Convergent biomolecular mechanisms between bipolar disorder and schizophrenia revealed by multiscale modeling of data from GWAS, eQTL, GO and ENCODE. *J. Berghout[1,2,3], H. Li[1,2,3], I. Achour[1,2,3], J. Li[1,2,3], L.L. Pesce[4,5], I. Foster[4,6], Y.A. Lussier[1,2,3,4,7].* 1) Center for Biomedical Informatics and Biostatistics, The University of Arizona, Tucson, AZ, USA; 2) BIO5 Institute, The University of Arizona, Tucson, AZ, USA; 3) Department of Medicine, The University of Arizona, Tucson, AZ, USA; 4) Argonne National Laboratory and University of Chicago, Chicago, IL, USA; 5) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 6) Department of Computer Science, University of Chicago, Chicago, IL, USA; 7) Institute for Genomics and Systems Biology, Argonne National Laboratory and University of Illinois at Chicago, Chicago, IL, USA.

Schizophrenia and bipolar disorder are highly comorbid conditions with strong evidence for shared genetic etiology underlying much of the heritable disease risk. We computationally modeled high dimensional data from genome-wide association studies (GWAS), expression quantitative trait loci (eQTL), gene product functional annotations (GO), and high-throughput locus functional annotations (ChIP-Seq, ENCODE project) to explore biological similarity and convergence between all possible pairs of GWAS polymorphisms associated with either or both diseases. This strategy allowed us to uncover high confidence and recurrent biological signals from intergenic SNPs as well as those within candidate gene regions. SNP pairs where both polymorphisms had been mapped to schizophrenia alone reproduced a strong antigen presenting and immune/inflammatory process signal driven by the Major Histocompatibility Locus (MHC: Chr 6: 29Mb-33Mb) that was not found in bipolar disorder. However, most other biological convergence was found between the two diseases, or for SNP pairs that were mapped in comorbid samples. This underlines the molecular and mechanistic similarity for bipolar disorder and schizophrenia even in the absence of GWAS positional overlap. 384 SNP-SNP pairs (involving 66 unique SNPs) had statistically significant evidence that they each influenced a common biological process or were each bound by the same transcription factor. 16 pairs of SNPs were found to regulate the expression of at least one of the same mRNA transcripts via eQTL, including four SNP-SNP pairs mapped across two different chromosomes and acting in *trans* (regulating expression of *RAB11FIP4*, *MED29* and *HIST1H2BN*). Common biological process annotations to both diseases included additional immune responses, chromatin architecture and histone biology, with overrepresentation of binding by the transcriptional repressor *CTCF* and chromatin remodeling factor *FOXA1*. The multiomic approach used in this case study is extensible to other diseases and syndromes, with a data management system currently being built to mine preprocessed results from 400 additional GWAS traits. .

1697T

CalibrE: Data-driven methodology for the calibration and comparison of RNA-seq data across tissues and cell types. *A.E. Byrnes[1,2,3], S. Lodato[3,4], J.M. Bloom[1,2,3], A. Bloemendal[1,2,3], C. Seed[1,2,3], P. Arlotta[3,4], B.M. Neale[1,2,3].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Medical and Population Genetics Program, The Broad Institute, Cambridge, MA; 3) The Stanley Center for Pyschiatric Reseach, The Broad Institute, Cambridge, MA; 4) Harvard Stem Cell Institute, Harvard University, Cambridge, MA.

RNA-seq is emerging as a powerful resource to aid in the interpretation of GWAS and sequencing studies. In particular, evaluating expression of GWAS genes may yield insights into which cell types and tissues are important in pathogenesis. However, a central impediment to realizing the value of RNA-seq data is that the variability in transcriptional landscape across different cell and tissue types is not well estimated and by extension controlled for in analysis. We present a data-driven statistical method for calibration and comparison of RNA-seq data from multiple cell- or tissue- types in order to correct for possible differences in overall transcription rate and transcriptional diversity. Our method, CalibrE, is based on the simple assertion that genes co-expressed as a coherent unit can be used to calibrate across tissue types; these genes must therefore have relatively constant ratios of expression to each other and so we use weighted graphical modeling to find such clusters for calibration. In both the Genotype Tissue Expression (GTEx) and Brainspan data, we find a cluster of genes encoding ribosomal components that are consistently co-expressed and as a consequence can be used to calibrate the overall transcriptional landscape of cells and tissues. We contrasted CalibrE with RPKM and TMM by assessing the rank-based correlation between each normalization approach on RNA-seq data with conventionally normalized expression microarray data on the same samples. We show that on average, CalibrE normalization correlates most strongly with the microarray data, which is less subject to sampling inherent in RNA-seq analysis. Ultimately, we aim to use RNA-seq data to prioritize cell types and tissues based on the relative abundance of expression of genes identified by genetic studies. We have performed CalibrE on a set of previously unpublished RNA-seq data from single neuron subtypes, at a variety of time points, isolated from mice. From these analyses, we see a clear enrichment of genes identified as having two or more loss-of-function (LOF) *de novo* mutations in a study of Autism Spectrum Disorder (ASD) in prenatal cerebral projection neurons, compared to other neuron types. This provides novel insights into the etiology and developement of ASD.

1698F

Genotyping and visualizing copy number variation from whole genome sequencing data. *J.T. Glessner[1,2], R.L. Collins[1,3], M.R. Stone[1], H. Brand[1,2], M.E. Talkowski[1,2,4,5].* 1) Psychiatric and Neurodevelopmental Genetics Unit, Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Department of Neurology, Harvard Medical School, Boston, MA 02115, USA; 3) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA 02115, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02141, USA; 5) Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

Copy number variation (CNV) is a highly penetrant source of genetic risk for human disease. Despite recent progress, routine detection methods remain limited for CNVs below the resolution of a traditional chromosomal microarray (<40kb), a subset encompassing >95% of all CNVs in an average human genome. Whole-genome sequencing (WGS) is the currently most sensitive approach for CNV discovery but many current algorithms suffer from low sensitivity and/or specificity. To refine and validate CNV signatures from WGS data *in silico*, we devised a CNV genotyping algorithm extensible to biallelic, complex, and multiallelic CNVs (i.e. copy number states >3 at a single locus). When evaluating a putative CNV site, our genotyping algorithm can be executed in two complementary modes. The first is designed for characterization of multiallelic CNV sites and uses c-means fuzzy clustering to yield sample-based copy state predictions and genotype probabilities. When applying this method to putative CNV calls from a WGS dataset of 160 human samples, we observed evidence that up to 23% of duplication loci may be multiallelic, which is a blind spot for most existing CNV algorithms. The second mode evaluates biallelic CNV sites using a unidirectional t-test with p-value permutation to evaluate the difference between a subset of samples predicted to harbor a CNV at the specified locus versus those samples predicted not to harbor the CNV. This test is modified for singleton CNVs where discriminatory power is insufficient (power<0.8). To assess the performance of our biallelic CNV genotyping algorithm, we used large-insert jumping library sequencing and chromosomal microarray as validation datasets for the WGS cohort of 160 individuals: upon incorporating CNV size, allele frequency, and inheritance profile as covariates, we estimated a 94% sensitivity and 95% specificity. Finally, we developed two approaches for CNV visualization from WGS data, the first output directly by our genotyping algorithm and the second, CNView, which is a stand-alone utility for CNV scoring, annotation, and normalized depth visualization across populations. Collectively, these tools provide an improved framework for distilling the massive number of candidate WGS CNV sites called by many detection algorithms to a reasonable number for automated visualization and molecular validation, and represent possible valuable additions to WGS analysis pipelines deployed for basic research and clinical applications.

1699W

The PhenX toolkit: Standard measures and bioinformatics for collaborative research. *T. Hendershot[1], H. Pan[1], W. Huggins[1], M. Phillips[1], S. Hwang[1], Y. Qin[1], V. Bakalov[1], J. Levy[1], D. Nettles[1], D. Maiese[1], R. Boyles[1], S.M. Dolan[2], L. Dixon[3], G.E. Swan[4], J.J. Prochaska[4], C.A. McCarty[5], M.L. Marazita[6], E.M. Ramos[7], C.M. Hamilton[1].* 1) RTI International, Research Triangle Park, NC; 2) Albert Einstein College of Medicine / Montefiore Medical Center, Bronx, NY; 3) Columbia University Medical Center, New York, NY; 4) Stanford University, Stanford, CA; 5) Essentia Institute of Rural Health, Duluth, MN; 6) University of Pittsburgh, Pittsburgh, PA; 7) National Human Genome Research Institute, Bethesda, MD.

The Web-based PhenX Toolkit, https://www.phenxtoolkit.org/, is a catalog of standard measures to facilitate collaborative biomedical research. Use of PhenX (consensus measures for *Phen*otypes and e*X*posures) measures helps ensure that phenotype and exposure data is collected using consistent methodology. Consistent data collection improves data quality and can enable data comparability across multiple sites in large cohorts (e.g., the Precision Medicine Initiative), facilitate combining data to validate clinically actionable variants, increase statistical power (e.g., studies of rare genetic conditions or gene-environment interactions), or compare treatment and outcomes among patients. We will present new tools that link PhenX to major informatics resources, including the database of Genotypes and Phenotypes (dbGaP) and Research Electronic Data Capture (REDCap). To help users find comparable data, a team of curators is mapping the ~22,000 PhenX variables to all 690 studies in dbGaP. We will present the process, results, and implications of this mapping effort. We will also present PhenX compatibility with the REDCap study design tool. REDCap modules are available for all protocols in the PhenX Toolkit providing an easy way to add PhenX data collection protocols directly to REDCap projects. REDCap is being used by more than 257,000 projects across numerous research areas, and is available for use on mobile devices. The PhenX Toolkit currently includes 475 measures from 23 research domains, with additional depth for Substance Abuse and Addiction, Mental Health, Tobacco Regulatory, and Sickle Cell Disease research. We will present new measures for Pregnancy and Early Psychosis (clinical and translational), as well as Core measures for Tobacco Regulatory Research studies. For each measure, the Toolkit provides detailed instructions (e.g., required training and equipment), features (e.g., Spanish language versions), and bioinformatics support (e.g., data dictionaries, REDCap compatible modules) to facilitate implementation and data analysis. Through these efforts, the PhenX Toolkit is evolving to increase content, utility and help establish connections among major informatics resources. *Funding provided by a Genomic Resource Grant (U41 HG007050) from NHGRI, with co-funding from NIDA. Supplemental funding is provided by NIMH, NHLBI, and the NIH ODP TRSP.*.

1700T

Substantial contribution of non-coding regulatory mutations to autism and identification of risk genes from whole-genome sequencing data of affected families. *Y. Liu[1], J. Li[2], E.A. Cicek[3,4], N. Knoblauch[1], Y. Jiang[2], R. Muhle[7], Y. Wang[5], E. Geller[8], J. Wu[2,6], J. Noonan[7], Z. Sun[2], X. He[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China; 3) Lane Center for Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA; 4) Computer Engineering Department, Bilkent University, Ankara, Turkey; 5) The institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; 6) Institute of Genomic Medicine, Wenzhou University, Wenzhou, China; 7) Department of Genetics, Yale School of Medicine, New Haven, Connecticut; 8) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Analyzing de novo mutations in protein coding genes from whole-exome sequencing (WES) data has emerged as a powerful tool for mapping risk genes of autism spectrum disorder (ASD). The impact of non-coding mutations in ASD, however, has been largely unknown. This represents a large gap in our understanding of genetics of ASD, as the majority of GWAS hits in a range of disorders fall into non-coding regions. To address this question, we performed meta-analysis of de novo mutations from whole-genome sequencing (WGS) data, consisting of more than 300 affected families. We found that DNMs are enriched within brain regulatory enhancers near putative autism genes and evolutionarily constrained genes. We estimate that de novo non-coding mutations make a larger contribution to the risk of autism than coding mutations, explaining about 50-70% of ASD genetic risk attributable to de novo mutations. By combining information of non-coding DNMs with published WES data, we identified five new ASD risk genes at a false discovery rate (FDR) < 0.1, and 12 at FDR < 0.2. We also show by simulation that WGS outperforms WES in power at comparable sequencing cost. Take together, our results demonstrate the pathogenic contribution of non-coding DNMs in ASD etiology and highlight some promising ASD risk genes. The analytic tools we provided in this study, for estimating contribution of de novo mutations to disease risk, for mapping susceptibility genes, and for power analysis, are applicable to any WGS studies on de novo mutations.

1701F

Improved detection and analysis of retrotransposon insertion variation in a family quartet including monozygotic twins discordant for schizophrenia. *M.E.O. Locke[1], C.A. Castellani[3], C.W. Gendron[2], S.M. Singh[2], M.J. Daley[1,2].* 1) Department of Computer Science, The University of Western Ontario, London, Canada; 2) Department of Biology, The University of Western Ontario, London, Canada; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Statement of Purpose Retrotransposons are a class of mobile genetic elements whose activity in germline and somatic cells has been linked to numerous phenotypes in humans. Studies have indicated that mobile element insertions(MEIs) of retrotransposon LINE1 play a role in central nervous system development and - along with another class of retrotransposon, human endogenous retroviruses (HERVs) - have been implicated in psychiatric disorders including schizophrenia. Existing next-generation-sequencing approaches to find MEI sites first align reads to the reference genome then leverage information from paired reads to find those that have one "anchor" read aligning to the reference genome and the other read aligning to a mobile element of interest (MEOI). These methods are limited by the requirement of aligning to the reference sequence first, which may miss reads that are not near-exact matches to the human reference genome used. We aim to improve the sensitivity of MEOI context detection to better assess both the context and quantity of MEOIs in a given genome dataset. Methods Used We have developed a novel approach to find sequence contexts flanking an MEOI by flipping the anchor read from the reference genome to the MEOI itself. This is accomplished by filtering all raw reads into those matching MEOIs and then performing reference-free de novo assembly of any read extending beyond the end of an MEOI. This method is highly sensitive, as it can find flanking regions not in the reference genome sequence. We have also developed a method to find the relative abundance of MEOIs, which uses read depth tempered by k-mer uniqueness. We used these methods to compare whole genome sequencing data generated by Complete Genomics of two sets of MZ twins discordant for schizophrenia and the parents of one set. This approach allowed us to measure differences in the quantity and context of LINE1s and HERVs. Results The software developed is applicable to any dataset from Complete Genomics including raw reads, and is extensible to any DNA sequence of interest larger than approximately 500bp. The software is written in C++ using the SeqAn library, and is available as open source. We have found that not all LINE and HERV locations are explained by Mendelian inheritance, and that there are significant differences in LINE1 and HERV sites between MZ twins that may explain their discordance for neurodevelopmental disorders including schizophrenia.

1702W

Integrative analysis of cell-type specific genomic annotations recovers complex physiological relationships between tissues. *R. Powles[1], Q. Lu[2], Y. Lu[3], B. Li[2], Q. Wang[1], J. He[4], H. Zhao[1,2,5].* 1) Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT; 2) Department of Biostatistics, Yale School of Public Health, New Haven, CT; 3) Department of Immunobiology, Yale School of Medicine, New Haven, CT; 4) Division of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT; 5) VA Cooperative Studies Program Coordinating Center, West Haven, CT.

Functional annotation of the human genome has been a constant pursuit of genomics consortia with an additional emphasis on tissue- and cell type-specificity in recent years. While current methods can offer some localization of functionality to broader systems (e.g., nervous or circulatory), many human diseases are malignancies of only a few tissues. To improve upon the resolution and specificity of current functional annotation methods, we use an unsupervised learning framework to consider an expansive collection of high-throughput annotations from Roadmap Epigenomics, including DNA methylation and chromatin modification integrated from 7 histone marks. Instead of focusing on a handful of broad systems, our model considers over 100 different cell and tissue types separately. We identify a diverse collection of cell-type specific non-coding regulatory elements including long non-coding RNA (lncRNAs), micro-RNAs (miRNAs), and enhancer regions for a variety of cell types. Additionally, these integrative annotations are capable of capturing relationships between cells of similar location, morphology, and/or functional role down to single cell and tissue types. We show that we can recapitulate functional relationships between cell types using genomic annotation in a variety of case studies. Our model's predictions of genomic functionality identify miRNAs with high expression specificity for skeletal muscle that are also found to have increased functionality in the right and left ventricles of the heart, but not in the thinner, lower resistance walls of the right atria. Similarly, lncRNAs with high expression specificity for intestinal tissue are shown to share functionality regions not only with other intestinal layers (e.g., colonic mucosa, sigmoid colon), but also tissues with similar exocrine and endocrine secretory roles like the liver and pancreas. We also identify different subsets of T-cells based on functional predictions in known T-cell lineage-specific non-coding regions, such as the interleukin-17 (*IL17*) locus. Finally, we evaluate the ability of our integrated annotations to better prioritize genome-wide association study (GWAS) variants, and compare our method with other popular variant annotation metrics in complex diseases such as schizophrenia. Our findings suggest that our cell-type specific annotation is essential to understanding the intricate relationships between different cell populations and their functional genomes.

1703T

CNV analysis identifies novel duplication at the *SLIT3* locus in FTLD. *A. Desai, M. March, C. Kim, T. Kelly, H. Hakonarson, P. Sleiman.* Center for Applied Genomics, The Children's Hospital of Philadelphia, PA Philadelphia, PA 19104.

Frontotemporal lobar degeneration (FTLD) is the second most common cause of dementia in individuals under age 65. FTLD manifests clinically with progressive behavioral and/or language deficits and has a prevalence of approximately 3.5 to 15 cases per 100,000 in 45 to 64 year olds. At autopsy, atrophy is evident in the frontal and temporal poles. In addition to neuronal loss and gliosis ubiquitin inclusion bodies are present. Approximately 50% of cases manifest as FTLD-TDP with neuronal and glial TDP-43 inclusions. To identify novel variants contributing to FTD we assessed genome wide copy number variation (CNV) between unrelated 567 FTLD-TDP cases compared to 3380 controls. These genotyping studies were performed using Illumina human 610-quad V1 beadchip. The CNV analysis was performed using PennCNV software based on the XHMM model. Using stringent quality control steps we filtered out potentially false positive CNVs based on the underlying quality score of sample, the count of CNV<100 for each sample. CNVs that survived the QC steps were assessed for association with FTLD-TDP using the ParseCNV package. ParseCNV generates probe-based statistics for a given CNV in both cases and controls which can be defined as a CNV region (CNVRs). We identified a statistically significant duplication at a locus on 5q34 that spans the *SLIT3* gene (P=1.7x10^{-8}, OR=56.95). The *SLIT3* duplication CNV was only observed in FTD cases and was absent from controls. Presence of the CNV has been validated by qPCR in all 9 unrelated FTD cases. *SLIT3* is a homolog of the Drosophila SLIT gene which is critical to the formation of the midline of the central nervous system. In mammals the three Slit proteins (Slit 1-3) have been shown to mediate repulsive signaling during axon guidance and neuronal migration. In addition to known neuronal role of SLIT3, a motor neuron expressed micoRNA, miR218, and it's enhancer have also been mapped to the locus, encoded in the *SLIT3* introns. SLIT3 duplications have been associated with major depressive disorder (MDD) and *de novo* mutations reported in autism. This study represents a comprehensive CNV analysis using the Illumina 610K platform. We report the identification of a novel variant contributing to FTLD susceptibility. These results are significant not only in shedding light on the contribution of CNVs to the pathophysiology of FTLD, but they also increases the range of neurological disorders associated with *SLIT3* mutations.

1704F

Predicting the cellular impact of genotype variant to prioritize disease hits. *F. Farhadi Hassan Kiadeh[1,2,3], J. Xu[4], C. White[4], D.A. Bennett[5], P.L. De Jager[4], S. Mostafavi[2,3,6,7].* 1) Department of Bioinformatics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 3) Child & Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 4) Department of Neurology, Harvard Medical School, Boston, MA, USA; 5) RUSH Alzheimer's Disease Center, RUSH University Medical Center, Chicago, IL, USA; 6) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 7) Department of Statistics, University of British Columbia, Vancouver, BC, Canada.

Over the last decade, genome-wide association studies (GWAS) have successfully identified many disease-associated variants for several common and complex diseases. However, identifying associated variants requires huge sample sizes, because of the small effect sizes of individual variants and the typical polygenicity of common disease. In order to increase the statistical power of GWAS, several recent approaches have been devised for limiting the 'search space' of GWAS to those SNPs that are predicted to have a regulatory impact on one or multiple cellular phenotypes [1,2]. In particular, previous approaches that "impute" gene expression from genotype data present a very attractive framework for aggregating genetic variants in a gene-specific manner. Such approaches not only result in increased statistical power, but also can identify potential mechanisms through which regulatory variants impact cellular processes. In this study, we develop a "gene-based" variant aggregation method that uses expression quantitative trait loci (eQTLs) data, in order to improve prioritize SNPs for testing in GWAS. Because of the tissue-specificity of cellular phenotypes, the choice of "tissue" is critical in building such a model. Therefore, we built such a model using eQTL data for two major human tissues, the DGN study for blood (n=922) and the ROSMAP study for cortex (n=500), and compare these models on prioritization of disease variants in several GWAS studies. Finally, we propose a comprehensive "gene-based" model that effectively aggregates genetic variants for joint expression imputation from blood and brain.You may contact the first author (during and after the meeting) at ffarhadi@cmmt.ubc.ca 1. Gamazon, Eric R., et al. "A gene-based association method for mapping traits using reference transcriptome data." *Nature genetics* 47.9 (2015): 1091-1098. 2. Barbeira, Alvaro, et al. "MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results." *bioRxiv* (2016): 045260..

1705W

A systematic evaluation of gene expression regulation changes in schizophrenic brains. *S. Jiang[1], C. Chao[2], E.S. Gershon[3], Y. Wang[4], G. Yu[4], C. Liu[1,2].* 1) Department of Psychiatry, the University of Illinois at Chicago, Chicago, IL, USA; 2) The State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; 3) Department of Psychiatry and Behavioral Neuroscience, the University of Chicago, Chicago, IL, USA; 4) Department of Electrical & Computer Engineering, Virginia Tech, Arlington, VA, USA.

When comparing cases to controls, change in expression of a gene could be manifested as alternations of absolute expression level and/or of relative ratios to its functional partners. The two alternations are related but also distinct. Both could contribute to disease progression. They represent property changes of one gene as an individual entity and as a community member. Therefore, considering only one can lead to missed opportunities. The former change has been the major focus in the past decades of studies. However, the latter one has obtained limited attention until recent. In this study, we assessed both differential expression and differential connectivity (as correlation levels with the other genes) in postmortem brains contrasting schizophrenics to controls. Also, we implemented a new algorithm, Differential COnnectivity and Differential Expression (DiCODE), which combined differential expression and connectivity weights, to evaluate the combined effects of expression and connectivity changes for each gene. DiCODE, together with differential expression and differential connectivity analyses, constitute our new comprehensive analysis method called TRIple Differentiation (TRID). Therefore, each gene was given three different weights so that ranked gene lists can be used for further analyses. We used two independent schizophrenia (SCZ) case-control gene expression datasets for prefrontal cortex (PFC) and one dataset for parietal cortex (PC) in the present study. Pair-wise correlation coefficients among gene lists of the three datasets by TRID are all close to zero. When gene set enrichment analysis was applied to the weighted gene lists, highly replicable schizophrenia-associated pathways were identified across the three datasets by TRID. Integrating the weighted gene lists with known human PPI networks, several schizophrenia-associated modules shared by the three datasets were identified by TRID. In these PPI modules, some genes including *PTN*, *HSPA8* and *SRPK2* were mapped to the genome-wide association signals of schizophrenia for all datasets by TRID. We further compared TRID with weighted gene co-expression network analysis (WGCNA), and learned that TRID is more sensitive to changes of individual genes that could be missed by WGCNA. Our data suggests the changes in schizophrenia are highly heterogeneous at individual gene level. TRID captured different genes of the same signaling pathways or regulatory modules impaired in different patients.

1706T

NetDx: An algorithm to classify patients based on similarity networks. *S. Pai, S. Hui, R. Isserlin, H. Kaka, G.D. Bader.* The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada.

Patient classification has widespread biomedical applications in clinical management, including tumour subtyping in cancer and the pharmacogenetic prediction of drug response. A general purpose, clinically useful prediction algorithm should be accurate, generalizable, be able to integrate diverse data types (e.g. clinical, genomic, metabolomic, imaging), handle sparse data and be intuitive to interpret. We describe netDx, a supervised patient classification framework based on patient similarity networks, that meets the above criteria. netDx models data as patient networks and uses the GeneMANIA machine learning algorithm[1] for network integration and feature selection. The results of the classifier are clinically and biologically interpretable, and can be visualized as an integrated patient similarity network. We will present two applications of netDx. In the first, we use a published gene expression dataset of medulloblastoma subtypes (N=103 tumours) and build a 4-class predictor for tumour subtype[2]. In the second, we predict case/control status in autism spectrum disorders based on the occurrence of rare copy number deletions in cellular pathways (N=3,291 patients)[3]; this predictor achieved better performance than previously published methods. Near-term extensions involve networks based on genotype similarity from dense SNP genotyping. Software will be made available as open-source R packages. For project updates, visit http://netdx.orgReferences: 1. Mostafavi S et al. (2008). *Genome Biol.* **9**:S4 2. Northcott PA et al (2011) *J Clin Oncol.* **29** (11):1408. 3. Pinto et al. (2014). *Am J Hum Gen.* **94** (5):677.

1707F

Deciphering the importance of non-coding elements deleted in patients with 2p15p16.1 microdeletion syndrome. *H. Bagheri[1,2], C. Badduke[1,2], Y. Qiao[1,2], S. Martell[1,2], S. Mostafavi[2], S. Lewis[2,3], C. Gregory-Evans[4], E. Rajcan-Separovic[1,2].* 1) Dept Pathology, UBC, Vancouver, BC, Canada; 2) Child and Family Research Institute, Vancouver, Canada; 3) Dept. Medical Genetics, UBC, Vancouver, BC, Canada; 4) Dept. of Ophthalmology, UBC, Vancouver, BC, Canada.

Introduction: The 2p15p16.1 microdeletion syndrome (OMIM 612513) is first described in two phenotypically similar individuals with intellectual disability (ID). Recently, our clinical, genomic and functional analysis (Bagheri et al., JCI Insight, March 2016), using patient cells and zebrafish knockdown, identified 3 driver genes (*XPO1*, *USP34*, and *BCL11A*) for the syndrome . However, out of 33 patients reported in the literature so far, deletions in 2 patients (from 56,853,162-60,380,981 and 57,606,726-59,619,316) did not contain any of the above 3 genes. This study examines whether non-coding elements in these two regions of 2p15p16.1 may regulate the 3 driver genes. **Materials & Methods**: Our approach comprised the *in silico* analysis of the enhancer elements involved in the two deletions without the 3 candidate genes, which were extracted using UCSC and VISTA enhancer browsers. Using several webtools (e.g. braineac.org) we identified expression quantitative trait loci (eQTLs), by searching for SNPs mapping to the two 2p15p16.1 deletions that alter the expression of the 3 driver genes. **Results**: We found that chromosome 2 contains the highest number of enhancers in the genome (104 of 897 VISTA-positive and 68 of 617 brain/neuronal-expressed enhancers) of which approximately one-third are present in the 2p15p16.1 region. The enhancers did not contain eQTL SNPs that regulate expression of the 3 candidate gene, or any genes from the 2p15p16.1 region. However, there were 24 eQTL SNPs in one of the deletions that regulate the expression of one of the driver genes (*BCL11A*). The second patient's deletion did not include these SNPs, but occurred together with 2 additional small non-coding deletions in the second intron of *BCL11A* and ~40Kb upstream of *BCL11A*. The intronic deletion is in the proximity to two SNPs (rs1427407 and rs75994488) that have been previously associated with schizophrenia, autism and fetal hemoglobin expression. **Conclusions**: We described an *in silico* approach to decipher the effects of non-coding elements on the expression of 2p15p16.1 genes. The regulatory elements in both deletions that do not contain exons of the 3 critical genes could be associated with *BCL11A* function. This strengthens the role of *BCL11A*, a chromatin remodeler, in the 2p15p16.1 microdeletion syndrome and ID, as previously suggested, based on the recurrence of mutations in this gene in patients with ID.

1708W

A review of clinical NGS accuracy for mitochondrial disorders: A comparison of NGS to secondary Sanger sequencing confirmation. *S. Dames[1,3], B. Brulotte[1], G. Pont-Kingdon[1], K. Eilbeck[3], R. Mao[1,2].* 1) ARUP Laboratories Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) University of Utah Department of Pathology, Salt Lake City, UT; 3) University of Utah Department of Biomedical Informatics, Salt Lake City, UT.

Introduction: The mitochondrial disorder (mt disorder) assay was the first clinical next generation sequencing (NGS) assay offered at ARUP. Initial validation quality control metrics required a secondary method fpr confirmation of reportable variants by Sanger sequencing. Subsequently, improvements in bioinformatics, capture technologies, and NGS sequencing platforms have been validated; however a secondary detection by Sanger is still utilized. In order to reduce turnaround time and assay costs, a comparison of concordance between Sanger and NGS for the mt disorders assay was performed to determine the overall accuracy of NGS compared to Sanger in a clinical setting. **Methods:** The mt disorder assay enriches the mitochondrial genome by long-range PCR and 108 nuclear genes associated with mt disorders by either RainDance (initial validation) or NimbleGen SeqCap technology (current methodology). All medical director Sanger variant requests that met minimum QC metrics based off of NGS data were compared to the final Sanger result (i.e. nucleotide call, zygosity, and plasmy). **Results:** Over 400 mt disorder assays have been performed at ARUP Laboratories resulting in 608 mtDNA and 421 nuclear Sanger requests. The overall concordance of NGS to Sanger sequencing for the mtDNA is 100% when heteroplasmy levels are greater than 25%. Nineteen of 608 mtDNA variants were not detected by Sanger sequencing (96.9% concordance). Of these 19 variants, all had heteroplasmy levels less than 25%. Two of these variants were further examined by allele-specific PCR and confirmed. Five of 421 nuclear gene variants were not confirmed by Sanger sequencing, yielding a concordance of 98.8%. Missed calls associated with the nuclear genes were found with variants that used probabilistic zygosity algorithms which were on the lower end of zygosity QC metrics. These false positive heterozygous variants were all homozygous. No homozygous variants were miscalled, and all heterozygous variants between calculated 25-75% alternate allele frequency were unconfirmed. **Conclusions:** Based on concordance data between NGS and Sanger sequencing, ARUP is currently considering different options bioinformatically and experimentally to reduce or eliminate the requirements of Sanger verification for all NGS assays. Examination of dual indexed libraries and/or optimization of QC metrics to provide a probabilistic metric for NGS variant calls are currently underway and will be presented.

1709T

Mitochondrial DNA clinical diagnostics through next generation sequencing. *R. Berutti[1], T. Wieland[1], T. Schwarzmayr[1], B. Lorenz-Depiereux[1], U. Ahting[1,2], M. Radivojkov-Blagojevic[1], P. Lichtner[1], T. Haack[1,2], T. Meitinger[1,2], H. Prokisch[1,2], T.M. Strom[1,2].* 1) Institute of Human Genetics , Helmholtz Zentrum München, Neuherberg, Bayern, Germany; 2) Institute of Human Genetics , Technische Universität München , Munich, Bayern, Germany.

The mitochondrial genome is an important target for clinical diagnostics. In the last one and a half year we analyzed the mtDNA of more than 8,000 samples referred for exome sequencing, 226 for targeted mtDNA analysis and 45 whole genomes.Sequencing was performed on the Illumina HiSeq and MiSeq platforms, reads were aligned using BWA-mem to a modified hg19 reference sequence, where the mtDNA sequence has been replaced with the revised Cambridge Reference Sequence (GenBank NC_012920). Contamination from NUMTs, nuclear DNA regions featuring high homology with mtDNA, is negligible. A full nuclear reference allows the discrimination of ambiguous reads. Moreover, on genomes and targeted sequencing (single amplicon of the mtDNA) the ratio of the mtDNA depth to the nuclear depth is exceptionally high. On exomes, enrichment kits not encompassing NUMT regions were used, minimising contamination.Variant calling was performed with GATK, for SNPs and short indels, and with Pindel for large deletions. Annotation uses gene definitions from ENCODE, and an extra tRNAs track. Interpretation of discovered mutations was supported by additional information from MITOMAP, like disease association and variant frequency.Detected mutations are stored in our in house variation database, which is accessible via a web front-end, providing annotation, quality, functional and frequency information.Overall, we obtained an average depth of 3800X for genomes, 45X for exomes, and 8500X for targeted sequencing. We identified on average 27 mutations per sample in the range of heteroplasmy over 15%. Within our cohort of 315 exome patients with suspect mitochondrial disorders, we found 25 disease causing mutations on the mtDNA from the exome and 41 from the targeted sequencing. The mtDNA of 14 samples was genotyped both with exome and targeted sequencing. The achieved concordance was of 98.9% on variant calls, with a sensibility of 92.3% for exomes. Heteroplasmy detection was performed down to values of ~20% for exome and ~10% for targeted DNA samples.Since most of our samples are of German origin, we additionally performed haplogroup analysis of our cohort, showing a haplogroup composition in agreement with a European population. We demonstrated the possibility of using exome sequencing data to reliably identify mitochondrial variants. Furthermore we were able to implement mtDNA sequencing for diagnostics to help solving unclear clinical cases of mitochondriopaties.

1710F

Next-generation sequencing in three Israeli families with Keratoconus. *W.A. Goar[1,2,3], P. Majdalani[5], C.C. Searby[2,3], S.S. Whitmore[1,3], A.P. DeLuca[1,3], A. Imtirat[4], E.M. Stone[1,3], R. Pavari[5], T.E. Scheetz[1,3], V.C. Sheffield[2,3].* 1) Opthalmology and Visual Sciences, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) Stephen A. Wynn Institute for Vision Research, University of Iowa, Iowa City, IA; 4) Department of Medicine, Ben-Gurion University in The Negev, Beersheba, Israel; 5) Microbiology, Immunology and Genetics, Ben-Gurion University in The Negev, Beersheba, Israel.

Keratoconus (KT) is the most common corneal dystrophy with an occurrence rate of 1 in every 2,000 people. KT causes the cornea to thin with age, becoming conical and leading to myopia, irregular astigmatism, and corneal scarring. The age of onset is generally the teenage years with stabilization in the third and fourth decades of life. Incidence of KT does not seem to be more prevalent in a particular gender, but does seem to occur more frequently and aggressively in Israel and Saudi Arabia. Currently, corneal transplantation is the only treatment for KT when visual acuity is no longer correctable by contact lenses. We hypothesize that KT is a genetically heterogeneous disease that is caused by mutations in one of several genes. Samples were obtained from 3 Israeli KT families (16 samples) and genotyped using an Affymetrix Genome-Wide SNP 5.0 microarray. The SNP data was analyzed for regions of autozygosity using PLINK to investigate the presence of homozygous regions of the genome shared between the affected family members. Three samples with KT (one for each family) were chosen for exome sequencing. The resulting variants were filtered based upon variant quality, predicted function, and population prevalence (1000 genomes, ExAC). Spurious variants (based upon our local set of 1000+ exomes) were removed to create a final variant list for each family, which was annotated with corneal expression (http://genome.uiowa.edu/otdb) to assist in prioritizing potential candidates. Using the autozygous regions, we have prioritized candidates and areas of the genome on which to focus our investigation. No plausible variations were found in these three families in genes previously reported to cause KT. In addition, no single gene with plausible disease-causing variations was shared across all three families. However, a few genes were found that are shared between two of the families and need further follow up. We also identified at least 3 areas of autozygosity that are shared between the 3 families. These areas need further examination to determine if they are harboring causative variants. Further work is needed to identify the causative mutations in these families. Due to the isolated population we will continue to pursue these through ascertainment of additional families and family members. This will allow us to further narrow the intervals of the genome in search of the causative mutations.

1711W

Prioritizing biology based on single-cell transcriptomics. *P. Timshel[1,2], T.H. Pers[1,2].* 1) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 1, København Ø 2100, Denmark; 2) Department of Epidemiology Research, Statens Serum Institut, 2300 Copenhagen, Denmark.

Background: A key challenge for gaining biological insights from genetic associations is to identify which genes and pathways explain the associations. We have previously shown that reconstituted gene sets augment biological interpretation of genome-wide association studies (GWAS; Pers et al., 2015). 'Reconstituted gene sets' are based on multi-tissue expression data and comprise probabilities of each gene's membership for a given pathway. However, a remaining major limitation is that these gene sets are not tissue or cell type specific. **Methods**: We introduce DEPICT for single-cell (DEPICT-sc), which uses single-cell gene expression data to more accurately highlight enriched pathways. We learn robust 'single-cell transcriptional components' (scTC) from single-cell transcriptomic data by applying matrix factorization models on data from >13,000 mouse retina cells (Macosko et al., 2015). Then we use the scTCs and >14,000 predefined gene sets to construct reconstituted gene sets. **Results**: We show that scTCs reflect genuine biological differences between retina cells. We show that the top scTCs enrich for relevant retina biology, such as 'detection of light stimulus' ($P < 2.23 \times 10^{-108}$) and 'phototransduction' ($P < 3.08 \times 10^{-66}$). To illustrate the relevance of DEPICT-sc for gaining biological insights from genetic associations, we integrate the reconstituted gene sets with GWAS data for age-related macular degeneration (AMD) and successfully prioritize 'response to reactive oxygen' for age related macular degeneration ($P < 1.25 \times 10^{-10}$). **Conclusion**: Our results suggest that single-cell gene expression data leads to more specific prioritization of likely etiologic gene sets and pathways. DEPICT-sc can easily be adapted to prioritize genes, pathways and cell types for other traits and diseases.

1712T

At-a-glance" human disease model and phenotype summary data at mouse genome informatics. *C. Smith, J. Eppig, The Mouse Genome Informatics Team.* Mouse Genome Informatics, Jackson Laboratory, Bar Harbor, ME.

The mouse is widely used as a model system to investigate the genetics of human disease. Coupling human sequence, phenotypes, pathogenicity variant calls, copy number variation, and other biological evidence with mouse model data can accelerate the identification and testing of therapeutic candidates. Mouse Genome Informatics (MGI, www.informatics.jax.org) provides data sets to correlate mouse phenotype with human diseases, clinical signs and symptoms. MGI catalogs all mouse mutant alleles, including published mouse mutations and mutations produced by large-scale mutagenesis programs. Point mutations (made by ENU and CRISPR) or knock-out (null) mutations now exist for most mouse genes. In MGI, mouse genotypes are annotated to phenotype and OMIM disease model descriptions, and include links to supporting gene information such as sequence, spatiotemporal expression, genomic location, biochemical function and process, sub-cellular topology, and mammalian gene homology. Thus, the model data at MGI aids in understanding biochemical pathways and pathological processes that can further advance the knowledge of underlying mechanisms of human genetic disease. Mouse Genome Informatics has recently revised the web-based gene-level presentation of these phenotype and disease model data. In addition to the number of alleles and variants for a gene, the number of phenotype annotations from single and multigenic genotypes, images and references, MGI now provides 'at-a-glance' graphical displays into systems affected by variants in genes, with quick links to more detailed information. A section on Human Diseases gives an overview of mouse models in orthologs of human genes. In this presentation, we will show how to access the fully integrated data sets at MGI starting with summary views, and provide examples of therapeutic targets identified using data generated from mouse models.

1713F

Non-cooperative facial recognition technology enables scalable accurate facial gestalt matching of individuals with the same syndromic form of intellectual disability. *T.E. Dudding-Byth[1], A. Baxter[1], E.G. Holliday[3,4], A. Hackett[1,4], S. O'Donnell[1], S.M. White[5], J. Attia[3,4], C. Nellaker[6,7,8], H. Brunner[9], B. de Vries[9], D. Koolen[9], T. Kleefstra[9], S. Ratwatte[10], S. Brain[11], B.C. Lovell[12].* 1) Hunter Genetics, Hunter New England Health Service, Newcastle, NSW, Australia; 2) GrowUpWell, Priority of Research Excellence, The University of Newcastle, Newcastle, NSW, Australia; 3) Hunter Medical Research Institute, Newcastle, NSW, Australia; 4) The University of Newcastle, Newcastle, NSW, Australia; 5) Murdoch Childrens Research Institute, Melbourne, Australia; 6) Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Radcliffe Hospital Oxford, United Kingdom; 7) Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom; 8) Big Data Institute, University of Oxford, Oxford, United Kingdom; 9) Department of Human Genetics and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; 10) The Department of Medicine, John Hunter Hospital, Newcastle, NSW, Australia; 11) Imagus Technology, Brisbane, Queensland, Australia; 12) School of ITEE, The University of Queensland, Brisbane, Queensland, Australia.

Massively parallel genetic sequencing allows rapid testing of known intellectual disability (ID) genes. However, the discovery of a novel ID gene often requires molecular confirmation in a second patient with an overlapping phenotype or similar facial gestalt. This study examined the accuracy of non-cooperative facial recognition technology for matching the facial images of non-identical individuals, each of whom was either affected with one of ten genetic syndromes (n~1145) or was unaffected (normal control, n~2000) (total 3145 images). The database was populated with images from published peer review manuscripts and publically available patient groups. We tested the accuracy of the facial recognition technology using the "leave one out" method, i.e., an individual image is removed from the database to be used as a test case for generating their top ten closest matches. For the analysis, we specified two research questions: 1) Using all photographs in the database, did the software correctly identify i) a top match, ii) at least one in the top five, or iii) at least one in the top ten with an individual from the same syndrome subgroup, more than would be expected by chance; 2) The relative accuracy of software-based matches and whether three experienced clinicians would have also considered the syndrome diagnosis based on the photograph alone. For the first question, expected frequencies were estimated via simulation, and frequencies compared using Chi-Square statistics. For the second question, Kappa statistics and McNemar's tests were used to assess agreement and compare the relative accuracy of software and clinician diagnoses. Statistical analysis performed to date on the first five of the ten syndromes shows that the software correctly identifies a top match, at least one correct match in the top five and at least one in the top ten far more often than expected by chance (for top match, one in five and one in ten, Chi-square statistics all > 800, >350 and >150, respectively; all P<0.00001). For the second question, we observed low agreement between the software and clinicians with higher accuracy of the software when results were discordant (most McNemar's Chi-square statistics >10, with P>0.01). The scalability of this technology on an international basis has the potential to enhance the efficacy of deep phenotyping platforms for identifying the second individual or a cluster of individuals with a similar facial gestalt. .

1714W

SolveBio FUSE: A very large genetic variant dataset generated with MapReduce. *J. Hull, P. George, D. Caplan, X. Xu.* SolveBio, New York, NY.

   Purpose: Genomic variant annotations can be found in a large number of databases containing a plethora of information ranging from clinical evidence & allele frequencies, to information curated from individual literature articles. In order to provide real-time, context-specific variant intelligence, we devised a method to generate specialized meta-datasets (FUSE datasets) containing relevant annotations from any number of sources. In addition, the meta-dataset includes data-provenance allowing it to act as a single source of truth within an application. The resulting FUSE datasets, indexed by genomic location, provide a solid foundation for novel genomic applications to be rapidly built & deployed. **Methodology**: The complete SolveBio FUSE dataset contains 385 million genomic variants comprised of multiple versions of 15 public genomic datasets. The dataset is automatically compiled using a proprietary Python application & Apache Hive which leverages MapReduce technologies. Using the SolveBio contextual knowledge hub, our solution takes an input of datasets & desired fields & dynamically generates schemas & a pipeline of Apache Hive queries to generate the final dataset. Our schema generators consolidate non-normalized fields across any number of datasets, & our Hive queries de-duplicate values from each while maintaining data provenance information. The final dataset is compatible with both SQL & NoSQL datastores. Given any genomic coordinate, querying the FUSE dataset returns a complete picture of a region or variant including all associated annotations. **Results**: We deployed our FUSE dataset to allow for intelligent search within SolveBio in Fall 2015. Since then, our dataset has been queried millions of times & has enabled SolveBio to build a contextual knowledge hub that leverages real-time search across a large number of datasets, a beacon system that conforms to the GA4GH guidelines, & a genomic entity recognition algorithm. Because of the ease in configuration & generation of the dataset & the parallel & scalable nature of MapReduce algorithms, we have been able to make rapid improvements & add new data to our dataset since its inception. The entire dataset can be generated in less than an hour. **Conclusion**: A single dataset containing a consolidated & versioned view of every genomic data point about variants creates a strong foundation for accurate, intelligent & novel applications for genomics in enterprise, research & clinical environments.

1715T

Variant-level information is critical for inferring the functional impact of spontaneous human mutations. *A. Saha Mandal[1,2,3], A. Mathankeri[1,2,3], A.P.J. De Koning[1,3,4,5].* 1) Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada; 2) Bioinformatics Graduate Program, University of Calgary, AB, Canada; 3) Alberta Children's Hospital Research Institute, Calgary, AB, Canada; 4) Cumming School of Medicine, University of Calgary, AB, Canada; 5) Department of Medical Genetics, University of Calgary, AB, Canada.

   Computational approaches for predicting the pathogenicity of genomic variants have been recently advanced by the development of novel training strategies (*e.g.,* CADD and EIGEN). Here we evaluate some of the pitfalls of these strategies and explore how they can be overcome. Despite being very successful at distinguishing highly penetrant ClinVar variants from common polymorphisms, we show that these data integration classifiers perform no better than PolyPhen2 at distinguishing benign from pathogenic variations *at the same genomic position*. This problem is far more difficult than simply identifying mutations at functionally important positions, and is arguably the central challenge in prioritizing rare and spontaneous mutations. Here we show that the integration of additional variant-level data substantially improves sensitivity and specificity over existing approaches. We present a new classifier, TAIGA, Transformation and Integration of Genomic Annotations, which implements a rotation forests algorithm for missense variant classification and can outperform CADD and EIGEN by a substantial margin. A lack of reliable training and/or testing data continues to be a limiting problem, however, we argue that an even greater impediment to further progress is the lack of sufficiently detailed predictor information at the variant level. It is thus an urgent priority to develop new variant-level scoring systems that can make fuller use of large-scale variation data and enable quantitative assessment of the functional deficit caused by human mutations.

1716F

Lynx – An integrated bioinformatics platform to support high-throughput translational biomedicine. *D. Sulakhe[1,2], B. Xie[1,4], S. Wang[1,3], M. DSouza[1], A. Taylor[1], S. Hashemifar[1,3], G. Agam[4], T.C. Gilliam[1,2], N. Maltsev[1,2].* 1) Human Genetics, The University of Chicago, Chicago, IL., 920 E. 58th str, Chicago, IL 606037; 2) Computation Institute, The University of Chicago, Chicago, IL., USA; 3) Toyota Technological Institute at Chicago, Chicago, IL, USA; 4) Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA.

   Advances of high throughput genomics now allow studying biology and medicine as complex networks of interacting genetic factors in relevant biological contexts. This integrative approach holds the promise of unveiling hitherto yet unexplored levels of molecular organization and biological complexity. To support the end-to-end analytical needs of various translational projects we have developed an integrated bioinformatics platform Lynx (http://lynx.cri.uchicago.edu), a web-based knowledge base and a knowledge extraction engine. Lynx contains the following components: (a) Lynx integrated knowledge base (Lynx KB) provides a vast compendium of biomedical information integrating phenotypic, genomic, proteomic and networks information from over forty public databases. (b) Lynx analytical workbench assists the user in extracting meaningful knowledge from LynxKB and experimental data using a variety of algorithms for enrichment analysis and network-based gene prioritization. Since the last release the Lynx analytical workbench has been supplemented with new tools for reconstruction and analysis of co-expression networks, as well as a novel Cheetoh algorithm for feature-and-network-based gene prioritization. These tools facilitate the prediction of high-confidence genetic factors and molecular mechanisms contributing to the phenotypes of interest and formulating the working hypotheses to be tested experimentally. Lynx is used by a variety of translational projects for the studies of neurodevelopmental and pulmonological disorders, and cancer. The example of Lynx analysis includes the identification of high-confidence candidate genes contributing to pathogenesis of spina bifida in consanguineous family. Our analysis has predicted the deleterious mutations in the SLC19A placental folate transporter in mothers of the affected children. These mutations resulted in the narrowing of the outlet channel and consequentially to the reduced folate permeation rate and folate deficiency in affected children. Lynx is freely available on the Web. The Service Oriented Architecture provides public access to LynxKB and its analytical tools via user-friendly web services and interfaces. Future developments of Lynx include inter alia the support of the contextual isoforms-based reconstruction of biological networks and the development of the project-specific environments customized for the needs of particular interest groups and projects (e.g. cancer biomedicine).

1717W

WISEexome: Detection of copy number variations in clinical exome sequencing data based on a within-sample comparison scheme. *M.M. Weiss[1], R. Straver[1], D. van Beek[1], E.A. Sistermans[1], M.J.T. Reinders[1,2].* 1) Clinical Genetics, Genome Diagnostics, VU University Medical Center, Amsterdam, Netherlands; 2) Delft Bioinformatics Lab, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

   Clinical exome sequencing (CES) can provide a molecular diagnosis in families with an unexplained genetic disease. Both SNPs and CNVs can be pathogenic and are searched for in diagnostic settings. At the moment most clinical labs perform both CES and array to be able to detect SNPs and CNVs (resolution ~20 kb). To be able to replace array with CES we developed a method, based on a combination of our previous work (WISECONDOR; https://github.com/rstraver/)) and a segmentation algorithm, to find aberrated exons in exome data. Our method fully depends on an internal comparison of DNA fragments per probe, rather than comparisons with a reference set of samples. This works by comparing the amount of reads mapped to a probe to a set of probes known to behave alike in terms of read depth, which is off-line learned through a set of known normal samples (in our case 212 samples). This approach provides both the effect size and z-score per probe. These are merged in differently-sized windows to scout the sample for aberrations of any size. A segmentation algorithm applied to detected aberrated regions then determines their exact start and end positions. All 18 known CNVs in our dataset of 24 samples were correctly identified. Size ranges from 4 kb to 5,2 Mb. The smallest CNV contained a deletion of 2 exons and a partial deletion of an exon in a neighboring gene. Results can be sorted by a score that indicates the reliability of a call, usually providing the true positives in the top 5. We are now in the process of retrospectively analyzing our diagnostic exome cohort (~500 trio analysis) and expect to detect small pathogenic CNVs that may have been missed by the routinely used array analysis (resolution of 20 kb). Our work provides an alternative method to find CNVs in exome data. We do not require that reference samples are resequenced in the same run, and we are able to detect variable-sized CNVs. Consequently, WISEexome has the potential to replace array analysis and maybe even MLPA tests in a diagnostic setting in a cost-effective way. .

1718T

SuperVario: An analysis framework, DBMS, and common repository for variant data. *A.S. Wenocur[1], L. Tian[2], K. Nguyen[1,4], P.M. Abou-Sleiman[1,2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Center for Dynamic Data Analytics, Rutgers University, Piscataway, NJ.

In many bioinformatic analysis pipelines, nucleic acid reads are mapped and submitted to a variant caller. Typically, the variant data are stored in flat files that often are difficult to interrogate as sample numbers increase , when queried by multiple investigators. A common problem encountered is that CNVs, point mutations, indels, SNPs, structural, and other variants are tagged for specific analyses and are not easily shared or retrieved for new *in silico* experiments. A solution to this is a centralized repository for genomic variants, which can accommodate multiple classes of variant data. We have developed a database of variants, allowing researchers to apply custom filters and visualize by frequency in a specialized genome browser. The initial prototype was an instance of Exac Browser, modified to load VCFs produced at our institution. The next experiment involved configuring a Hadoop stack to store variant data. Currently, SuperVario is a lightweight but powerful C++ framework, analysis tool, and specialized DBMS for storing variant annotations and clinical metadata, with an open architecture that can accommodate replication, sharding, concurrency, and highly efficient serialization.

1719F

Leveraging PheWAS in fine-mapping complex trait GWAS loci. *F. Sathirapongsasuti, P. Fontanillas, V. Vacic, A. Auton.* 23andMe Inc., Mountain View, CA.

The interpretation and application of complex trait loci identified in genome wide association studies (GWAS) requires the ability to fine-map the causal variants and causal genes. So far, the focus of fine-mapping has been on GWAS loci of a single trait. However, large genotype-phenotype databases have enabled interrogation of variants across multiple phenotypes, an approach known as phenome-wide association studies (PheWAS). We can leverage association patterns of multiple related phenotypes to disentangle complex trait GWAS loci. We explored the effectiveness of combining association signals across phenotypes when fine-mapping a locus by analyzing all loci associated with multiple uncorrelated phenotypes in the 23andMe dataset. The 23andMe database contains genotypes of more than one million research participants and more than 1,200 phenotypes. Using this approach, the size of the credible sets was reduced substantially. As a positive control, we found that the reduced credible set at the FTO locus included the known causal variant rs1421085. We highlight the THADA locus where we combined signals from type 2 diabetes, Crohn's disease, polycystic ovary syndrome, and prostate cancer, in order to reduce the credible set to a manageable subset. Our results suggest that leveraging multiple phenotypes from a PheWAS analysis is beneficial for fine mapping complex trait loci..

1720W

MAPPIN: A method for predicting the pathogenicity and disease inheritance modes of missense variants. *N. Gosalia[1,2], O. Gottesman[1], V. Arunachalam[1], J.D. Overton[1], J.G. Reid[1], J. Staples[1], C. O'Dushlaine[1], F.E. Dewey[1], A.N. Economides[1,2], S. Balasubramanian[1].* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY; 2) Regeneron Pharmaceuticals Inc., Tarrytown, NY.

Whole-exome sequencing (WES) is increasingly being used to identify causal mutations leading to a disease or phenotype. According to the Human Gene Mutation Database (HGMD), approximately 45% of these disease-causing mutations across 5,700 genes are missense substitutions *(Stenson et al., Hum Genetics, 2014)* and, on average, an exome contains 10,500-13,500 nonsynonymous SNVs (nsSNVs) *(Levy et al., PLoS Biol, 2007; Ng et al., PLoS Genetics, 2008; Wheeler et al., Nature 2008; Wang et al., Nature 2008)*. Therefore, it is critical to differentiate potential disease-causing variants from the tolerant ones. Several algorithms are currently in use to predict the pathogenicity of missense mutations, however they are unable to distinguish between autosomal-dominant and autosomal-recessive disease-causing mutations *(Li et al., PLoS Genetics, 2013)*. Here we present MAPPIN (Method for Annotating, Predicting Pathogenicity, and mode of Inheritance for Nonsynonymous variants), a method to annotate and predict the pathogenicity of nsSNVs. We developed a classifier to distinguish between dominant, recessive, and benign missense variants using a random forest algorithm based on over 100 features including conservation, protein domains, biological networks, and allele frequency distribution. This method classifies missense mutations into three distinct groups: tolerant, haploinsufficient (dominant), or recessive with a multiclass AUC of 0.95. The algorithm performs well even under a two-class system similar to that of SIFT or Polyphen, with AUCs of 0.95 (tolerant vs. pathogenic) and 0.94 (haploinsufficient vs. recessive). We applied MAPPIN to a set of mutations derived from the Centers for Mendelian Genomics (CMG) studies *(Chong et al., Am J Hum Genetics, 2015)* and Deciphering Developmental Disorders Study *(Nature, 2015)* for which the prediction accuracy for pathogenicity <u>and</u> mode of inheritance was 73.8% and 78% respectively. To the best of our knowledge this is the first algorithm that is able to classify missense variants as dominant or recessive-disease causing.

1721T

Phenotypic-driven prioritization of trio-based whole genome sequencing data for congenital disorders. *A. Khromykh[1], S. Shah[2], D. Richards[2], D.L. Bodian[1], N.S. Hauser[1,4], R. Iyer[1], R. Baveja[3,4], M. Kane[1], T. Vilboux[1], K.C. Huddleston[1], J. Niederhuber[1], B.D Solomon[1,4,5].* 1) Inova Translation Medicine Institute, Inova Fairfax Hospital, Falls Church, VA; 2) QIAGEN Bioinformatics, Redwood City, CA; VA; 3) Fairfax Neonatal Associates, Falls Church, VA; 4) Inova Children's Hospital, Falls Church, VA; 5) Virginia Commonwealth University School of Medicine, Richmond, VA.

Congenital anomalies and genetic disorders are a leading cause of early childhood morbidity and mortality. The goal of our IRB-approved 'Impact of Genetic Disorders' study is to use trio-based whole genome sequencing (WGS) to elucidate the genetic and genomic factors that contribute to the etiologies of congenital anomalies and other disorders that present early in life. Our analyses are bolstered by our database of >7,000 WGS (along with other biological data and clinical information) derived from multiple ongoing trio-based genomic studies. We used the QIAGEN's Ingenuity Variant Analysis to analyze cases and performed analyses both with and without inputting standardized phenotypic data. The Ingenuity Variant Analysis software classifies variants using draft ACMG-recommended assessment guidelines based on a pre-curated knowledge base of biomedical literature and clinical evidence allowing standardized and scaled annotation, filtering, classification and reporting of observed clinically relevant variants. For cases in which a genetic cause for disease is suspected based on family history, but no causal variant has been previously identified, we use QIAGEN's Ingenuity Variant Analysis to identify gene variants that may perturb biological pathways/networks upstream of known causal genes in described cases with overlapping phenotypes. Data generated from an initial subset of 46 trios reveals findings of interest in 32 (70%), including *de novo* as well as dominant and recessive inheritance patterns. We compare the results of analyses with and without the use of phenotypic data to prioritize variants. Our study demonstrates the potential of the trio-based WGS analysis to identify molecular etiologies for previously undiagnosed patients, and helps provide data about the effect of phenotypic data to prioritize identified variants.

1722F

MaCHTools improves and streamlines the GWAS workflow to determine the genetic basis of D, a latent variable representing the dementing process. *J. Mitchel[1], R. Barber[1], N. Phillips[1], J. Tilson[2], D. Royall[5,6,7,8], K. Wilhelmsen[2,3,4], Texas Alzheimer's Research and Care Consortium.* 1) Molecular and Medical Genetics, University of North Texas Health Science Center, Fort Worth, TX; 2) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 3) Department of Genetics, University of North Carolina, Chapel Hill, NC; 4) Department of Genetic Medicine, University of North Carolina, Chapel Hill, NC; 5) Department of Psychiatry, The University of Texas Health Science Center, San Antonio, TX; 6) Department of Medicine, The University of Texas Health Science Center, San Antonio, TX; 7) Department of Family and Community Medicine, The University of Texas Health Science Center, San Antonio, TX; 8) South Texas Veterans' Health System Audie L. Murphy Division GRECC, San Antonio, TX.

   Reliance on commercially available genotyping arrays in genome-wide association studies (GWAS) limits coverage of the genetic variation in a study cohort. Imputation allows for additional genotypes to be included in the study, increasing coverage by orders of magnitude. MaCH is a popular and well-validated Markov chain-based haplotyping program for genotype implementation, but it has limited usability. MaCHTools is a Linux tool for handling and manipulating large genotype data sets that facilitates imputation via MaCH. MaCHTools checks input files for user-defined quality control measures, and arranges, orders and cleans datasets in preparation for imputation. MaCHTools' latest version includes a user-friendly graphical user interface. At the end of the MaCHTools-MaCH workflow, the user is left with observed genotypes, imputed genotypes, and associated and quality statistics between the all genotypes and the trait or disease of interest. MaCHTools has been validated using known datasets, and this study aims to use MaCHTools to find genetic variants that are associated with "D", a latent variable that represents the dementing process. Employing structured equation models, variance in cognitive task performance can be separated from variance unrelated to a dementing process. This "dementia-relevant" variance labeled "D", is a continuous variable that has been validated in several cohorts, including the HAAS, the NACC and the Texas Alzheimer's Research and Care Consortium (TARCC), in which this GWAS was performed. This GWAS found 24 SNPs that reached genome-wide significance with a minor allele frequence greater than 5% in association with "D" and several "D" homologs in TARCC participants.

1723W

Fast and accurate detection of circular DNA reveals its prevalence in human tissue and blood. *M. Mohiyuddin[1], H.D. Møller[2], N.B. Asadi[1], J.F. Halling[2], P. Plomgaard[3], H. Pilegaard[2], H.Y.K. Lam[1], B. Regenberg[2].* 1) Roche Sequencing, Belmont, CA 94002, USA; 2) Department of Biology, University of Copenhagen, Denmark; 3) Department of Clinical Biochemistry, Rigshospitalet, Denmark.

   We have recently shown that extrachromosomal circular DNA (eccDNA) is commonly present in the eukaryotic model, S. cerevisiae. It derives from chromosomal DNA and forms by DNA circularization through homologous recombination, non-homologous end-joining or other mechanisms. It can potentially be an intermediary for copy-number variations (CNVs) which have been implicated in several diseases, cancer, aging as well as contributing to genomic diversity. eccDNAs in S. cerevisiae were recorded by a sensitive, genome-scale enrichment and detection method for eccDNA, named Circle-Seq. eccDNAs of sizes ranging from 1kb to 38kb were identified and covered 23% of the S. cerevisiae genome including thousands of genes. In another study, tens of thousands of small eccDNAs were identified in mouse and human cell lines further confirming the abundance of eccDNA in eukaryotes. In this work, we enhanced Circle-Seq to work efficiently on the human genome. Our method used a combination of alignment signals involving discordant read-pairs, large soft-clips, secondary alignments and read-depth to detect eccDNAs. We used the enhanced Circle-Seq to conduct a case-control study on the impact of physical activity on eccDNA abundance in blood and muscle samples. The study involved blood and muscle tissue DNA samples from two groups of 8 men with ages from 60 to 65 years. These groups comprised of men who had either been physically active or inactive all their life. Our approach identified tens of thousands of eccDNAs ranging in size from 52b to 53kb. eccDNA counts in muscle samples (from 1000 to 5000 per million cells) were significantly higher than those in blood samples (from 200 to 1000 per million cells). The size distribution of eccDNAs was similar for muscle samples across the two groups. However, for blood samples, we found that the physically inactive group had more than twice the abundance of eccDNA elements for sizes ranging from 100b to 1kb. PCR validation confirmed 11 out of 12 tested eccDNAs of sizes from 400b to 35kb from multiple samples. In addition, exogenous plasmids were added in varying ratios to all samples prior to eccDNA purification and were detected at an average rate of 98.8% in blood samples. The analysis was fast and took less than three days for the 32 samples on a single node. Our results show that eccDNA is common in human blood and tissue DNA and also indicate a possible link between physical inactivity and enrichment of eccDNA in human cells.

1724T

Moving towards the biophysical genome: Biophysically-driven, high-accuracy driver discovery and variant interpretation. *C.L. Araya[1,4], A. Colavin[2,4], D.M. Fowler[3], M.P. Snyder[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Biophysics, Stanford University, Stanford, CA; 3) Genome Sciences, University of Washington, Seattle, WA; 4) Jungla Inc., San Francisco, CA.

Genomic sequences contain rich information associating genetic variants to the molecular mechanisms that underpin human health and disease, but our ability to uncover such associations is critically reliant on our understanding of the molecular functions encoded in the genome. Existing functional annotations in the genome rely on coarse, linear descriptions of genes that ultimately describe the transcriptional nature but not the functional nature of biomolecules in the genome. While structural data continues to increase, these today exist for only 20% of protein-coding genes and frequently describe as little as 15% of the structure of the each protein. We have developed methods that yield and leverage a biophysical view of how function is encoded within genes and the genome, without the requirement for structural data. Analyzing >4,700 tumor genomes from 21 cancer types, we recently demonstrated the discovery of novel cancer-drivers recurrently affecting precise molecular functions in as many as 10%, 14%, and 15% of melanoma, endometrial, and bladder cancers, respectively (Araya et al. 2016, Nature Genetics). Coherent molecular alterations often associated with changes in signaling. We extend these insights to describe a novel class of analytic solutions that accurately predict the impact of variants in specific disease-associated protein-coding genes with high accuracy. Our approach is a marked departure from existing in silico predictors; instead of generating a single predictor for the entire genome, we focus on generating solutions for specific gene-disease associations to reflect the diverse molecular mechanisms of disease-causing mutations. We apply this approach to generate solutions for 15 genes of critical clinical importance, including BRCA1, KCNQ1 and KCNH2, for which many individually rare missense variants are present in the population. Importantly, our classifiers achieve high diagnostic sensitivity without sacrificing specificity, vastly outperforming existing in silico variant impact predictors. Taken together, our gene-specific variant classifiers provide an opportunity to vastly improve the diagnostic value of existing genetic tests..

1725F

False negatives are a significant feature of next generation sequencing callsets. *D.M. Bobo[1], M. Lipatov[1], J.I. Rodriguez-Florez[2], A. Auton[3], B.M. Henn[1,4].* 1) Ecology & Evolution, Stony Brook University, Stony Brook, NY; 2) Department of Genetic Medicine, Weill Cornell Medical College. New York, NY, 10021, USA; 3) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA; 4) Graduate Program in Genetics, Stony Brook University, Stony Brook, NY, 11794, USA.

Short-read, next-generation sequencing (NGS) is now broadly used to identify rare or *de novo* mutations in population samples and disease cohorts. However, NGS data is known to be error-prone and post-processing pipelines have primarily focused on the removal of spurious mutations or "false positives" (FP) in downstream genome datasets. Less attention has been paid to characterizing the fraction of missing mutations or "false negatives" (FN) in these datasets. We design a phylogeny-aware tool to determine false negatives [PhyloFaN] and describe how read coverage, multi-allelic variants and reference bias affect the FN rate. Using thousand-fold coverage NGS data from both Illumina HiSeq and Complete Genomics platforms derived from the 1000 Genomes Project, we first characterize the false negative rate in human mtDNA NGS genomes. The false negative rate for the publically available callsets is 17-20%, even for this extremely high coverage haploid data. We demonstrate that high FN rates are not limited to mtDNA by then comparing autosomal data from 28 publically available full human genomes to intergenic Sanger sequenced regions for each individual. We examine both low-coverage (5x-10x) Illumina and high-coverage (~50x) Complete Genomics genomes. We show that the FN rate varies between ~6%-18% for these datasets and that false-positive rates are considerably lower (<3%). The FN rate is strongly dependent on calling pipeline parameters, as well as read coverage. Our results demonstrate that missing mutations are a significant feature of genomic datasets and imply additional fine-tuning of bioinformatics pipelines is needed. We provide a tool which can be used to quantify the FN rate for haploid genomic experiments, without additional generation of validation data.

1726W

MIDAS - Multiple Integration of Data Annotation Software. *R. Brumm, S. Eck, M. Schmuck, H.G. Klein.* Bioinformatics, Centre of Human Genetics, Martinsried, Bavaria, Germany.

   The implementation of Next-Generation Sequencing in a clinical diagnostic setting opens vast opportunities through the ability to simultaneously sequence all genes contributing to a certain indication at a cost and speed that is superior to traditional sequencing approaches. Especially in the case of rare, heterogeneous disorders this may lead to a significant improvement in diagnostic yield. On the other hand, the practical implementation of NGS in a clinical diagnostic setting involves a variety of new challenges which need to be overcome. Among these are the generation, analysis and storage of unprecedented amounts of data, strict control of sequencing performance, validation of results, interpretation of detected variants and reporting. Here we present MIDAS (Multiple Integration of Data Annotation Software), a central software system for data integration in a diagnostic laboratory. MIDAS is a modular constructed software system to integrate data from Laboratory Information Management System (LIMS), data from the routine Sanger sequencing workflow as well as NGS sequencing results and correlates the identified variants with the patients' phenotypical features. The phenotype is systematically recorded using the Human Phenotype Ontology standard nomenclature. In particular, Genotype-Phenotype correlations identified in one patient are made available for all other cases, to aid the interpretation and build a comprehensive knowledge base. MIDAS is implemented in Java using hibernate for database access and JavaFX as graphical user interface. Its architecture is designed modular with the basic architecture including a dynamic module loader, a user management with LDAB connection and a basic search form. According to available modules the user management and search form are adjusted; granting access for module specific views. It is possible to load every module combination, since all modules are able to run as standalone. As an advantage of this architecture, other data, such as arrayCGH or MLPA can easily be integrated by implementing new modules.

1727T

Validation of a novel, high-throughput HLA-I genotyping method based on phased, full-length sequencing. *R. Chhibber[1], D. Monaco[1], D. Wilkins[1], R. Perry[2], E. Hunter[1], J. Tang[3], D. Dilernia[1].* 1) Emory Vaccine Center, Emory University, Atlanta, GA; 2) Department of Epidemiology, University of Alabama, Birmingham, AL; 3) School of Medicine, University of Alabama, Birmingham, AL.

   Background: Current HLA genotyping methods are routinely based on partial sequencing of HLA genes or haplotype inference based on short-read Next Generation Sequencing (NGS). Obtaining the full-length sequence of each individual allele can only be achieved using phased, long-read NGS technology, but multiple technical challenges preclude its wide-spread implementation. **Objective:** To develop a novel HLA-I (A, B and C) genotyping method based on PacBio-sequencing of the complete gene and to validate this approach in Mother-Father-Child trios. **Materials and methods:** HLA class I alleles from trios were resolved using a novel NGS-based method in which SMRT technology (Pacific Biosciences) was used in conjunction with previously published computational algorithms for data analysis (Dilernia *et al* NAR 2015) to assemble the entire length of each HLA allele. The NCBI SBT interpretation tool was used for allele assignments. Accuracy was assessed by visual comparison of the sequences using Genious R9. **Results:** A total of 174 alleles were sequenced in parents to allow validation of 96 alleles obtained from offsprings. In all, 92 distinct alleles in the offsprings were readily confirmed against the parent's alleles (no mismatches). Among the four inconsistencies, three were likely missing as a result of limited number of reads in gene-specific PCR amplicons. For the remaining questionable trio sequences, we found a C*08:02:01:01 in a child instead of either C*02:01:01:01 or C*07:02:01:01, as expected from the parents' alleles. Interestingly, we found nine novel alleles of which four were among offsprings that were consistent with parental sequences (B*15:18x, A*34x, A*03x and B*35:01:01x). **Conclusions**: Our HLA-I genotyping method is able to provide the highest possible allelic resolution, with a potential to detect novel variations in sequences that are rarely covered by conventional methodology.

1728F

Detecting heterozygous Copy Number Variants (CNV) deletions in haploinsufficient genes in Clinical Whole Exome Sequencing (cWES). *T. Chiang[1], S. White[1], T.J. Wu[1], W. Bi[2], P. Liu[2], F. Xia[2], Y. Yang[2], M. Bainbridge[1], E. Boerwinkle[1,3], D. Muzny[1,2], E.B. Venner[1], W. Salerno[1], R.A. Gibbs[1,2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Human Genetics Center, University of Texas Health Science Center, Houston, TX.

Identifying CNVs in cWES is becoming routine in the clinical setting. ClinVar contains 20,546 CNV losses, 9,768 of pathogenic significance. We previously reported methods to reliably detect homozygous deletions in cWES using a dual normalization approach that examines PCA and RPKM Read Depth (RD) values to call copy number losses. Here, we extend these methods (AtlasCNV) to detect heterozygous deletions by using a Firth Logistic Regression (FLR) model trained on exon targets whose PCA-RD is less than 3 standard deviations below the mean (diploid), and RPKM-RD values less than 75% of the median RPKM. FLR models were performed for each exon of the exome capture design (VcromePK2) and the model parameters were applied to test samples. AtlasCNV results were compared with CNV calls from Illumina array (HumanExome-12v1, cSNP) for 40 samples, each with a heterozygous deletion (also detected independently on other microarrays). AtlasCNV confirmed all 40 validated CNVs (6,898 total exons). We also analyzed AtlasCNV results for 4,956 exomes, identifying 279,983 deleted exons (mean:56 exon/sample), with 95% heterozygous deletions (266,355). Evaluating these CNVs for biological significance, 55% of the samples (2708/4956) have heterozygous deletions in genes that are described in ClinGen as haploinsufficient genes linked to dosage pathogenicity. Out of 376 curated genes in ClinGen with evidence of dosage pathogenicity (289), or genes associated with autosomal recessive phenotype (85), we have samples identified with candidate heterozygous deletions in our cohort that are associated with 350 of these genes (93%). Ongoing work is focused on refining AtlasCNV precision. While approximately 50% of the exome map is not covered by the cSNP array, in regions with overlapping coverage, the concordance between cSNP and WES data for deletions is 17% at target resolution. However, 41% of exon targets in these regions have only one probe, limiting the efficacy of cSNP data in assessing the specificity and resolution of WES data. Validation data based on AtlasCNV calls was also assessed among clinical exomes in 511 Trios. Among all the heterozygous deleted targets in probands, concordant in cSNP and WES data, 68% (990/1465) follow a consistent mendelian inheritance pattern with their parents. In sum, these data demonstrate the utility of calling heterozygous CNVs in the clinical setting, and incorporating this analysis in known disease genes will boost the diagnostic rate.

1729W

Allele-specific expression from single-cell RNA-seq data. *K. Choi, N. Raghupathy, S.C. Munger, G.A. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

In diploid cells, the two allelic copies of a gene can differ in the timing and the level of their expression as determined by genetic, environmental, and stochastic factors. But until recently allele-specific expression (ASE) studies were restricted to analysis of whole tissues or population of sorted cells. The application of single-cell RNA Sequencing technologies (scRNA-Seq) opens a whole new perspective on the dynamics of allelic expression that is lost in whole tissue samples. We are now aware that allele expression occurs in bursts over short time intervals. And, as a result, scRNA-Seq data delivers a snapshot of asynchronous cells in which genes are randomly expressed from either the maternal or paternal allele (a.k.a. random monoallelic expression or RME) in many cells. However scRNA-Seq data poses multiple technological challenges that result from amplification of the cellular RNA and limited size of the RNA pool in individual cells. High level of sampling noise often coupled with low depth of coverage causes larger variance in quantified allelic expressions. Due to inefficient reverse transcription process, alleles may drop out from measurements irrespective of their abundance. Low expressed alleles are frequently missed from underlying Poisson process. Sorting out the combined effect of technical dropouts and sampling noise on top of cell-to-cell stochasticity is key to understanding the genetic regulation and dynamics of allelic expression. To overcome these problems, we developed a mixture model in an empirical Bayes framework. Our model tackles data sparsity in each individual cell by partially referring to information from other cells of a kind, and controls overdispersion by allowing cellular heterogeneity on maternal/paternal allele proportion via a common prior. Our approach also classifies genes according to the cell proportions in maternal monoallelic, paternal monoallelic, and biallelic expression states resulting from RME. We found that ASE quantification based only upon uniquely-aligning reads reports hundreds of false monoallelic expressions. The timing of maternal and paternal allelic bursts was positively correlated. Our classifier identified hundreds of genes that are highly dynamic in their allele usage patterns during different stages of mouse blastocyst development.

1730T

Linking 25,000 eMERGE participants with highly-accurate imputed HLA regions to electronic health records. *D. Crosslin[1], A. Gordon[2], P. Devi[1], A. Burt[2], D. Carrell[3], X. Zheng[4], M. Hayes[5], S. Hebbring[6], K. Kiryluk[7], S. Stallings[8], L. Bastarache[9], J. Denny[9,10], E. Phillips[10,11], D. Velez[12], T. Edwards[10], D. Roden[9,10,11], B. Namjou-Khales[13], J. Harley[13], I. Kullo[14], M. de Andrade[15], E. Larson[3], G. Jarvik[2,16], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 2) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 3) Group Health Research Institute, Center for Health Studies, Seattle, WA; 4) Department of Biostatistics, Division of Medical Genetics, University of Washington, Seattle, WA; 5) Department of Medicine, Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago IL; 6) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 7) Department of Medicine, Columbia University, New York NY; 8) Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 10) Department of Medicine, Vanderbilt University, Nashville, TN; 11) Department of Pharmacology, Vanderbilt University, Nashville, TN; 12) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 13) Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 14) Department of Internal Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 15) Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 16) Department of Genome Sciences, University of Washington, Seattle, WA.

The Electronic Medical Records and Genomics Network (eMERGE) has developed eMERGEseq, a CLIA-grade sequencing panel of 106 genes comprising the ACMG 56 and site-requested genes focused on clinical actionability and approximately 1500 SNPs. The goal is to return these results through the electronic health record (EHR) for clinical decision support for 25,000 participants over a three-year period. Of the 1500, we selected 275 SNPs common among Illumina 1M Duo, OmniQuad, OmniExpress, 660K and 550K platforms to assist in the imputation of the HLA region using HIBAG-HLA genotype Imputation with attribute BAGging [Zheng 2013]. This is includes *HLA-A, B, C, DRB1* and *DQB1* regions. Alternative to tagging strategies, HIBAG combines the concepts of attribute bagging with haplotype inference from unphased SNPs and HLA types. Attribute bagging is a method for improving the accuracy and stability of classifier ensembles deduced using bootstrap aggregating and random subsets of variable, with pre-fit classifiers for European, Asian, Hispanic, and African ancestries [Zheng 2013]. In addition to returning results to the EHR, another goal of eMERGE is genetic discovery, and we believe having 25,000 participants with imputed common variants will provide unprecedented resolution for this region, and ultimately genotype-phenotype associations. Once such phenotype is PheWAS, which is a technique to harness a wide range of physiological and/or clinical outcomes and phenotypes by using algorithms to parse EMR data [Denny 2013]. We hypothesize that we will identify the HLA region as having strong pleiotropic effects, and will be well-powered with 25,000 participants to detect these associations. The first 1000 participants with colorectal cancer are being sequenced at the time of this abstract proposal, and will provide pilot data for this study. Many of the participants will have orthogonal data of these regions including exomes, genomes, and Sanger sequencing, which will provide validation and/or comparison for the imputation results. In addition, there are *HLA-B* tagging SNPs on eMERGEseq, which will allow comparison of imputation using tagging and attribute bagging using HIBAG. Having high-confidence calls in this region in this sample size linked to EHRs will prove beneficial both in discovery and implementation science.

1731F

Comparative analysis of global gene expression profile in mouse strains after hypoxic challenge either susceptible to or resistant to chronic hypoxia-induced pulmonary hypertension (PH). *N. Dasgupta[1], P. Hale[1], M. Pauciulo[1], T. LeCras[2], P. Pastura[2], W. Nichols[1].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Pulmonary arterial hypertension (PAH), characterized by high pulmonary artery pressure and remodeling of the small pulmonary arteries, can progress to right heart failure and death if untreated. While previous studies have identified *BMPR2* and other genes in the pathogenesis of PAH, data suggest additional genetic factors contribute to the disease. The goal of our study is to identify novel genes and/or genetic modifiers and pathways contributing to the development of PH. Phenotyping of RVSP and RVH in 33 inbred strains after hypoxic exposure identified one high responding (PL/J; RVSP=85mm Hg) and two low responding (MRL/MpJ, RVSP=37mm Hg; FVB/nJ, RVSP=47mm Hg) strains. To compare responses at the molecular level in PL/J compared to MRL/MpJ and FVB/nJ, RNASeq and miRNASeq was performed using total RNA prepared from lung tissue. Global profiles of the differentially expressed mRNAs (DEGs) and microRNAs (DEmiRs) were studied for all 3 strains at 3 different exposure times (3days; 1week; 3weeks) using 0hrs as baseline. Total altered DEGs from all 3 strains and exposures were stratified into chronic and acute response using 3way Venn diagrams. There were almost twice as many upregulated DEGs compared to the downregulated DEGs in all strains and time points except at 3 weeks for MRL and FVB. PL/J had the maximum number of DEGs at all 3 time points. Similar analyses performed on the miRNAs showed that DEmiRs in PL/J and FVB/nJ strains had a comparable profile at 3 days and 1 week. The number of DEmiRs almost doubled at 3 weeks. In contrast, the number of DEmiRs in MRL/MpJ was sharply reduced at 3 weeks as compared to 3 days and 1 week. There were 35 pathways common across the 3 strains. 71 pathways were shared between MRL/MpJ and FVB/nJ but absent in PL/J. No unique pathways were identified in PL/J. IPA analyses showed the top functional pathways of the acute and chronic response DEGs were inflammation which constituted 19-24% of the total DEGs in each strain/exposure length while cell growth and proliferation DEGs constituted 17-29%. Our results suggest a differential pattern of gene expression in mouse strains demonstrating differences in response to hypoxia. The predicted interactions between DEmiRs and DEGs suggest regulation of the observed differential gene expression under hypoxic conditions. Additional studies are warranted to validate these interactions and determine their possible contribution to the development of chronic hypoxia-induced PH.

1732W

Robust lineage reconstruction from high-dimensional single-cell data. *S. Garcia[1,2], G. Giecold[1,2], E. Marco[1,2], L. Trippa[1,2], G.-C. Yuan[1,2,3].* 1) Biostatistics and Computational Biology, Dana-Farber Cancer Institute, MA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; 3) Harvard Stem Cell Institute, Cambridge, MA 02138, USA.

Single-cell gene expression data provide invaluable resources for systematic characterization of cellular hierarchy in multi-cellular organisms. However, cell lineage reconstruction is still often associated with significant uncertainty due to technological constraints. Such uncertainties have not been taken into account in current methods. We present ECLAIR (Ensemble Cell Lineage Analysis with Improved Robustness), a novel computational method for the statistical inference of cell lineage relationships from single-cell gene expression data. ECLAIR uses an ensemble approach to improve the robustness of lineage predictions, and provides a quantitative estimate of the uncertainty of lineage branchings. We show that the application of ECLAIR to published datasets successfully reconstructs known lineage relationships and significantly improves the robustness of predictions. ECLAIR is a powerful bioinformatics tool for single-cell data analysis. It can be used for robust lineage reconstruction with quantitative estimate of prediction accuracy.

1733T

Ultrafast comparison of personal genomes. *G. Glusman, M. Robinson, D.E. Mauldin, Big Data for Discovery Science (BDDS) Center of Excellence.* Institute for Systems Biology, Seattle, WA.

We present here an ultra-fast solution to the problem of comparing genomes across sequencing technologies and reference genome versions; our solution also preserves privacy. Comparing genomes can be a slow, complicated and error-prone bioinformatic procedure. Personal genomes are large, consisting of millions of variants represented relative to a reference genome. The position of a genomic variant depends on the reference version used, can be "zero based" or "one based", and the chromosomes may be labeled using different schemes (e.g., "NC_000003.12", "chr3" or just "3"). Sequencing the same genome using different technologies can yield different results, as each technology has its own biases. Even sequencing the same genome repeatedly with the same technology can give somewhat different results due to the stochastic nature of genome sequencing, batch effects, or differences in the computational pipelines used. We have developed a method for transforming a standard genome representation (i.e., a list of variants relative to a reference) into a "fingerprint" of the genome involving both common and rare variants. This transformation does not require knowledge of the technology, reference and encoding used, and yields fingerprints that can be readily compared without requiring 'lifting over' coordinates. Fingerprint sizes can be tuned to adjust sensitivity for different applications. Thanks to their reduced size, computation on the genome fingerprints is fast and requires little memory. This enables scaling up a variety of important genomic analyses. Computing a fingerprint from a whole-genome VCF typically takes 10-15 seconds; comparing two fingerprints, 0.5 milliseconds. Thus, a massive all-against-all comparison of 11,726 genome fingerprints (68,743,675 pairwise comparisons) took just 10 CPU hours. The results clearly separate comparisons of the same genome, comparisons of closely related individuals, and comparisons of unrelateds. Within a family, relationships up to second cousin are recognizable. The fingerprints also enable reconstruction of population structure in a fraction of the time, and facilitate a variety of crucial tasks like selecting suitable controls for a set of cases and computing kinship matrices. Since the original genome representation cannot be reconstructed from its fingerprint, our method has significant implications for privacy-preserving genome analytics. This Big Data for Discovery Science work is supported by NIH 1U54EB020406.

1734F

Envision: Inferring the functional effects of non-synonymous variants using experimental results from large-scale mutagenesis. *V.E. Gray, R.J. Hause, J. Shendure, D.M. Fowler.* Genome Sciences, University of Washington, Seattle, WA.

Current computational predictors for mutational effect focus on the binary consequences of mutations, *e.g.,* deleterious or not. Recently, technological advances have afforded high-throughput methods to quantify mutational effects on protein function. Here, we leverage large-scale mutagenesis data sets comprising tens of thousands of quantitative mutational effect scores for several proteins and protein domains to train a computational tool for predicting mutational effect scores. Our tool, Envision, was trained using gradient boosting machine learning and uses evolutionary conservation, biochemical, and structural annotations to predict both categorical and quantitative effects of single amino acid mutations. Envision is highly accurate both for classification and regression in 10-fold cross validation. We validated Envision in several ways, including on large-scale mutagenesis data not included in model training and on other mutational databases like the Protein Mutant Database. In all cases, we find that Envision outperforms other predictors, except when those predictors were trained on the testing data in question.

1735W

Novel dog genome content revealed by pseudo-*de novo* assembly of unmapped sequence reads. *L.A. Holden[1], M. Arumilli[2], M.K. Hytönen[2], H.T. Lohi[2], K.H. Brown[1].* 1) Department of Biology, Portland State University, Portland, OR, USA; 2) 2Research Programs Unit, Molecular Neurology, University of Helsinki and Folkhälsan Institute of Genetics, Helsinki, Finland.

Dog breeds have distinct genetic backgrounds due to historical breeding practices aimed to produce desired traits. Along with the intense selective breeding for behavioral and physical characteristics comes the predilection for certain diseases. Many of the diseases found in purebred dogs also exist in humans such as diabetes, cancer, and inflammatory disease. Not only are dogs known as man's best friend, but they also serve as a unique tool for studying the genetic basis of diseases. The current dog reference genome is based on a single female Boxer and remains poorly annotated—especially breed-specific variants—yet is widely relied upon to understand the genomic components of simple and complex diseases. To examine genomic content that may not be captured by the reference genome, we evaluated whole genome sequence reads from 26 Border Collie individuals that did not align to the Boxer breed dog reference genome. We trimmed raw sequence reads on quality score and aligned trimmed reads to the reference genome producing 6.7 billion paired reads and 1.1 billion orphaned reads, of which 92% and 99% aligned to the reference genome, respectively. 7% of the Border Collie paired reads aligned discordantly and 1% of both paired and orphaned reads did not align to the Boxer reference genome. We assembled discordant and unmapped reads *de novo* for each individual. Primary assembly of contigs per individual resulted in an average of 454,934 contigs covering 143.5 Mb with an N50 of 301. All primary assembly contigs were again assembled *de novo* into a secondary assembly (all individuals combined) producing 124,488 contigs covering 125.9 Mb of non-redundant sequence with an N50 of 1158. BLASTn of 1000 randomly selected secondary assembly contigs against NCBI's non-redundant nucleotide database showed that 28% of contigs have no known matches in sequenced species and 9% of contigs preferentially align to the human genome over the dog genome. This study reveals novel genomic content present in unmapped sequence reads in the Border Collie that does not exist in the Boxer reference genome. Extending the dog genome by including information found in unmapped reads may prove critical in understanding the breed identity and the genetic component of diseases shared by humans and dogs.

1736T

A functional gene network model for pathogenic bacteria. *H. Kim[1,2], M. Lee[2], I. Lee[2].* 1) Korea Institute of S&T Evaluation and Planning, Seoul, South Korea; 2) College of Life Science and Biotechnology, Yonsei University, Seoul, South Korea.

   Purpose *Xanthomonas oryzae pv. oryzae* is a gram-negative pathogenic bacteria that causes bacterial blight. The purpose of this sutdy is to construct a functional gene network model for *Xanthomonas oryzae pv. oryzae*, dubbed XooNet. Through the predictive network model, we expect to provide better insights and explanations for infection mechanism. And also, we expect that this approach is extended to human pathogenic bacteria research.   Methods Diverse omics data types were incorporated to construct XooNet. First, we inferred co-functional links from 4 distinct omics data, i.e., co-expression, co-occurrence of prtein domains, and two types of genomic contexts between *Xanthomonas oryzae pv. oryzae* genes. In addition, co-functional links from 6 omics data of other bacteria networks, i.e., EcoliNet and PseudomonasNet, were trasnferred by associalogs. Then, all the co-functional links derived from ten distiguished omics data types were integrated to XooNet using a variant of naïve Bayesian integration.   Results XooNet is comprised of 103,426 co-functional links among 3,608 protein coding genes, which covers about 82% genome of *Xanthomonas oryzae. pv oryzae*. For assessment of prediction power, benchmarking test and receiver operating characteristic curve analysis were performed with genes involved in known biological pathways. As a result, we have confirmed that XooNet implements diverse biological pathways of interest and is useful to predict unknown gene functions.

1737F

VarSome: A free knowledge base and aggregator for human genomic variants. *C. Kopanos, V. Tsiolkas, M. Albarca, C. Chapple, A. Massouras.* Saphetor, Lausanne, Switzerland.

   There is now a wealth of information available about human genomic variants. However, this is usually distributed across multiple, independent databases, which greatly complicates the collection of available information on a given variant. Here, we present VarSome, a **novel knowledge base and aggregator for human genomic variants**. It provides an intuitive web-based interface, offering access to the more than 14 billion items of variant and gene annotation we have accumulated in less than a second per query.VarSome provides functional annotation, calculated in real time using the RefSeq and Ensemble transcript databases; it correctly matches (and displays) equivalent indels to annotations; it displays disease and phenotype associations, drug interactions, relevant clinical trials, population frequencies, and prediction scores from a multitude of sources. The large number of external sources is regularly updated to their most recent release.Variants can be searched in Varsome in many different ways including dbSNP rsIDs, explicit descriptions using HGVS notation or genomic coordinates (e.g. BRAF:V600E, chr6-161127501-A-G) or even by pasting a line from a VCF file. VarSome is not limited to known variants: the user may obtain functional annotation for any variant.Users can post comments on specific variants which will be visible to anyone else who visits that variant. This provides an easy and centralized way of sharing information about known or novel variants. VarSome also provides stable, unique links to any variant, known or not, facilitating the sharing of variants between collaborators.VarSome is freely available to the community on varsome.com.

1738W

Localized structural frustration for evaluating the impact of sequence variants. *S. Kumar[1,2], D. Clarke[1,2], M. Gerstein[1,2,3].* 1) Molecular Biophysics and Biochemistry, Yale University; 2) Program in Computational Biology and Bioinformatics, Yale University; 3) Department of Computer Science, Yale University,New Haven, CT 06520, USA.

   With rapid decline in sequencing cost, a large amount of effort has been invested in cataloging protein-coding variants in various cohorts. Growing sequence datasets are increasingly uncovering large numbers of rare single-nucleotide variants (SNVs) in coding regions of the genome. This rarity renders conventional variant-phenotype associations unfeasible when trying to evaluate the potential deleteriousness of SNVs. Protein structures may help to provide the needed means for inferring otherwise difficult-to-discern rare variant induced phenotypic effects. Previous approaches to evaluate the effect of coding variants on protein structures have primarily focused on quantifying global stability changes. However, local perturbations can severely impact functionality (such as catalysis, allosteric regulation, interactions and specificity) without strongly disrupting global stability. In this work, we utilize localized frustration as a metric to quantify the coding variant induced local structural alterations in protein structures. Many of our observations in this study are intuitively consistent; such as the finding that disease-associated SNVs tend to induce strong changes in localized frustration. However, we also find intriguing results in the context of cancer-associated genes: SNVs in oncogenes induce stronger perturbations at the surface, whereas those in tumor suppressor genes (TSGs) induce stronger perturbations in the interior. These findings clearly highlight the efficacy of quantifying local structural changes to assess the functional impact of coding variants. We believe that our study is opportune considering the imminent growth in rare genomic variation datasets through various public and private sequencing initiatives.

1739T

Highly accurate alignment and variant calling by using population genome graphs. *W. Lee[1], J. Browning[1], A. Dolgoborodov[2], K. Ghose[1], A. Jain[2], A. Komoroczi[2], D. Kural[1], B. Pollex[2], G. Rakocevic[3], V. Semenyuk[1], J. Spencer[2], I. Spulber[2], R&D Seven Bridges Genomics.* 1) Seven Bridges Genomics, Cambridge, MA; 2) Seven Bridges Genomics, London, UK; 3) Seven Bridges Genomics, Belgrade, Serbia.

   The 1000 Genomes Project (1000GP) has analyzed genomes from twenty-six populations. A key insight from 1000GP is that most of the variants in an individual are shared by the population. NA12878 is a sample from CEU and EUR as sub- and super-population. Among NA12878's four million variants, there are only 1.3% uniquely belonging to NA12878 in CEU. In other words, other samples in CEU share 98.7% variations with NA12878 which is the largest number compared with other populations. Against EUR, NA12878 has only 0.8% unique variants and the numbers for other super populations are 4.4%, 1.4%, 1.5% and 1.1% for EAS, SAS, AFR and AMR respectively. This has led to the hypothesis that, by incorporating known variants into the current linear reference, we can improve alignment, variant calling and genotyping. To address the challenges of population-scale genomics, Seven Bridges is pioneering the development of graph genomes. A directed acyclic graph (DAG) is constructed by combining the linear reference genome with a list of variants in a population. The backbone of a graph is the linear reference genome while each branch on the graph indicates a genomic variant, guiding an alternative path for read alignments. We then work on a graph read mapper that maps reads onto a DAG representing a reference population genome. The mapper is designed to be interoperable with existing bioinformatics pipelines and accepts inputs in standard format and produces alignments as BAM files. FASTA with VCF are used for graph construction while FASTQ provides short reads of a sample. Reads are mapped to the graph reference using a global search algorithm followed by a local mapping algorithm. Furthermore, highly sensitive genotype information can be provided by giving the graph path of each alignment. We simulated sixty million paired-end reads with 3.4 million variants including SNPs and INDELs. Since having a perfect graph for aligning a sample is impossible we then simulated different sizes of graphs for benchmarking, either larger or smaller. Each graph may have only fifty percent variants overlapping with the sample. Calling variants by GATK, BWA's alignments achieves 0.889 precision-recall while the graph alignments have 0.948, 0.915 and 0.88 based on the perfect, imperfect and none graphs respectively. Graph alignments of NA12878 shows that several false possible SNPs reported in dbSNP are caused by INDELs nearby which demonstrates the power of graph genomes. .

1740F

InterVar: Clinical interpretation of genetic variants by ACMG 2015 guidelines. *Q. Li, K. Wang.* Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA.

   Statement of purpose: The American College of Medical Genetics and Genomics (ACMG) published in 2015 the updated standards and guidelines for the clinical interpretation of sequence variants, based on 28 pieces of evidence. However, for many clinical labs, implementing these variant scoring rules into a standardized workflow is difficult with available informatics tools or datasets. An automated tool or web service that combine all annotation databases and offer a one-stop shop for human interpreters is urgently needed, which can help human interpreters quickly understand the clinical significance of genetic variants. **Methods:** We present such a tool, InterVar (clinical <u>Inter</u>pretation of genetic <u>Var</u>iants), to fill these unmet needs. InterVar can take a VCF file and automatically generate necessary evidence based on the current ANNOVAR annotation results. InterVar mainly consists of two steps: 1) automated interpretation and 2) manual re-interpretation by user-supplied domain knowledge. The companion web server called wInterVar(wintervar. wglab.org) is developed for searching or re-interpreting all possible missense variants. **Results**: (I) We performed automated interpretation by InterVar on a dataset of 9,305 *de novo* variants from 12 published neurodevelopmental disorders studies with 6,515 cases and 959 controls. Comparing cases versus controls, we detected a strong enrichment of pathogenic/likely pathogenic (P/LP) variants (P=0.0087), based on InterVar's automated interpretation. There are no strong enrichment of P/LP variants using SIFT and PolyPhen. (II) We analyzed the entire ClinVar (version 2016-03-02), and compared their annotations with InterVar. InterVar is largely consistent (93.1%) with ClinVar on the benign variants, but has large differences in the pathogenic category. (III) We interpreted previously reported 616 clinically actionable variants in 112 actionable genes. For InterVar results, only 34 (5.5%) variants were classified as pathogenic/likely pathogenic variants. The vast majority of the "pathogenic" variants annotated by HGMD or other studies are probably not pathogenic, which is consistent with recent literature. In summary, InterVar can facilitate human reviewers to interpret the clinical significance of variants in human genetic studies. InterVar is especially useful to address severe congenital or very early onset developmental disorders with high penetrance.

1741W

New resources and updates at the Database of Genomic Variants (DGV). *J.R. MacDonald, M. Zarrei, J. Whitney, D. Merico, S.W. Scherer.* The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

   The Database of Genomic Variants (http:/dgv.tcag.ca) was established in 2004, and has served as a publicly accessible, curated catalogue of structural variation (SV) found in the genomes of control individuals. The current version of DGV consists of 72 published studies, comprising >6 million entries ranging from 50bp to 3 Mb in size, identified in >36,000 unique genomes. The latest release includes data from 5 new studies, and includes novel collections with > 3,000 samples from previously under-represented populations. Currently over 75% of the genome is covered by CNV and this estimate is likely inflated as a result of contributions from earlier studies, using low resolution platforms such as BAC CGH, which have been shown to overestimate the CNV boundaries. To address the overabundance of CNV, we have generated a new resource that utilizes the most current and high-resolution studies to generate a Gold Standard set of CNVs. Each study in DGV was evaluated for sensitivity and specificity and 32 studies were selected as they were found to have a low false positive rate with accurately defined boundaries. A cross study meta-analysis was performed to generate a reference set of variants found in the > 19,000 samples tested in these studies. This allowed us to leverage large unique cohorts, and complementary approaches across many studies to achieve a well-defined set of variants. To reduce the number of false positives in this dataset, the curated Gold Standard variants were required to have support from more than one study and found in more than one sample. This will help to reduce technological biases, and also remove sample specific artefacts by compensating for individual study biases. This new resource contains a total of 38,185 CNV regions which were identified from over 5 million sample level CNVs used as input. This new resource will aid the interpretation of new CNV findings for both clinical and research applications.

1742T

CNV detection from targeted next-generation panel sequencing data in routine diagnostics. *A.M. Nissen, C. Rapp, M. Locher, A. Laner, A. Benet-Pagès, E. Holinski-Feder.* Medical Genetics Center Munich, Munich, Bavaria, Germany.

In times of next generation sequencing (NGS), a single analysis approach to detect SNVs and CNVs from the same data source would be of great benefit for routine diagnostics. However, CNV detection from exon-capture NGS data has no standard methods or quality measures so far. Current approaches depend solely on read depth data which is systematically biased by exon capture. Different enrichment methods and variation within sequencing runs makes data normalization a challenging task. We developed a bioinformatics method to detect exonic CNVs using read depth data derived from targeted NGS panels. The development of our pipeline is based on: (1) utilization of more than one detection tool, (2) different reference sets for different kits, (3) normalization against samples from the same sequencing run, (4) definition of special quality thresholds for single exon events. In detail, our CNV pipeline is a combination of the open source tools ExomeDepth, Clamms, Canoes and Codex, and an in-house developed method that utilizes Mann-Whitney U test and heterozygosity check. All tools are based only on read depth data wherefore capture efficiency per exon was assessed with a reference set of CNV negative patients to identify reliable regions. In these regions, a CNV is called if at least two out of the five methods are concordant for the respective CNV. The pipeline showed a CNV detection sensitivity of 77.46% and precision of 94.83% in a test set of 71 patients with pathogenic CNVs previously detected by MLPA. We correctly detected 42 deletions and 13 duplications with high confidence. False negatives (22.54%, 16 CNVs) were proven to be one exon events or located in pseudogene regions and were detected by only one of the detection tools (mostly ExomeDepth). To minimize false negatives, the pipeline is currently adjusted to call single exon events under specifically adapted quality parameters with only one out of the five tools. Performance of the pipeline was successfully tested with different enrichment systems (Agilent, Illumina) and data from external laboratories. However, this method is only applicable for selected genes and non-reliable regions as well as single exon CNVs with low significance should be confirmed by another method.

1743F

Integrating mean and variance heterogeneities to identify differentially expressed genes. *W. Ouyang[1,2], Q. An[2,3], J.Y. Zhao[3], H.Z. Qin[1,2].* 1) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA; 2) Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA; 3) Department of Epidemiology, Tulane University,, New Orleans, LA.

In functional genomics studies, mean heterogeneity tests have been widely employed to identify differentially expressed (DE) genes with distinct mean expression levels under different experimental conditions. Variance heterogeneity (aka, the difference between condition-specific variances) of gene expression levels is simply neglected or adjusted for as an impediment. The mean heterogeneity in the expression level of a gene reflects one aspect of its distribution alteration; and variance heterogeneity induced by condition change may reflect another aspect. Change in condition may alter both mean and some higher-order characteristics of the distributions of expression levels of susceptible genes. In this report, we put forth a conception of mean-variance differentially expressed (MVDE) genes, whose expression means and variances are sensitive to the change in condition. We mathematically proved the null independence of existent mean heterogeneity tests and variance heterogeneity tests. Hereby, we proposed an integrative mean-variance test (IMVT) to combine gene-wise mean heterogeneity and variance heterogeneity induced by condition change. The IMVT outperformed its competitors under comprehensive simulations of normality and Laplace settings. For moderate samples, the IMVT well controlled type I error rates, and so did existent mean heterogeneity test (i.e., the Welch t test (WT), the moderated Welch t test (MWT)) and the separate tests on mean and variance heterogeneities (SMVT), but the likelihood ratio test (LRT) severely inflated type I error rates. In presence of variance heterogeneity, the IMVT appeared noticeably more powerful than all the valid heterogeneity tests. Application to the gene profiles of peripheral circulating B raised solid evidence of informative variance heterogeneity. After adjusting for background data structure, the IMVT identified novel experiment-wide significant MVDE genes, while all the valid competitors did not identified any experiment-wide significant DE genes. Our results indicate tremendous potential gain of integrating informative variance heterogeneity after adjusting for confounders. The proposed informative integration test better summarizes the impacts of condition change on expression distributions of susceptible genes than do the existent competitors. Therefore, particular attention should be paid to explicitly exploit the variance heterogeneity induced by condition change in functional genomics analysis. .

1744W

ClinGen Pathogenicity Calculator. *R. Patel, N. Shah, A. Jackson, R. Ghosh, P. Pawliczek, S. Paithankar, A. Baker, K. Riehle, H. Chen, X. Feng, S. Milosavljevic, S. Plon, A. Milosavljevic. Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.*

   The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) recently provided guidelines for sequence variant interpretation. To enable automated large-scale application of the ACMG/AMP guidelines in a consistent way, we have developed the ClinGen Pathogenicity Calculator which automatically calculates provisional variant classifications based on evidence provided by users. The Calculator also makes reasoning explicit by documenting (a) applicable ACMG/AMP rules, (b) evidence codes applied for the variant and (c) links to data provided by users to support variant classification. The Calculator is accessible at http://calculator.clinicalgenome.org along with use cases and accessory services. Using the current Calculator interface, a registered user may (1) identify an allele already present in the ClinGen Allele Registry (CAR); (2) register new alleles; (3) enter ACMG/AMP-style evidence tags and links to supporting data for each tag; (4) generate ACMG/ AMP guideline-based pathogenicity classification; (5) store and share reasoning behind each conclusion; (6) generate variant reports in PDF format. The Calculator design is flexible and allows customization of guidelines, for specific genes and conditions. For example, we customized the calculator for diagnosis of inborn metabolic disorders by including evidence codes and rules that take into account metabolite levels and additional evidence codes to support the diagnosis of MYH7 related cardiomyopathy. So far there are ~100 registered users of the calculator from various parts of the world. We expect the ClinGen Pathogenicity Calculator to facilitate the improved use of the ACMG/AMP variant guidelines and sharing of variant classification.

1745T

ClinGen Allele Registry. *P. Pawliczek[1], R. Patel[1], A. Jackson[1], N. Shah[1], C. Bizon[3], S. Rynearson[6], T. Nelson[5], S. Harrison[2], D. Azzariti[2], B. Powell[7], R. Freimuth[4], L. Babb[2], A. Milosavljevic[1]. 1) Baylor College of Medicine, Houston, TX; 2) GeneInsight, Sunquest Information Systems; 3) RENCI, Chapel Hill, North Carolina; 4) Mayo Clinic Rochester Minnesota; 5) Geisinger autism and developmental medicine, Lewisburg, PA 17837; 6) UCGD,University of Utah; 7) University of North Carolina,Genetics Medicine, Chapel Hill, North Carolina.*

   Current allele naming systems define alleles in the context of specific reference sequences. It is therefore hard to collate information about an allele from various sources when they refer to the same allele in the context of different contigs, genome assemblies, transcripts, or amino acid sequences. Moreover, naming systems such as HGVS may provide numerous equivalent representations of the same allele such as indel in the context of the same reference sequence. Unique variant identifiers may currently be obtained only upon a lengthy manual process of submitting the variant to one of a few databases that provide "de-duplication" or "canonicalization" services as part of their submission or data curation processes. To address the need for fast and simple programmatic allele registration services and thus catalyze the growth of knowledge about genetic variation, we developed the ClinGen Allele Registry. The registry provides query, registration and "canonicalization" services via an interactive user interface at http://reg.clinicalgenome.org . To facilitate linking of allele data within ClinGen and globally, all the functionalities of the registry are also exposed for programmatic access via JSON-LD (W3C standard compliant) REST APIs documented at reg.clinicalgenome.org . When provided an HGVS expression, the registry returns an existing or newly-registered allele identifier within a few seconds. The canonical identifier corresponds to a dereferencable URI/URL that may be used to programmatically retrieve allele representations across genome assemblies and transcripts and obtain alternative identifiers of the allele across key resources such as ClinVar and dbSNP. The registration process includes validation and canonicalization by alignment of the allele against genome assemblies and transcripts (NCBI and ENSEMBL). The registry implements conceptual and resource models for alleles documented athttp://datamodel.clinicalgenome.org/development/allele/. We illustrate registry features by showing how it may be used to accomplish the following: (1) Resolve identity of repetitive indels in ClinVar; (2) Collate information about an allele from different sources using Linked Data methods; (3) Provide programmatically accessible registry services to applications such as the ClinGen Pathogenicity Calculator.

1746F

Disentangling expression across genes, tissues and individuals by Bayesian tensor decomposition. *I. Pe'er, S. Yang, M. Wang.* Computer Science, Columbia University, New York, NY.

   Expression profiles are standardly analyzed as matrices that run genomewide and across multiple samples. Modern data allows distinguishing samples across individuals vs. across tissues. The resulting three-dimensional tensor structure obviates the faults of two-dimensional analysis, that either averages the data along one of the tensor axis, thus losing information, or treats it as an unstructured set of profiles, introducing redundant parameters.   We propose a fully-Bayesian model for low-rank decomposition of expression tensors. The underlying assumption is that measured expression profiles tally weighted combinations of idealized ones specific to biological processes. Weights are determined by tissue and separately, by individual. Expression tensors are thus approximated by sums of outer products of vector factors. Inferring vector coefficients separately using Gibbs sampling overcomes the computational limitation of inferring such a model, while using a prior that encourages similarity facilitates sharing information across tissues, genes and individuals. The Bayesian formulation conveniently sidesteps arbitrary parameter-tuning. Initialization is conducted using two-dimensional methods akin to principal component analysis.   We demonstrate accuracy of the approach on simulated data, and apply it to real RNA-seq data of brain regions from GTeX. We validate the discovered factors in an independent dataset by the CommonMinds consortium.

1747W

Data-driven computational modeling of three-dimensional genome architecture: Lessons from CTCF-and RNAPII-mediated chromatin interactions. *D. Plewczynski[1,3], P. Szalaj[1,3,4], Z. Tang[2], P. Michalski[2], M. Kadlof[1], P. Wroblewski[1], M. Sadowski[1], G. Mazzocco[1], O. Luo[2], Y. Ruan[2].* 1) Centre of New Technologies, University of Warsaw, Warsaw, mazovia, Poland; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) Centre for Innovative Research, Medical University of Bialystok, Białystok, Poland; 4) Hasselt University, Belgium.

   Chromosomal folding are important features of genome organization, which play critical roles in genome functions, including transcriptional regulation. Using 3C-based mapping technologies to render long-range chromatin interactions has started to reveal some basic principles of spatial genome organization. Among 3D genome mapping technologies, ChIA-PET is unique in its ability to generate multiple datasets (in a single experiment), including binding sites, enriched chromatin interactions (mediated by specific protein factors, like CTCF), as well as non-enriched interactions that reflect topological neighborhoods of higher-order associations. The multifarious nature of ChIA-PET data represents an important advantage in capturing multi-layer structural-functional information, but also imposes new challenges in multi-scale modeling of 3D genome.We applied an advanced ChIA-PET strategy combined with computational modelling to comprehensively map higher-order chromosome folding and specific chromatin interactions mediated by CTCF and RNAPII with haplotype specificity and nucleotide resolution in different human cell lineages. We find that CTCF/cohesin-mediated interaction anchors serve as structural foci for spatial organization of constitutive genes concordant with CTCF-motif orientation, whereas RNAPII interacts within these structures by selectively drawing cell-type- specific genes towards CTCF-foci for coordinated transcription. Haplotype-variants and allelic-interactions have differential effects on chromosome configuration influencing gene expression and may provide mechanistic insights into functions associated with disease susceptibility. Our 3D-genome simulation suggests a model of chromatin folding around chromosomal axes, where CTCF is involved in defining the interface between condensed and open compartments for structural regulation.Finally, we present 3D GeNOme Modeling Engine (3D-GNOME) - a web service which generates 3D structures from 3C data and provides tools to visually inspect and annotate the resulting structures, in addition to a variety of statistical plots and heatmaps which characterize the selected genomic region. 3D-GNOME simulates the structure and provides a convenient user interface for further analysis. 3D-GNOME is freely available at http://3dgnome.cent. uw.edu.pl/ providing unique insights in the topological mechanism of human variations and diseases.

1748T

GATTACA: Lightweight species deconvolution in metagenomes. *V. Popic[1], V. Kuleshov[1], S. Batzoglou[1], M. Snyder[2].* 1) Computer Science, Stanford University, Stanford, CA; 2) Genetics, Stanford University School of Medicine, Stanford, CA.

Despite rapid advances in sequencing technologies, determining the genomic sequences of bacterial species within complex metagenomes poses significant challenges. Existing approaches require specialized binning (Iversen et al, 2012; Albertsen et al., 2013) or complex library preparation techniques (Adey et al., 2014; Kuleshov et al, 2016), which may be expensive or inaccurate. Recent work has shown that genomes of individual bacteria can be accurately recovered from cohorts of metagenomes by clustering genomic sequences using their co-abundance patterns across multiple samples (Nielsen et al. 2014; Alneberg et al., 2014). To estimate abundance, these tools map reads of each sample against each genomic sequence (e.g. pre-assembled metagenomic contig). This methodology assumes the availability of a large cohort of samples and of sizable compute and storage resources to perform read alignment. To alleviate these requirements, we introduce GATTACA, a lightweight alignment-free tool that enables fast binning of genomic sequences from a single metagenomic sample with no significant loss of accuracy. At a high level, GATTACA builds a small index of kmer counts from a reference panel of metagenomic samples; it then uses this index to estimate co-abundance levels of arbitrary genomic sequences and cluster them using variational mixture models. The GATTACA index enables orders-of-magnitude savings in space and time for abundance estimation. We tested GATTACA on both synthetic and real data used to benchmark earlier methods aimed at cohorts of metagenomes. On the synthetic species dataset of Alneberg et al. (2014), GATTACA produces clusters with an accuracy that is comparable with their tool (CONCOCT); unlike CONCOCT, GATTACA takes as input only a single metagenomic sample and a 10GB index and runs on a standard laptop computer. Furthermore, using an index derived from 100 samples of the Human Microbiome Project study, GATTACA accurately identifies clusters of contigs corresponding to the genome of Bifidobacterium animalis within samples of the study by Nielsen et al. (2014). In summary, GATTACA enables researchers to accurately cluster genomic sequences associated with individual bacteria at a level of accuracy that was possible only using a large cohort of samples and substantial computing resources. Thus, GATTACA makes accurate metagenomic binning accessible to a much wider range of researchers.

1749F

The missing parent problem: Attempting parental reconstruction through sibling analysis. *B.N. Pusey, K.R. Chao, M. Warburton, C.C. Lau, W.A. Gahl, D.A. Adams, T.C. Markello.* Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD.

Since 2009, the NIH Undiagnosed Diseases Program (UDP) has been routinely carrying out genome scale genotyping and exome, sequencing for its patients and their nuclear families, and to date, it has sequenced and analyzed over 400 families. However, for more than 100 of these families data was not available for one or both parents, rendering the analysis for these families exponentially difficult, if not impossible. While these alternative families make up 25% of the UDP cases, they account for less than 5% of the diagnosed cases. It can therefore be surmised that missing parents result in reduced variant filtration power. The resulting data contains a greater number of false positives, which constrains the practical curation and subsequent diagnostic rate. Using genotype data, a minimum of 3 siblings, and the remaining parent, we can partially reconstruct the sequence of the missing parent, provided all three siblings have the same biological parents. This approach relies on the identification of regions within the siblings that can only have been co-inherited from the missing parent. With each extra available sibling, the portion of the missing parent that can be reconstructed increases asymptotically. The strategy uses two sets of discrete mathematical expressions to define informative loci and designate region of the genome as attributable to the missing parent, excludable from the missing parent, or inconclusive. A BED file defining these regions incorporated into subsequent variant analysis steps. The resultant analysis, while still containing regions of indeterminate parentage, allows for more complete phasing of compound heterozygous variant pairs and an overall reduction of noise during filtration (i.e. suppression of false de novos). This technique has the potential to salvage previously difficult cases by striving toward a maximum Mendelian filtration potential. This research is supported by the Intramural Research Program of the National Human Genome Research Institute and Common Fund of the National Institutes of Health.

1750W

Karp: Accurate and fast classification of microbiome taxonomy using pseudomapping. *M. Reppell, J. Novembre.* Human Genetics, University of Chicago, Chicago, IL.

 The application of next generation sequencing technology to microbiome DNA samples has had a transformative impact on microbial ecology and human health. The ability to uncover associations between microbes, their environment, and host genetics is dependent on the accurate quantification of the organisms present in the sequenced samples. Many methods have been developed to perform this quantification task; few scale to the size and complexity of modern metagenomic data. Methods that scale well often sacrifice accuracy to achieve reasonable performance. Pseudomapping, a rapid k-mer based approach for query sequence matching, provides a novel way to achieve both speed and accuracy in read assignment. Pseudomapping was originally implemented in the software Kallisto and has shown superior accuracy to other classification methods. However, Kallisto ignores sequencing quality information that could help assign multiply-mapped reads more accurately. This led us to develop Karp, a program that leverages the speed and low memory requirements of pseudomapping with an EM algorithm that uses sequencing base quality scores to quickly and accurately classify the taxonomy of pooled microbiome samples. Using simulated 16S reads we tested the accuracy and performance of Karp relative to Kallisto and USEARCH for large datasets (1M reads) with a range of community compositions and read lengths. Karp consistently had higher accuracy than comparison methods. With 75bp single-end reads on average Karp's errors were 11% less than Kallisto and 55% less than USEARCH. With 301bp paired-end reads, Kallisto's strict pseudomapping threshold lowered performance, and on average Karp's errors were 90% smaller than Kallisto's and 58% smaller than USEARCH's. Karp also improved the accuracy of summary statistic estimates, the average error for pairwise beta diversity was 45% and 80% smaller than Kallisto and USEARCH's estimates, respectively. While Kallisto set the benchmark for computational speed, quantifying 1M reads in <5 minutes and 10GB RAM with 12 cores, Karp is computationally very feasible, requiring <90 minutes and 10GB. USEARCH required <75 minutes and 4GB. Modern sequencing provides us with unprecedented access to the microbial world, but its output requires careful interpretation. Quantifying the taxonomy of a pooled sample is difficult, but with the development of Karp we provide a scalable and accurate solution, of value to every researcher performing microbiome sequencing.

1751T

Motif scraper: A tool for efficient sequence motif searches in FASTA files using degenerate nucleotide queries. *E.D.O. Roberson.* Medicine and Genetics, Washington University, St. Louis, MO.

 FASTA files are a common storage format for sequence data, including reference genomes, personal genomes, cDNA sequences, and *de novo* assembly outputs. However, it is not easy to search these files for specific sequence motifs, such as miRNA binding sites, transcription factor sites, and genomic editing sites capatible with different CRISPR-based technologies. I created motif scraper to address this missing utility. The tool is open-source and freely licensed. The code is written in Python for cross-platform compatibility and ease of use. Motif scraper accepts one or more nucleic acid sequence queries, allowing for IUPAC degenerate bases. The text query is converted to a regular expression, substituting the degenerate bases for their appropriate standard bases. Motif scraper then searches for the degenerate motif using the regex module. The algorithm is amenable to multiprocessing, and can automatically spread the word load across the specified number of processors. This allows for scanning of an entire FASTA genome in minutes for any degenerate motif of interest. The utility of a motif search goes beyond the canonical use of searching reference genomes. FASTA formatted files are commonly generated using de novo assembly from both DNA and RNA, and will increasingly become important as personal genomes, and particularly phased personal genomes, are created. This tool allows for efficient search of personal genomes to annotate such features as new transcription factor or miRNA binding sites produced by an individual combination of rare and private variants. These types of changes in particular cannot be determined from available databases and require calculation per-individual.

1752F

Haplotype resolved SV assembly: Producing gold-standard SV calls on diploid genomes using PacBio and 10X sequencing data. *O. Rodriguez[1], F. Sedlazeck[2], M. Pendleton[1], C.T. Watson[3], R. Sebra[1], A.J. Sharp[1], C. Chin[4], M.C. Schatz[2], A. Bashir[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Computer Science, John Hopkins University, Baltimore, MD; 3) Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY; 4) Pacific Biosciences, Menlo Park, CA.

Long read sequencing has dramatically improved our ability to examine large-scale genome architecture. Human de novo assembly using Pacific Biosciences (PacBio) sequencing routinely achieves N50s in the Megabases. Additionally, multiple studies have shown that long-read technologies provide more insights into both the number and the complexity of structural variants (SVs). However, traditional de novo assembly and read-mapping approaches appear unable to exhaustively characterize SVs and disambiguate zygosity. For example, these approaches have been observed to exhibit a high-rate of Mendelian violations in trio analyses. Here, we utilize long-read PacBio sequencing, and long-read sequencing supplemented with barcoded short-reads from 10X genomics, to provide highly contiguous and complete assembly of phased haplotype contigs (haplotigs). Novel algorithms are then applied to these haplotype resolved sequences to substantially improve SV inference. We employ our approach on high-depth sequencing data on EBV-immortalized B lymphoblastoid cell lines from an Askenazi Jewish trio (Coriell samples GM24835, GM24149, GM24143) provided as part of the Genome in a Bottle (GIAB) Consortium. A combination of PacBio-only and hybrid approaches are used to build paired maternal and paternal contigs (haplotigs) containing phased SVs and SNVs; together they enable haplotig contigs and scaffolds with N50s exceeding 100kb and 10Mb, respectively. This method also produces more complete calling of heterozygous, events, especially highly complex events, by disambiguating maternal and paternal variation prior to SV resolution. Fewer Mendelian violations are observed, while the absolute number of SV calls increases within the >70% of the genome that appears haplotype separable. Over 20,000 SVs are observed in the proband alone, nearly a third of the total SVs that have been catalogued in all 2,504 individuals within the Phase 3 1000 Genomes project. With the improved knowledge of where high-quality haplotype separation occurs, we can provide far more robust set of SVs which will improve our understanding of structural variation in the genome and provide a training set further method development. We believe this not only represents a critical genomic resource, but moves the community away from biases induced by overreliance on a single reference genome and even closer to automatic, complete, diploid personal genomes.

1753W

ASPIREdb – An interactive web-based system for exploration, analysis and visualization of complex phenome-genome datasets. *S. Rogic[1], P. Tan[1], A. Zoubarev[1], C. McDonald[1], F. Lui[1], G. Charathsandran[1], M. Jacobson[1], M. Belmadani[1], J. Leong[1], T. Van Rossum[1], E. Portales-Casamar[1], Y. Qiao[2,3], K. Calli[3], X. Liu[4,5], M. Hudson[4], E. Rajcan-Separovic[2], S. Lewis[3], P. Pavlidis[1].* 1) Michael Smith Laboratories, UBC, Vancouver, British Columbia, Canada; 2) Department of Pathology, BC Child and Family Research Institute, UBC, Vancouver, Canada; 3) Department of Medical Genetics, BC Child and Family Research Institute, UBC, Vancouver, Canada; 4) Department of Psychiatry, Queen's University, Kingston, Canada; 5) Ongwanada Resource Cente, Kingston, Canada.

One of the main challenges facing researchers studying complex genetic disorders using next generation sequencing technologies is identification of causal variants. Considering additional patient information and accumulated knowledge from public data repositories is necessary to enable sifting through thousands of predicted variants and identifying candidates for further validation. To address this challenge, we developed ASPIREdb, an open-source, easy to use, interactive, collaborative web application that allows researchers to search, organize, analyze and visualize variants and phenotypes associated with a set of human subjects. The data model supports all main types of genomic variants and phenotypes are recorded using Human Phenotype Ontology, whose controlled vocabulary and hierarchical structure facilitate computational inference and linking to external resources. ASPIREdb features a powerful query builder for filtering subjects and/or variants using genotype and phenotype characteristics in conjunction with data from integrated databases and external sources. It also implements several standard statistical methods for analysis and exploration of data, such as burden analysis and phenotype-based clustering. Projects can be shared, allowing researchers to work collaboratively to build queries, analyze and annotate the data. In addition to table-based view of the data, ASPIREdb provides various graphical representations: selected genomic regions and associated, color-coded variants can be viewed using interactive ideogram; the Phenotype Heatmap tool provides a visual summary of phenotype values for selected subjects phenotypic data; and the Variant Report tool plots the distributions of various variants' attributes. ASPIREdb is linked to UCSC Genome Browser, which allows the user to view selected genomic variants in the context of a large collection of aligned annotation tracks. It is also integrated with other web-based databases developed in our lab to provide information on gene networks and differential gene expression patterns of the genes associated with the variants and phenotypic associations of these genes with diseases and phenotypes. In addition, built-in reference databases of common and disease-associated variants, such as DGV and DECIPHER, can be used for variant annotation and filtering.

1754T

Identification of causal genes for rare genetic disorders using whole genome and whole exome sequencing. *Y. Shen[1], A. He[1], W. Zhang[1], N. Thiessen[1], Y. Ma[1], A. Mungall[1], R.A. Moore[1], W. Gibson[1,2], M.A. Marra[1,2], S.J.M Jones[1,2].* 1) Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

There are over 7000 types of rare diseases known, with many more as-yet undescribed diseases thought to exist. Although individually rare, collectively rare diseases affect 10% of the North American population, and most patients present in childhood. These disorders manifest a variety of medical issues including birth defects, intellectual disability, growth abnormality, and organ failure. It is estimated that approximately 80% of rare disorders have a strong genetic basis, but the majority of their causal genes are unknown. Finding the causal genes has the potential to aid in diagnosis and treatment, as well as to provide an understanding of the molecular basis of the disease. Here we describe a bioinformatics platform that analyses genome and exome sequencing data to identify causal genes in rare genetic disorders. We designed and built a bioinformatics pipeline that integrates a variety of tools to detect small mutations and indels, copy number alterations, and structural variations. Identified aberrations are then filtered against public and in-house databases to detect rare genomic events, annotated with potential functional impact and pathogenicity, and ranked based on the population frequency and biological relevance. Using this pipeline, we have identified multiple genes causing or related to rare genetic disorders, such as *EZH2* in Weaver Syndrome, *EED* in a novel syndrome similar to Weaver syndrome, *ATP1A3* in CAPOS (Cerebellar ataxia, Areflexia, Pes cavus, Optic atrophy and Sensorineural hearing loss) syndrome, and *SGOL1* in CAID (Chronic Atrial and Intestinal Dysrhythmia) syndrome. Our study design also enables collection and comparison of sequencing data from patients with unusual phenotypes, such that updated analyses comparing undiagnosed patients with phenotypes similar to each other can be done.

1755F

High frequency of transcriptomics studies with sample mix-ups. *L. Toker[1,2], M. Feng[1,2,3], P. Pavlidis[1,2].* 1) Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada; 2) Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada; 3) Graduate Program in Genome Sciences and Technology, University of British Columbia, Vancouver, BC, Canada.

Concern about the reproducibility and reliability of biomedical research has been rising. An understudied issue is the prevalence of sample mislabeling, one impact of which would be invalid comparisons. We studied this issue in a corpus of human transcriptomics studies by comparing the provided annotations of sex to the expression levels of sex-specific genes. We identified apparent mislabeled samples in 46% of the datasets studied, yielding a 99% confidence lower-bound estimate for all studies of 33%. In a separate analysis of a set of datasets concerning a single cohort of subjects, 2/4 had mislabeled samples, indicating laboratory mix-ups rather than data recording errors. While the number of mixed-up samples per study was generally small, because our method can only identify a subset of potential mix-ups, our estimate is conservative for the breadth of the problem. Our findings emphasize the need for more stringent sample tracking, and that re-users of published data must be alert to the possibility of annotation and labeling errors.

1756W

Database for preeclampsia v2.0: Network visualizations and variants. *A. Uzun[1,2], D. Armanious[3], K. Deragon[1], J. Schuster[1,2], J. Padbury[1,2].* 1) Pediatrics, Women and Infants Hospital, Providence, RI; 2) Warren Alpert Medical School, Brown University, Providence, RI; 3) Computer Science, Brown University, Providence, RI.

Preeclampsia is a multi-system hypertensive disorder of pregnancy, characterized by variable degrees of maternal symptoms including elevated blood pressure, proteinuria and fetal growth retardation that affect 2-8 % of deliveries. It is a complex disease and one of the most common causes of fetal and maternal morbidity and mortality worldwide. We built a Database for Preeclampsia (dbPEC) including curated articles, clinical features, genes, and preeclampsia phenotypes. The first version of dbPEC was released on October 2015 and included 899 articles and 602 genes. It is publicly available (http://ptbdb.cs.brown.edu/dbpec). We are now releasing an updated version (v2.0) where users can visualize the interaction between the genes in our database as network maps. We extracted the data from BioGRID (version 3.4.136), a database of physical and genetic interactions, and built a visualization tool implemented using a combination of server-side Java 8 and PHP, along with client-side HTML5 and JavaScript. The Java backend uses a modified force-directed layout algorithm that performs a fine-grained layout for the connected components of the graph and a coarse-grained layout on the entire graph. On the client side, the JavaScript displays the graph on an HTML5 Canvas element and makes it interactive by handling various input events. In this updated version, we also include information on more than 100 SNPs from the curated papers in which genes were experimentally validated. By recent updates and expansions, we believe dbPEC will continue to serve as not only a preeclampsia resource but also will provide useful datasets for analysis of high-throughput experiments and genome-wide association studies.

1757T

A genome and phenome wide landscape of diagnoses in Geisinger Health System for 38,662 individuals. *A. Verma[1,3], S.S. Verma[1,3], A. Lucas[1], D.N. Hartzel[1], D.R. Lavage[1], J. Leader[1], H.L. Kirchner[1], M.D. Ritchie[1,2], S.A. Pendergrass[1].* 1) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 2) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 3) The Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA.

Genome-wide genetic studies typically focus on a genetic map of one single disease or a limited phenotype domain, whereas Phenome-Wide Association Studies (PheWAS) provide a comprehensive approach for investigating the association between genetic variants, diseases and quantitative traits. Most PheWAS to date have used a small to moderate number of SNPs for association with phenotypic data. For this study we performed a comprehensive genome-wide PheWAS in DicovEHR cohort, investigating the association between 632,574 genetic variants with minor allele frequency > 1% from 38,662 consented unrelated adult European American participants in the Geisinger Health System MyCode® biorepository and 541 ICD-9 derived case-control diagnoses, to gain insights on the genome-wide landscape of associations. We compared our results to a database of all other existing published GWAS and PheWAS results to date. The most significant association for our PheWAS was between SNP rs9273363 in *HLA-DQB1* and type I diabetes (ICD-9 250.01, p-value = 1.36 x 10^{-77}) replicating previous associations with type 1 diabetes. We also found associations with type 2 diabetes, hypothyroidism, and morbid obesity replicating previously reported associations with similar traits. We identified a series of novel associations, including a missense variant, *HLA-DRB5* SNP rs701884 associated with multiple sclerosis (ICD-9 340, p-value = 5.28x10^{-23}) and *PIGU* intronic SNP rs17305573 associated with actinic keratosis (ICD-9 702.0, p-value = 6.38x10^{-12}). Other possible novel PheWAS results include highly significant genetic associations (p-value < 1x10^{-8}) with phenotypes such as vitamin D deficiency (rs1037378, ICD-9 268.9), acute bronchitis (rs1868256, ICD-9 466.0), mitral valve disorder (rs4852126, ICD-9 424.0), macular puckering (rs12634684, 362.56), and hyperpotassemia (rs983574, 276.7). We are also evaluating our PheWAS results using regulatory and non-protein-coding information from GENCODE and ENCODE. The sweeping comprehensive nature of this PheWAS allows for novel hypothesis generation, the identification of phenotypes for further study and phenotypic algorithm development, identification of cross-phenotype associations, multiple phenotype pathway based analyses, and the identification of the networks existing between phenotypes, genes, and genetic variation.

1758F

Benchmarking distributed data warehouse solutions for storing genomic variant information. *M.S. Wiewiórka[1], D.P. Wysakowicz[1], M.J. Okoniewski[2], T. Gambin[1].* 1) Institute of Computer Science, Warsaw University of Technology, Warsaw, Mazowieckie, Poland; 2) Scientific IT Services, ETH Zurich, Zurich, Switzerland.

Genomic-based personalised medicine incorporates storing, analysing and interpreting of genomic variants as its central issues. When thousands of patients sequenced exomes and genomes are becoming available, there is the growing need for the efficient database storage and querying. The answer may be application of modern distributed datastores and query engines. Still, their properties for large genomic variant databases have not been explored yet. To investigate the utility of modern columnar data stores and distributed querying engines., we have constructed a prototypic large genomic variant database, which has been built according to data warehousing paradigms adjusted to current massively parallel computing architectures using robust big data technologies . Proposed solution consisted of centralized Hive Metastore providing an abstraction layer over pluggable storage backends (e.g. HDFS with different columnar file formats - ORC, Parquet or Kudu storage system). Various query engines like Hive, SparkSQL, Presto or Impala have been evaluated. Finally application of distributed OLAP cube system - Kylin have been verified. To construct a benchmark comparison of technical solutions, the database has been loaded with simulated variant data generated for 50,000 individuals using the allele frequency information from ExAC database. Then in order to enable its time efficiency comparisons, a set of queries both requiring efficient sequential scans and random access patterns (range queries) has been run. The queries are modeling the ad-hoc approach to finding information in the data warehouse, which may be essential for both: research and medical use of the genomic variant databases. The result is a multi-sided insight into the utility of genomic data warehouse systems. It is also the proof that genomic data warehouses can be practically applied in the personalised medicine, once various technical issues related to their design, scaling and query optimization can be solved. The quantitative benchmarking results give the detailed and practical recommendations on using particular formats, querying engines, database structures and query optimization. The results of the research indicate clearly that Kudu may outperform other storage systems. Studies present also recommendations for choosing apropriate distributed query engines in specific use cases. The detailed description of the project can be found at https://github.com/ZSI-Bio/variantsdwh

1759W

Algorithmically optimized gene selection for genomic disease studies. *E. Williams, Y. Tian, Y. Zhu, C. Munroe, C. Shtir.* Thermo Fisher Scientific, Carlsbad, CA.

Selection of genes to include in genomic studies of disease remains a difficult task. Current methods rely on expert opinion or manual search engine use. With these methods, the process and result are neither repeatable nor scalable. To remedy this situation, we created the Informative Genetic Content (IGC) system, which enables the algorithmic selection of genes for inclusion in such studies, given one or more diseases to target.   The IGC system stands on three components: a database associating diseases with genes and other diseases, an algorithm to rank the genes under consideration for inclusion in a panel, and a module that clusters genes by families of diseases.   The first component, the database, maps diseases to associated genes and scores each of these mappings according to the strength of the relationship. The database also maps diseases to other diseases, such that groups of diseases or hierarchical relationships between diseases can be identified.   The second component enables the ranking of candidate genes when multiple diseases are of interest. The algorithm accounts for the common situation where two or more diseases are associated with the same gene with varying strengths of association, weighting and combining the scores across the diseases associated with each gene.   The final component, the gene clustering module, groups genes by pathogenic pathways, should the user want to consider targeting a broader family of diseases affected by a closely related set of genes.   We validated the IGC system through comparisons of our automated gene selections with expertly curated gene panel designs. We found a high degree of overlap between the IGC's gene selection and the gene lists chosen by experts, supporting the viability of our system. Together with the scalability and repeatability enabled by its automation, the IGC system greatly improves the gene panel selection process and therefore advances targeted genomic studies.

1760T

Accurate and robust estimation of DNA contamination against mis-specified population allele frequencies. *F. Zhang[1], G. Abecasis[1,2], H.M. Kang[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI,USA.

Detecting and estimating inter-sample DNA contamination became a crucial quality assessment step to ensure high quality sequence reads and reliable downstream analysis. Our previous method *verifyBamID*, which provides DNA contamination estimates only from sequence reads and population allele frequency at selected variants, has become a part of standard sequence processing pipeline in many large sequencing centers, effectively flagging largely contaminated samples (e.g >3%). However, it has been known that the DNA contamination from *verifyBamID* could be underestimated if the population allele frequencies are mis-specified.   Small amount of contamination (e.g. 1-3%) could also affect the quality of variant calling and genotype accuracy of sequenced genomes, and it becomes more crucial for deeply sequenced genomes which is expected to provide extremely high genotype accuracy. Accurate estimation DNA contamination levels will allow us to more accurately assess the quality of sequenced genomes to determine objective QC criteria, and also will enable us to correct for the contamination in genotype calling. However, it is cumbersome and imprecise to manually annotate the genetic ancestry of every sample prior to sequencing,   Here we describe a novel method to robustly detect and estimate DNA contamination by modeling sequence reads by individual-specific allele frequencies. Instead of starting from a single population allele frequency, our method reference takes the singular value decomposition of genotypes from a diverse reference population (e.g. HGDP or 1000 Genomes), and jointly estimates principal component coordinates and contamination levels of a sequence individuals together based on a mixture model. The key idea is to precisely model individual-specific allele frequencies from principal component coordinates and estimate DNA contamination based on the allele frequencies iteratively.   We applied our method to 1000 Genomes datasets by simulating contamination levels from 1% to 20% and comparing the contamination estimates obtain from different methods. When using pooled allele frequencies, as opposed to population-specific allele frequencies, we observed that the contamination levels are underestimated by 18%, 35%, 41%, and 70% for CEU, YRI, FIN, and CHS populations, respectively. Using our new method, the underestimation bias was reduced to 1-3%. Our method is implemented in the rapid sequence quality assessment software tool FASTQuick.

1761F

Leveraging an advanced knowledge base of biological pathways and network analytics to identify disease-causing mutations from clinical genome and exome sequence data with increased efficiency and accuracy. *S. Shah[1], A. Krämer[1], S. Tang[1], A. Muthiah[1], T. Vvuong[1], G. Eley[2], B. Solomon[2], A. Khromykh[2], A. Joecker[1], R. Felciano[1], D. Richards[1].* 1) QIAGEN Bioinformatics, Redwood City, CA; 2) Inova Translational Medicine Institute, Falls Church, VA.

Identifying genetic variants underlying rare inherited diseases from Clinical Genome and Exome Sequence (CGES) data can be both challenging and time consuming, with a significant amount of time and resources invested in variant calling, annotation, and interpretation workflows. Accurate determination of true disease-causing variants requires thorough technical validation as well as in-depth assessment of both variant- and gene-level evidence supporting for causation. Here we present a Hereditary Disease and Cancer variant calling and interpretation solution that enables highly efficient identification of disease causing variants with superior accuracy and sensitivity. The high performance is achieved with a novel, streamlined, one step, end-to-end workflows that are able to identify disease causing mutations in hereditary and rare diseases as well as driver variants in cancer. All workflows were optimized for specific exome and genome applications in hereditary diseases or cancer and use large-scale causal network analytics derived from the QIAGEN Knowledge Base. The Knowledge Base is a large structured collection of observations in various experimental contexts with over 11 million integrated findings that are manually curated from the biomedical literature and third-party databases. Pathways and networks constructed from these findings that include mutations, biological interactions, and functional relationships, enable users to discover and prioritize the most interesting and relevant hypothesis and quickly visualize the causal and regulatory networks most closely associated with the disease or phenotype of interest. Additionally, we have developed and integrated a scoring method into our workflow to rank variants, using syndrome inference from phenotypes provided to accurately identify known or novel variants in disease causing gene(s). To showcase these features and capabilities, we will present the results from whole genome sequence data analysis and interpretation from 30 trios, with a child afflicted with severe congenital abnormalities, for which the causal variant was previously known. In 5 out of 6 of the cases analyzed so far, we were able to successfully reduce the number of candidate variants for follow-up to 5 or less, including the previously identified mutation.

1762W

Candidate gene prioritization with Endeavour. *A. Ardeshirdavani*[2,3]*, L. Tranchevent*[1]*, S. ElShal*[2,3]*, D. Alcaide*[2,3]*, A. Aerts*[2,3]*, D. Auboeuf*[1]*, Y. Moreau*[2,3]. 1) INSERM U1210, CNRS UMR5239, Laboratoire de Biologie et de Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, Université de Lyon, Lyon, France; 2) Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics Department, KU Leuven, Leuven, Belgium; 3) iMinds Future Health Department, KU Leuven, Leuven, Belgium.

Genomic studies and high-throughput experiments often produce large lists of candidate genes among which only a few are truly relevant to the disease, phenotype, or biological process of interest. Gene prioritization tackles this problem by ranking candidate genes by profiling candidates across multiple genomic data sources and integrating this heterogenous information into a global ranking. We describe an extended version of our gene prioritization method, Endeavour, now available for 6 species and integrating 75 data sources. Users can prepare a prioritization in four simple stages. In the first stage, the species with which to work is selected among the six available species (H. sapiens, M. musculus, R. norvegicus, D. melanogaster, C. elegans, and D. rerio). Then, the set of seed genes is prepared in the second stage. These are the genes that are already associated with the process of interest. The third stage consists in selecting the data sources to use for the prioritization. Finally, the set of candidate genes is defined in the last stage. The hypothesis is that one of these candidate genes is associated with the biological process of interest, but this association is yet unknown. The single output is an ordered list of the candidate genes with on top the most promising candidate genes. In an ideal situation, the gene yet unknown associated with the process of interest should be ranked first or at least among the first genes. Validation of our results indicate that this extended version of Endeavour efficiently prioritizes candidate genes. The Endeavour web server is freely available at *https://endeavour.esat.kuleuven.be/*.

1763T

An approach for integrated gene and variant prioritization. *A. Chhibber, Y. Fu, N. Bani Asadi, H.Y.K. Lam.* Roche Sequencing, Belmont, CA 94002.

Exome sequencing has become a key tool both in clinical practice and translational research for disease diagnosis and as a tool for understanding disease biology. However, identifying causal variants for a particular disease out of the tens of thousands of variants in a typical human exome continues to be a challenge. Methods that rely on variant filters on features such as population allele frequency, conservation, and functional effect can miss variants that fall below conservative thresholds. Variants can be ranked by predicted pathogenicity using a number of different algorithms, but as every individual carries hundreds of potentially pathogenic variants, identifying those events relevant to a particular phenotype remains a challenge. In this study, we highlight the value of incorporating knowledge about the particular disease or phenotype under investigation when prioritizing pathogenic variants in an exome, and present a novel approach for disease-specific gene and variant prioritization. Network-based approaches have been widely applied to the prioritization of genes utilizing human protein-protein interaction (PPI) networks; in particular, methods that employ a random walk along PPI networks to prioritize genes have been very effective. Our approach extends these methods in a manner similar to that used by the Topic-Sensitive PageRank approach in the web search community; each "topic" is a different set of potential disease genes, drawn from phenotype or gene ontology terms specified by the user, and if available, known genes associated with the disease and additional genes related to those genes (for example, genes in the same family, with the same protein domain, in the same pathways, regulatory relationships, etc.). Topics are weighted by the information content of each set and combined into a single gene-based score. These gene-based scores are then combined with variant scores representing predicted variant pathogenicity, giving a single ranked variant list per exome that takes into account both the likelihood the variants disrupt their protein functions as well as the likelihood the gene is involved in the disease under study. We demonstrate that incorporating information about the disease context can aid in variant prioritization.

1764F

The ENCODE analysis pipelines: Repeatable and shareable analysis tools for ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments. *J.S. Strattan[1], T.R. Dreszer[1], B.C. Hitz[1], U. Baymuradov[1], E.T. Chan[1], J.M. Davidson[1], I. Gabdank[1], J. Hilton[1], C.A. Sloan[1], F. Tanaka[1], C. Thomas[1], Z. Weng[2], A. Kundaje[1], J.M. Cherry[1].* 1) Department of Genetics, Stanford University Medical School, Palo Alto, CA; 2) University of Massachusetts Medical School, Program in Bioinformatics and Integrative Biology, Worcester, MA.

The comparability of experiments often depends on comparable experiment protocols. This applies equally to the physical manipulation of samples and to the analytical methods that transform raw data to results. Fourteen ENCODE labs have contributed over 4000 replicated ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments on nearly 200 cell types. To ensure that the results from these experiments can be compared, the ENCODE Data Analysis Center (DAC) have specified common data processing protocols for ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments. The Data Coordination Center (DCC) is implementing these specifications as processing pipelines, deploying these pipelines to a cloud-based platform, processing all ENCODE ChIP-seq datasets, and making the pipelines available for anyone to use. The results of these analyses, and metadata describing them, are distributed through the ENCODE Portal and illustrate general methods of accessing and interpreting ENCODE data. The cloud-based deployment of the ENCODE analysis pipelines illustrates how the DCC is creating transparent and reusable analysis tools for ENCODE data and for any primary data from experiments performed with similar protocols. The ENCODE Portal is https://www.encodeproject.org/. The DCC codebase is freely available at https://github.com/ENCODE-DCC/. .

1765W

Visualizing human genome in time and space. *W. Zheng[1], J. Zhao[1], L. Zhou[2], H. Xu[1], J. Tang[2,3].* 1) School of Biomedical Informatics, U Texas Health Science Center at Houston, Houston, TX; 2) Department of Computer Science and Engnieering, University of South Carolina, SC 29205; 3) Key Laboratory of Systems Bioengineering of the Ministry of Educatioon, Tianjin University, Tianjin, 300072, PR China.

The high order nuclear organization of mammalian genome plays significant roles in important cellular functions such as gene regulation and cell state determination. The influx of new details about the higher-level structure and dynamics of the genome requires new techniques to model, visualize and analyze the full extent of genomic information in three dimensions. While existing genome browsers have been proven as successful genome information management and visualization tools, these browsers are based on two-dimensional visual interface with limited capacity to represent structural hierarchies and long-range chromosome interactions, particularly across non-contiguous genomic segments.   We created the first model-view framework of eukaryotic genomes, *Genome3D* (http://genome3d.org), to enable integration and visualization of genomic and epigenomic data in a three-dimensional space. Our physical genome model implicitly contains all levels of structure and hierarchy, and provides an underlying platform for integrating multi-scale genomic information within three dimensions. Our viewer uses a hierarchical model of the relative positions of all nucleotide atoms in the cell nucleus. We further developed a game engine based Genome3D browser that has better performance, is platform independent and can be configured to allow users to access and visualize 3D genome models on a remote server.   In this works, we integrated various genomic databases with Genome3D software, providing a multi-scale genome information visualization system to explore and navigate eukaryotic genome in 3-dimension. The new system, iGenome3D, provides a wide spectrum of tools, ranging from model construction to spatial analysis, to decipher the relationships between 3D conformation of the genome and its functional implication. The incorporation of literature allows users to quickly identify key features from PubMed abstracts for genes in the displayed 3D genome structure. The seamless integration of UCSC Genome Browser allows genetic and epigenetic features from the 2D browser to be visualized in 3D genome structure. iGenome3D can also output 3D genome model in various forms, including one that allows these models to be explored in modern virtual reality environments such as Oculus. Eukaryotic genomes can be analyzed from a completely new angle in iGenome3D that enables researchers to make new discoveries from a truly multi-scale exploration.

1766T

SJ-Variant: An interactive genomic web portal for clinical variant annotation and visual interpretation. *A. Patel, E. Rampersaud, M. Rusch, D. Hedges, X. Zhou, S. Newman, M.R. Wilkinson, C. Pepper, S. Rice, J. Becksfort, J. Downing, J. Zhang.* St Jude Children's Research Hospital, Memphis, TN.

Clinical classification of genetic variants requires the assimilation of disparate data coupled with detailed review by an expert panel. Delivering this information to diverse audiences of clinicians, research scientists, and counselors, in an informative yet accessible format can also be challenging. To address this challenge, we developed SJ-Variant (https://pecan.stjude.org/pecan_variant_public.html), an interactive web-portal that facilitates clinical panel review and semi-automated variant classification based on ACMG-AMP (2015) guidelines. SJ-Variant houses integrated genomics data, including variant population frequency, recurrence in tumor, and functional predictions, along with additional automated and manually curated annotations. Currently, we are using SJ-Variant for variant classification at St. Jude Children's Research (SJCRH) for cancer-related genes, and in the **C**linical **Re**search in **A**LS and Related Disorders for **Th**erapeutic Development (**CReATe**) Consortium for ALS-associated genes. These classifications, along with user-definable evidence and reference information, can serve as an important resource for researchers and clinicians who themselves may not have the time or expertise to carry out the clinical panel review process. A novel feature of SJ-Variant is the incorporation of ProteinPaint, a visualization tool we recently developed for mapping variants onto isoform-specific protein diagrams (Zhou X. Nature Genetics.2016). SJ-Variant allows manually curated aspects of variant data aggregation, such as publication citations and interpretations to be captured in a structured manner for future retrieval and re-use. Currently, SJ-Variant hosts 800 germline variants with pathogenicity classification determined in a study of 1,120 pediatric cancer patients (Zhang J., NEJM 2015) and a catalogue of >100 germline variants from published ALS research and clinical studies. SJ-Variant will provide the clinical genomics community with an intuitive and easy-to-use interactive portal for tracking decisions on variant classification which may be useful when providing clinical recommendations to patients regarding the importance of known genetic variants. SJ-Variant improves the accessibility of genomic information to the biomedical community in an organized and non-technical manner.

1767F

RolyPoly: A fast and accurate method for identifying trait-relevant cell types using GWAS summary statistics and single-cell gene expression. *D. Calderon[1], D. Golan[2,3], T. Raj[8,9], A. Bhaskar[2], D. Knowles[6], A. Fu[7], J. Pritchard[2,4,5].* 1) Biomedical Informatics, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Statistics, Stanford University, Stanford, CA; 4) Department of Biology, Stanford University, Stanford, CA; 5) Howard Hughes Medical Institute, Stanford University, Stanford, CA; 6) Department of Radiology, Stanford University, Stanford, CA; 7) Department of Statistical Science, University of Idaho, Moscow, ID; 8) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 9) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

The identification of relevant cell types for complex traits is important for prioritizing costly, labor-intensive follow up studies. Genome-wide association studies (GWAS) summary statistics implicitly contain information pointing to the tissues and cell types relevant to the trait. However, to our knowledge, there is no polygenic and unbiased method to identify trait-relevant cell types using gene expression information. Here we describe a polygenic method RolyPoly, which connects gene expression and GWAS summary statistics to prioritize cell types for a trait. RolyPoly inference uses a novel method of moments estimator relying on a polygenic GWAS gene score and takes advantage of asymptotic composite likelihood theory for producing calibrated standard errors in order to accurately prioritize important cell types while minimizing false positive errors. We demonstrate the robustness of our method through simulation and validate previously known tissue-trait associations using the GTEx gene expression data from 27 tissues and GWAS summary statistics for over 15 complex traits. Moreover, using single-cell gene expression data from the brain, we discover newly linked pathogenic cell types, including a significant enrichment for myeloid cell types such as microglia ($p < 5e\text{-}7$) for Alzheimer's disease, and replicating fetal cortical cells ($p < 5e\text{-}5$) for Schizophrenia. Our approach represents a powerful framework for understanding the effect of common variants on cell types or cell states, and a useful tool for prioritizing pathogenic cell types for functional follow-up studies. We implement this method in the R package rolypoly.

1768W

Partitioning heritability of complex human disease by minor allele frequency. *K. Hartman[1], R. Hernandez[2].* 1) Biological & Medical Informatics, University California, San Francisco, San Francisco, CA; 2) Bioengineering and Therapeutic Sciences, University California, San Francisco, San Francisco, CA.

   In recent years the genetic heritability of traits have been broken down into the contributions from different parts of the genome, in a technique known as partitioning of heritability. Rather than trying to determine effects of individual loci, this approach seeks the contribution in aggregate of many single-nucleotide polymorphisms (SNPs). This has previously revealed certain regions of the genome contributing disproportionately more heritability per SNP, such as sites of DNase-hypersensitivity and highly conserved non-coding elements. However, the many of regions used in previous studies differ from the background genome in terms of minor allele frequency distribution. This difference could dramatically affect the estimates of heritability coming from these regions both due to bias in the estimation procedure and the etiology of the diseases studied. This work will investigate the relationship between heritability and minor allele frequency using several previously published data sets and through genotypic and phenotypic simulation. Preliminary results show that bins containing fewer SNPs and lower frequency SNPs are subject to greater variance in their estimates of heritability.

1769T

Weighted-likelihood inference of genomic autozygosity in The 1000 Genomes Project Phase III. *A. Blant[1], M. Kwong[1], Z. Szpiech[2], T. Pemberton[1].* 1) University of Manitoba, Winnipeg, Manitoba, Canada; 2) University of California, San Francisco, United States.

   Genomic regions of autozygosity (ROA) reflect homozygosity for haplotypes inherited identical-by-descent from an ancestor common to both parents. Long ROA most likely arise from a recent ancestor; shorter ones from a more distant ancestor. Their length distributions in contemporary populations provide invaluable insight into how population history and sociogenetic practices influence genomic variation patterns. Further, while ROA have underlain the identification of recessive Mendelian disease genes for over two decades, they are now gaining importance in complex diseases studies for the identification of recessive trait variants. We previously reported an ROA detection method that uses a logarithm-of-the-odds (*LOD*) score of autozygosity in a sliding-window framework. This method is sensitive to the effects of linkage disequilibrium (LD) and recent recombination and mutation events onto similar haplotype backgrounds that give the incorrect appearance of autozygosity leading to false-positive ROA detection. In this study, we report an improved *LOD*-based ROA detection method that incorporates population allele frequencies and per-genotype error rates as well as controls for LD and the probabilities of recombination and mutation within a window to provide improved specificity in ROA detection. Using whole-genome (WGS) and whole-exome (WES) sequencing and Illumina BeadChip data available for Phase 3 of The 1000 Genomes Project, we show that ROA identification with *wLOD* becomes more robust with increasing SNP density. While appreciable variability in short ROA is observed across the different data types, intermediate and long ROA are instead highly consistent. Intriguingly, in addition to distinguishing autozygous windows from non-autozygous windows, *wLOD* also appears able to distinguish ROA due to consanguinity from ROA due to endogamy. We replicate the observation that long ROA are enriched for deleterious variation carried in homozygous form, and extend this to show that the degree of enrichment is directly proportional to the predicted strength of deleteriousness. Finally, by directly comparing per-window *wLOD* scores between pairs of populations we identify genomic regions enriched for autozygosity signals in each of the 26 populations. These findings provide additional insights into the role of population history and cultural processes in the evolution and maintenance of recessive disease loci and provide a new tool that can be used to identify them.

1770F

Quantification of sensitive information leakage from genomic linking attacks. *A.O. Harmanci[1,2], M. Gerstein[1,2,3].* 1) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 2) Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 3) Department of Computer Science, Yale University, New Haven, CT.

Genomic privacy has recently gained much attention. Most studies on privacy focus on identification of scientific study participants by protection of genetic variants. The phenotypic data is generally assumed to be free of individual identifying information. However, an adversary can use phenotypic information and predict genotypic information utilizing the publicly available datasets that contain phenotype-genotype correlations, e.g. quantitative trait loci (QTL). The predicted genotypes can be cross-referenced with genotype datasets and be linked to the phenotype datasets. These links can then be used to reveal sensitive information, such as disease status, that is stored within genotype/phenotype datasets. The availability of very high dimensional phenotype datasets, such as gene expressions, increases the likelihood of a successful linking attack. We present methods for quantifying leakage of information that can be used to characterize individuals. We also present a three-step protocol for instantiation of linking attacks. This enables systematic analysis of these attacks. We finally present an instance of these attacks that utilizes a simple yet highly effective genotype prediction that uses extremes of gene expressions. We finally study the effectiveness of this attack under different scenarios.

1771W

Using polygenic approaches to identify novel associations in existing datasets: Examples from HIV host genetic studies. *R.A. Power[1], C.W. Thorball[2], I. Bartha[2], T. de Oliveira[1], P.J. McLaren[3], J. Fellat[2].* 1) Africa Centre for Population Health, University College London, Durban, South Africa; 2) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland; 3) JC Wilt Infectious Diseases Research Centre, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg Canada.

Polygenic approaches to GWAS have been successful in confirming the heritability of complex traits and in understanding aetiology. Here we highlight their potential contribution to the identification of novel risk variants for complex traits in two ways: 1) by identifying homogenous subgroups within a phenotype, allowing a more informed SNP association analysis; and 2) by highlighting SNPs of pleiotropic effect for genetically overlapping phenotypes. Using genome-wide data from 6,315 HIV+ individuals part of the International Collaboration for the Genomics of HIV (ICGH), we first analysed heterogeneity in set point viral load (spVL) – a quantitative trait correlated with disease progression. GWAS of spVL have solely identified variants within the major histocompatibility (MHC). Using GCTA-GREML analysis we showed that while variation in spVL below the median spVL was more heritable compared to variation above the median (h^2=23% vs. 10%), the MHC almost exclusively controlled spVL variation below the median (h^2 excluding MHC: 3% vs. 8%). By restricting to only those individuals with above median spVL, GWAS identified a novel SNP outside of the MHC (rs11555096) at genome-wide significance (p= 2.43e-8). We then used LD Score Regression to test for phenotypes overlapping with HIV susceptibility, a binary trait associated with homozygosity for a 32 bp deletion in *CCR5*. By looking at 25 phenotypes with publically available GWAS summary statistics, we found two with significant genetic overlap: schizophrenia (rG=0.17, p=0.0026) and ulcerative colitis (rG=0.24, p=0.0057). BUMHBOX analyses suggested genetic overlap in both instances was not driven by heterogeneity within the HIV population but by genome-wide pleiotropy. Both these diseases have >100 associated loci, and we tested if these variants significantly predicted HIV acquisition. Three SNPs were significantly associated after correction for multiple testing. Two of these were eQTLs for genes for proteins known to be related to HIV biology: rs1819333 in *CCR6* (p=0.0002) and rs4932178 in *FURIN* (p=0.0003). While these results are preliminary and require further replication, they highlight the potential of polygenic methods to help in the identification of associations with individual variants. Such methods are diverse but crucially often leverage existing, often publically available, data. This makes them cost effective means for gaining new insights into the genomic architecture of complex traits.

1772T

High-resolution relationship inference via multi-way identical by descent sharing. *M.D. Ramstetter[1], J. Blangero[2], A.L. Williams[1].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX.

Inferring relatedness among study samples is required to correctly perform genome-wide association studies, linkage analysis, and population genetic studies, but existing methods utilize only pairwise relatedness measures and need improvement. Current pairwise relatedness approaches perform well for close (e.g., first degree) relatives, but their accuracy drops as the relationship between two samples grows more distant. Due to the stochastic nature of Mendelian inheritance and recombination, individuals with the same pedigree-based relationship to a given individual vary in the amount of DNA shared with that person, leading to variability in the inferred relationship. This feature—while potentially confounding to traditional approaches—can be exploited by combining identical by descent (IBD) sharing information among sets of close relatives. We propose a new method that extracts and combines IBD sharing information from sets of closely-related individuals to provide more accurate relationship inference between these individuals and their shared relatives. The approach works by first computing pairwise IBD sharing information using Refined IBD and then identifying a confident set of closely-related individuals (e.g., full siblings, avuncular pairs, etc.). Next, we locate individuals that share IBD segments with these close relatives, and we infer relatedness between the close relatives and each such individual. The inference works by implicitly reconstructing the IBD sharing profile between the close relatives' (ungenotyped) common ancestor and each of the more distantly related samples. Assuming that the close relatives' IBD segments all descend from one common ancestor, we can estimate the IBD sharing profile of that ancestor by forming the union of the IBD segments among the close relatives. Descent from one ancestor can be tested by ensuring that all overlapping IBD segments have the same underlying haplotypes. Our analyses show that substantial accuracy improvements are possible via reconstruction of an ancestors' IBD sharing profile. For example, the relationship between sets of siblings and their fifth degree relatives (e.g., second cousins) can be estimated with 67.6% accuracy; by contrast, estimation using the parent of a set of siblings gives an accuracy of 85.3%. We present accuracy results using real data for relationships among 2,485 San Antonio Family Studies individuals that includes hundreds of close and distant relatives.

1773F

Impact of annotation parameter choice on exome sequence analysis pipeline output for application to metabolic disorders in newborn screening. *Y. Wang, A.N. Adhikari, S.E. Brenner.* University of California, Berkeley, CA.

Public health newborn screening (NBS) for rare metabolic disorders uses mass spectrometry (MS/MS). Identifying disease-causing variants from sequencing presents a promising opportunity for early detection and diagnosis of disorders, to complement or perhaps eventually supersede MS/MS for newborn screening. Although the efficiency of genomic data acquisition has increased dramatically, accurate identification of pathogenic variants in the human genome remains a challenge. It is not immediately obvious whether protocols designed for diagnostic clinical exomes can be readily applied for population-level screening, as these applications are quite different. Diagnostic sequencing is performed on patients likely affected by genetic disease and typically has a yield below 50%. By contrast, NBS must identify ~99% of affected individuals, yet must also have >99.5% specificity since ~99.97% of newborns are unaffected by metabolic disorders. Furthermore most clinical exome sequencing uses trios and involves multiple layers of manual review and clinical expertise, while NBS has access to individuals only, and must be overwhelmingly automated due to scale. The primary outcome of a screening pipeline is whether an individual is reported to "screen positive" or "screen negative" for any relevant disorder. Most current variant interpretation methods employ pipelines with multiple annotation features, but there is limited systematic knowledge of which parameters are more sensitive to alterations. We generated a battery of interpretation (>90) pipelines by perturbing different parameters commonly used in clinical sequencing, including the choice of variant callers, phasing methods, MAF thresholds, population reference and disease databases, pathogenicity predictors, and inheritance models. We evaluated which individuals and what fraction of the population screened positive from each pipeline (without knowing which individuals are actually affected). We systematically explored which parameters proved more sensitive to perturbations. Such perturbations can be revealing in terms of evaluating the robustness and dependencies among the parameter choices, and yielded several unanticipated choices especially sensitive to perturbation. These results will inform development of an analysis protocol to be used on the NBSeq project, which will study ~1600 exomes of California newborns found to have metabolic disorders, as well as a subset of false positive by using MS/MS in NBS.

1774W

Integrated genomic analysis with IOBIO. *A. Ward, C. Miller, T. Di Sera, Y. Qiao, G. Marth.* Human Genetics, University of Utah, Salt Lake City, UT.

As next-generation sequencing becomes ubiquitous and research MDs, clinical diagnosticians and other non-bioinformaticians become more actively involved in genomic data analysis, easy to use tools that require no installation, data upload, or computational expertize, providing users an interactive, visually intuitive environment are increasingly necessary.We present a genomic data analysis paradigm built on the innovative IOBIO platform. Starting with aligned DNA sequence and variant calls (the typical outputs of institutional level genomic pipelines), IOBIO web-based applications are used to identify problems with underlying DNA libraries (e.g. excessive PCR duplicates), sequence alignments and variant calls. Having established data quality, gene-level variant prioritization is performed using real-time functional annotation, and comparisons with multiple external variant databases and population allele frequencies are executed. False negative detection is provided using interactive, on-demand variant calling with Freebayes in conjunction with known alleles. Variant lists from external variant prioritization tools and gene-lists from phenotype-driven algorithms can be generated natively within the application or imported from external sources. All of this is achieved without the need for expensive, time-consuming data uploads or lengthy analyses.The power of this workflow to generate data-driven questions, and provide rapid methods to answer those questions is demonstrated using two examples; a real-world clinical example of a family trio, where the proband presents with the lactic acidosis phenotype, and a single individual with a family history of heart disease. Data quality is investigated using the **bam.iobio** and **vcf.iobio** applications. **Taxonomer.iobio** is used to understand the metagenomic composition of the samples and **gene.iobio** interrogates the list of variants (prioritized variants having been generated using **Gemini**), in candidate genes identified by **Phenolyzer**. We finally present some of the features and applications that will soon be augmenting this workflow making it a "one-stop-shop" genomic analysis platform for all users of sequencing data.

1775T

Structural variant calling combining Illumina and low-coverage Pacbio. *A. Carroll.* DNAnexus, Mountain View, CA.

Structural variant calling combining Illumina and low-coverage Pacbio   Detection of large genomic variation (structural variants) has proven challenging using short-read methods. Long-read approaches which can span these large events have promise to dramatically expand the ability to accurately call structural variants. Although sequencing with Pacific Biosciences (Pacbio) long-read technology has become increasingly high throughput, generating high coverage with the technology can still be limiting and investigators often would like to know what pacbio coverages are adequate to call structural variants.   Here, we present a method to identify a substantially higher fraction of structural variants in the human genome using low-coverage pacbio data by multiple strategies for ensembling data types and algorithms.   Algorithmically, we combine three structural variant callers: PBHoney by Adam English, Sniffles by Fritz Sedlazeck, and Parliament by Adam English (which we have modified to improve for speed). Parliament itself uses a combination of Pacbio and Illumina data with a number of short-read callers (Breakdancer, Pindel, Crest, CNVnator, Delly, and Lumpy).   We show that the outputs of these three programs are largely complementary to each other, with each able to uniquely access different sets of structural variants at different coverages. Combining them together can more than double the recall of true structural variants from a truth set relative to sequencing with Illumina alone, with substantial improvements even at low pacbio coverages (3x – 7x). This allows us to present for the first time cost-benefit tradeoffs to investigators about how much pacbio sequencing will yield what improvements in SV-calling.   This work also builds upon the foundational work of Genome in a Bottle led by Justin Zook in establishing a truth set for structural variants in the Ashkenazim-Jewish trio data recently released. This work demonstrates the power of this benchmark set – one of the first of its kind for structural variation data – to help understand and refine the accuracies of calling structural variants with a number of approaches.

1776F

NCBI's database of Genotypes and Phenotypes: dbGaP. *M. Feolo, R. Bagoutdinov, M. Bihan, S. Dracheva, L. Hao, Y. Jin, M. Kimura, R. Lapoint, M. Lee, N. Popova, S. Pretel, S. Sharopova, S. Stefanov, A. Stine, A. Sturcke, Z. Wang, W. Wu, M. Xu, L. Ziyabari, S. Sherry.* dbGaP, NIH/NLM/NCBI, Bethesda, MD.

Statement of PurposeThe National Center for Biotechnology Information's database of Genotypes and Phenotypes (dbGaP) is an NIH-sponsored repository charged to archive and distribute information produced by genome-scale studies investigating the interaction of genotype and phenotype. The data submitted to dbGaP include individual-level molecular and phenotype data; analysis results; medical images; general information about the study; and documents such as research protocols and questionnaires. The molecular data deposited to dbGaP include array-based, sequence-based, and imputed genotypes, as well as expression data and all types of next-generation sequencing (NGS), which are processed and distributed by NCBI's Sequence Read Archive (SRA) or by various NIH trusted partners. Many studies also include a wealth of phenotypic data pertaining to study participants with any combination of medical images, demographic, clinical, or exposure variables, plus the statistical measures of the allelic and gene association tests.MethodSubmitted data are validated for consistency and format using both computational and human curation. Valid submissions are accessioned with stable, unique identifiers that are versioned for successive updates, making it possible to cite the primary data in a very specific and stable way. Once processed, the metadata for each study is indexed for public browsing. Individual level data is packaged by consent limitations and made available for approved investigators via a controlled access system. This process supports data sharing policies for both NIH and scientific journals, and is utilized by several large scale NIH programs including The Alzheimer's Disease Sequencing Project (ADSP), Grand Opportunities - Exon Sequencing Project (GO-ESP), and Genotype-Tissue Expression (GTEx).ConclusionThe dbGaP provides unprecedented access to a vast array of genome scale studies. The submitted studies are funded both by NIH and other funding agencies worldwide. The use of the data has been cited in over a thousand published manuscripts. Through controlled access, approved researchers may obtain complete, versioned, and accessioned statistical analyses, study documents, medical images, individual phenotypes, and molecular measures. This presentation will summarize the current data content; describe dbGaP submission and access procedures; and describe how dbGaP has been successfully integrated with other archives/systems.

1777W

Practical interface and data structure in three generation family cohort study. *S. Kazuro, T. Takai, H. Metoki, S. Ogishima, M. Ishikuro, S. Mizuno, S. Nagaie, T. Obara, Y. Sato, M. Miyashita, C. Yamanaka, M. Kikuya, S. Kuriyama, H. Tanaka.* Tohoku-University, Tohoku Medical Megabank, Sendai, Miyagi, Japan.

This paper describes an input interface and data structure, used for the Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study (the BirThree Cohort Study). In the field of genetic statistics, only the relationship between a child and his/her parents is used to describe family structure. This description form has enough information to use for genetics, however, it is insufficient for the operation of the BirThree Cohort Study. We have to register 70,000 of pregnant women and their family without contradiction, and all of their data must be surely input by more than 150 Research Coordinators assigned to more than 50 medical institution from a Web based DB interface, for the BirThree Cohort Study. We enhanced the basic data structure, add a structure to describe the family role. We defined a unit, consists of seven people of "Pregnant (Mother), new born (child), Father, Grand Mother, Grand Father, Grand Mother-in-law, Grand Father-in-law" for family data input. In addition, we are always maintaining this complex data structure by a system that investigates the contradiction. .

1778T

A fast and easy to use framework for automatic biological knowledge base construction. *J. Lever[1,2], M. Jones[2], S.J.M. Jones[1,2].* 1) Bioinformatics, University of British Columbia, Vancouver, BC, Canada; 2) Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada.

Systems biology research relies on well maintained and up-to-date knowledge bases for a wide variety of information, e.g. genetic interactions. However some areas of biology research lack the required databases and researchers need to search raw text of articles for the relevant data. Manually building a robust knowledge base requires significant investment of manpower and time. Automated knowledge base construction techniques that can be used by domain users are required. These allow a computer to read all relevant literature, identifying all relevant knowledge, and can be updated easily with the latest literature. We introduce a generalised framework for easy knowledge base construction for biological systems. It is based upon our VERSE text mining tool which excelled in the BioNLP '16 Shared Task, winning the competition for extracting knowledge on bacteria habitats. The first stage involves users identifying specific types of target entities (e.g. genes, disease) and relations of interest (e.g. treatments). PubMed abstracts and PubMed Central Open Access full text articles are searched for sentences that discussed multiple target entities. A fast and intuitive interactive annotation system allows the user to select sentences that describe relations of interest. Our VERSE text mining tool is then used to train a classifier to identify relations of interest and applied to all remaining text. The knowledge extracted using VERSE is then available for easy search by the user. As an example of the power and ease of this framework, we developed a knowledge base for cancer-related genes known as CancerMine. This resource associates specific cancer types with known driver, oncogenic and tumor suppressive genes. The initial version was developed from scratch in under three weeks, with most of the time spent computationally in order to process all available biomedical text and optimise the machine learning classifier. Using the competition-winning VERSE tool, a robust and easily updated knowledge base can be generated for any biological problem in which information is scattered within abstracts and accessible full text articles. This framework will be prove immeasurably useful to researchers in domains that lack definitive and well-updated knowledge base resources.

1779F

Integrated database "dbTMM" for genomic prospective cohort study in Tohoku Medical Megabank toward personalized healthcare. *S. Ogishima, K. Shimokawa, S. Mizuno, S. Nagaie, T. Takai-Igarashi, S. Nagase, N. Nakamura, A. Hozawa, F. Nagami, J. Sugawara, S. Kuriyama, S. Kure, Y. Suzuki, N. Minegishi, M. Nagasaki, K. Kinoshita, J. Yasuda, N. Yaegashi, H. Tanaka, M. Yamamoto, ToMMo dbTMM Project.* ToMMo, Tohoku University, Sendai-shi, Miyagi, Japan.

To revitalize medical care and realize personalized healthcare in the disaster area of Great East Japan Earthquake, the Tohoku Medical Megabank (TMM) project has developed a large-scale biobank with conducting prospective genome-cohort study. In our prospective cohort study, we will recruit 150,000 people at Tohoku University, satellites, health clinics, and Iwate medical university, and have collected specimen and health survey data (laboratory tests, questionnaire data, and physical measurements) as baseline investigation. As for follow-up cohort study, we will collect clinical data as electronic health records provided by hospitals. Our biobank is an "integrated" biobank having both biobank and sequence facility for genomic and omics analysis to provide data with highly common interest instead of distributing the raw specimen. Integrated biobank provides not only health survey and clinical data but also genomic and omics data. We have developed the integrated database what we call "dbTMM" which integrates both health survey data and genomic data as integrated biobank. We have built Release 1.0.0 (Mar 31, 2016) which consists of genomic data (SNV data over 2.1 million SNV sites) and health survey data (laboratory tests, questionnaire data) for 1,070 participants. Researchers in universities, institutes, and companies can use data from dbTMM after approval from data access committee. Our integrated database "dbTMM" realizes ultra-fast search of large-scale health survey and genomic data, and shows statistical signature of narrowed-down cohort; e.g., if narrowed-down population has statistically significant bias in sex and alcohol intake, sex and alcohol intake will be indicated to users automatically as statistical signature. This statistical signature is expected to give insights and hints to researchers. Our integrated database "dbTMM" is expected to be a platform for data sharing to promote researches aiming at realization of personalized healthcare.

1780W

Prioritizing genes for the genetics community using GTR and ClinVar to close the interpretive gap. *W.S. Rubinstein, M. Landrum, J. Lee, D.M. Maglott, M. Benson, G. Brown, C. Chen, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, B. Kattman, K.S. Katz, A. Malheiro, M. Ovetsky, G. Riley, A. Sethi, R. Tully.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

Background The interpretation of sequence variation is the grand challenge of genomic medicine. As more sequence data are generated, the 'interpretive gap' in our understanding about the clinical significance of variation grows wider (Cutting, 2014; PMC3882730). About a quarter of variants in ClinVar have pathogenicity interpretations different enough to potentially change clinical management (e.g., variant of uncertain significance (VUS) vs. pathogenic). The actionability of a gene is moot for any variants which cannot be reliably interpreted. Methods Data in GTR and ClinVar were used to suggest priorities for ClinVar submission and methods to evaluate the pathogenicity of variants. GTR has >34,000 orderable tests. The number of tests offered for a gene reflects its clinical relevance. ClinVar's gene-specific report (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/gene_specific_summary.txt) includes numbers of VUS and variants with conflicts. ClinVar maintains a table of ACMG's 56 genes for reporting secondary findings. Results GTR has 4,129 tested genes, with the number of tests ranging from 1 to 313. All 56 ACMG genes are tested, with a significantly higher number of tests per gene than the average. ClinVar has submissions with variants specific to >4,500 explicitly reported genes, but is missing 658 GTR tested genes. Six ACMG genes have fewer than average (N=34) ClinVar submissions: *SDHC, CACNA1S, NF2, MYLK, PCSK9, SDHAF2*. 72 labs in GTR offer 292 tests for these genes. ClinVar has 2,398 genes with conflicts (average 12 per gene, range 1 – 1,758) and 565 genes with VUS (average 6 VUS per gene, range 1 – 347). The number of submissions is strongly correlated with the number of conflicts and VUS. We will also report a weighting approach to prioritize genes with more conflicts and VUS. Conclusions We encourage all who register their tests in GTR to submit their variant interpretations to ClinVar to help close the interpretive gap for all genes being tested. Submissions for the six under-represented ACMG 56 genes are especially important because the risk/benefit ratio is higher for secondary variants uncovered in the absence of a reported phenotype. However, additional submissions alone are unlikely to reduce conflicts and resolve VUS. Robust types of evidences, such as functional assays and correlation of clinical phenotype to established gene effects, are particularly needed for clinically relevant genes and welcomed as part of ClinVar submissions.

1781T

Introducing the Vertebrate Gene Nomenclature Committee (VGNC). *S. Tweedie, B. Yates, B. Braschi, K.A. Gray, R.L. Seal, E.A. Bruford.* HUGO Gene Nomenclature Committee (HGNC), European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom.

Standardised gene nomenclature provides an essential resource for all researchers. However an ever-increasing number of vertebrate genomes are being sequenced and the data released into the public domain without any systematic annotation or gene naming. There are only six vertebrate model organisms with an established gene nomenclature committee (mouse, rat, chicken, Anolis, Xenopus and zebrafish), all of which base their gene names on those approved by the HUGO Gene Nomenclature Committee (HGNC) for human genes. The Vertebrate Nomenclature Committee (VGNC) is a new initiative that extends the remit of the HGNC to approve consistent gene names and symbols across vertebrates. Our naming strategy for each vertebrate species starts by identifying a high confidence set of genes with 1:1 human orthologs using our HCOP tool (http://www.genenames.org/cgi-bin/hcop), which aggregates orthology predictions from multiple sources. These 1:1 orthologs are named in a semi-automated manner, with the human gene nomenclature being transferred to the orthologous gene. Genes with non-consensus orthologs, members of complex gene families, pseudogenes and RNA genes require additional manual curation. Our prototype species for VGNC naming has been chimpanzee and we have already named over 10K protein-coding chimp genes with a 1:1 human ortholog. During this process we have taken the opportunity to improve the consistency of our names and taken care to minimise transfer of species-specific information. This naming process will soon be expanded to other species, including dog and cow. An online vertebrate gene nomenclature portal has been created that stores, displays and makes this new nomenclature data accessible both to individual researchers and available for dissemination to other resources including Ensembl and NCBI Gene. Further information and requests for individual vertebrate gene names and symbols can be made via: http://www.genenames.org..

1782F

Next-generation sequencing (NGS) vs. Sanger sequencing: Defining quality parameters, reliability scores and thresholds to exclude false positive variants in a diagnostic setting. *M. Ziegler, Y. Dincer, S. Eck, J. Schulz, H.G. Klein.* Center of Human Genetics, Martinsried, Germany.

Separating true positive from false positive variants is a big challenge especially in huge exome datasets. Therefore, thresholds for NGS quality parameters (e.g. average quality, coverage, read balance) should be determined for effective filtering and maximum precision. We statistically analyzed 387 kb of data from 19 patients with various phenotypes (arrhythmia, hearing loss, nephrotic syndrome, thoracic aortic aneurysm and dissection, porphyria) sequenced simultaneously via Sanger sequencing and NGS. Statistics were performed using the in-house NGS-Pipeline (Agilent SureSelect Target Enrichment and Illumina MiSeq/NextSeq and CLC Genomic Workbench) with Sanger sequencing analysis (ABI3730 and JSI SeqPilot) as reference. The sensitivity (99.6 %), specificity (100.0 %), accuracy (100.0 %) and precision (96.6 %) were high for variants with coverage above 20. Integrating mutation confirmation data and variants with coverage from 10 to 20 in the analysis reveals 18/290 false positive called variants. On the basis of this data, a formula and reliability threshold was determined via statistical models. The reliability formula was then integrated in the in-house variant database for validation. When tested on further NGS confirmation data, the threshold reveals high precision and also represents a user-friendly application. Continuing analysis of confirmation data to increase prediction precision of the reliability threshold should be considered.

1783W

Updating continuously changing variant information in routine diagnostics. *B. Liesfeld[1], D. Schanze[2], R. Ewald[1], S. Leye[1], R. Abou Jamra[3].* 1) Limbus Medical Technologies GmbH, Rostock, Germany; 2) Institut für Humangenetik, Otto-von-Guericke-Universitaet Magdeburg, Magdeburg, Germany; 3) Institut für Humangenetik, Universitaetsklinikum Leipzig, Leipzig, Germany.

The ever increasing number of detected variants poses a great challenge for routine diagnostics. Complex software systems have been developed to integrate information from hundreds of different sources to annotate variants from a patient. New information emerges every day and previously trusted information becomes outdated at an increasing speed. While clinicians strive to provide highest-quality assessment to their patients based on the knowledge they have at the moment of diagnosis, it is extremely difficult to routinely review all previous cases in regular intervals. The large number of variants detected in every patient and the increasing number of patients submitted to genetic testing call for an automated and scalable system to perform regular reviews. We evaluate a system architecture that integrates the variant assessment service VARVIS with the ALLEXES network which provides access to a comprehensive number of different reference data sources. VARVIS is used to manage and maintain genotypic and phenotypic data of every patient of an institution. The ALLEXES network provides reference data for genomic variants across public data sources, e. g. ClinVar, as well as pathogenicity scores. In addition, the ALLEXES network links all connected VARVIS services so that variant data can be exchanged without compromising data privacy. All variants of a patient are annotated with data from the ALLEXES network. The VARVIS service highlights any inconsistencies of variant classification in comparison to local historic data or to data provided by the ALLEXES network. The data provided by ALLEXES includes allelic frequency data and variant classifications from other institutions connected to the ALLEXES network. At the same time the patient is registered with the ALLEXES network with all de-identified phenotypic and genotypic data. The improvement of this architecture over state-of-the-art genomic solutions is that the central ALLEXES network has the capability to notify any connected institution about changes in reference data that may affect the diagnosis of a patient. Rules can be configured flexibly such that the level of sensitivity to changes can be adjusted according to the standards of the local institution. The management and review is integrated into the graphical user interface of VARVIS. This system architecture is in use at several clinical genetics institutions and proves to be invaluable for the delivery of highest quality of care to our patients.

1784T

Distribution of machine images for rapid deployment of NCBI and community bioinformatics tools. *B. Busby.* NCBI-NLM, NIH, Bethesda, MD.

There is an increasing need to provide instruction on and rapid deployment for NCBI tools used on the LINUX command line, including EDirect, standalone BLAST, and the SRA toolkit, as well as community tools that are integral to using NCBI NGS data (e.g. samtools), and very popular tools that can be challenging to configure with our data structures (e.g. GATK). Setting up such environments offsite is often a challenge. We are releasing machine and docker-type images -- including popular reference genomes -- that can deploy NCBI resources and some community tools that work with them. These machine images will be launchable on several cloud platforms (e.g. Amazon and Google) and distributed computing environments (e.g. CyVerse). Please visit the NCBI learn page (ncbi.nlm.nih.gov/learn) for more information. .

1785F

A personal genome analysis course for undergraduates. *D.J. Matthes.* Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN.

As we enter the era of personal genomics, where it will soon be common medical practice to determine the genome sequence of each patient, and where it is already possible to inexpensively obtain ones genotype at nearly a million commonly polymorphic loci, the need for personal genome-oriented genetics education has arrived. Here I describe a personal genome analysis course for undergraduates in which students (1) are genotyped by a direct-to-consumer sequencing company, (2) work in a collaborative team-based learning course to analyze how their own variants reveal their ancestry and (imperfectly) predict their traits, and (3) write their autogenomography – a collection of stories each student finds within their genome. The structure of the course, the 20 genome analyses students carry our, examples of student work, and an assessment of student learning and achievement of course objectives will be discussed. Briefly, students improve their performance on select questions on the Genetics Concept Assessment, performed uniformly well (80%) and above on a final exam of technical issues in personal genome analysis, and research and write a 50+ page document (their autogenomography) that demonstrates the level of genomic understanding they have attained. While some students grappled with ancestry results that differed from what they expected, or with the implication of learning their carrier status for some significant diseases, the overall experience, even for these students, was a transformative and extremely positive one. In anticipation of the day when nearly every biology department will offer a personal genome analysis course, here I present a template and course materials for consideration, discussion and adaptation.

1786W

Whole-exome sequencing uncovers candidate mutations for genetic suppressors of Rett syndrome symptoms in *Mecp2* **mice.** *R. Zeng[1], J. Ruston[1], M.J Monica[1,2].* 1) Genetics&Genome Medicine, SickKids hospital, Toronto, ON, Canada; 2) Department of Molecular Genetics, The University of Toronto, Toronto, Ontario Canada.

Rett syndrome is a rare neurodevelopmental disorder characterized by progressive neurological decline and breathing abnormalities after birth. Ninety-five percent of patients carry mutations in the X-linked gene methyl CpG binding protein2 (*MECP2*). The *Mecp2*-null mouse is a useful model for Rett syndrome because its symptoms include limb clasping, tremors, uneven breathing, impaired coordination and hypo-activity. To demonstrate the power of suppressor genetics to identify therapeutic pathways, we used a dominant *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen to identify loci that suppressed symptoms in *Mecp2*-null mice. A Squalene epoxidase (*Sqle*) nonsense mutation was identified in one suppressor line, which suggested the cholesterol homeostasis pathway as a potential target for the treatment of Rett syndrome [1] We have extended the suppressor screen to isolate 46 dominant suppressor lines that show amelioration of health assessment traits. To identify gene mutations that implicate pathological disease processes and suggest new therapeutic treatments, we sequenced samples by using whole-exome DNA sequencing. To identify candidate genes, we sequenced the G1 founder and 2 - 4 samples from G3 offspring that inherit the suppressor traits. Sequencing was performed on an Illumina Hiseq 2000 system. Hard-filtering for the variants was based on a minimum SNP quality of 20 and read depth of 5. To narrow down the sequence variants (SVs), we removed common polymorphisms reported in the Wellcome Trust Sanger Institute mouse SNP database. Since it is unlikely to induce recurring variants by ENU, we further filtered SVs by removing shared polymorphisms present in other suppressor lines . By using these strategies, we reduced the potentially causative SVs from tens of thousands to a range of 22-77 mutations in different lines. Here, I report a missense mutation in Ectonucleoside triphosphate diphosphohydrolase3 (*Entpd3*), which is shared in the G1 founder and his trait-improved progenies in suppressor line 136. qRT-PCR reveals that the expression of *Entpd3* is significantly up-regulated in eight week old *Mecp2*-null mouse brain, which might indicate a perturbation in the nucleotide-hydrolyzing pathway in *Mecp2*-null mice. Altogether, we conclude that exome sequencing alone without mapping is an efficient and fast way to identify gene mutations induced by ENU mutagenesis. [1] Buckovecky CM, et al. (2013) *Nature Genetics* 45(9): 1013-1020 .

1787T

Use of electronic health records to generate one of the largest family cohorts with detailed longitudinal phenotypic data. *S. Hebbring[1], X. Huang[2], R. Elston[3], G. Rosa[4], J. Mayer[5], Z. Ye[5], T. Kitchner[1], D. Page[2].* 1) Human Genetics, Marshfield Clinic, Marshfield, WI; 2) Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 4) Department of Animal Science, University of Wisconsin-Madison, Madison, WI; 5) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI.

Diseases are the result of genetic and/or environmental factors. Understanding the etiology of human disease is necessary to advance precision medicine. Pedigree analysis is a longstanding approach to gain insight into the underlying genetic factors in human health. In the last decade, pedigree analysis has predominantly given way to large population-based studies on samples of independent persons as a result of high-throughput genotyping methodologies and the ability to collect significant numbers of unrelated cases and controls. A reverse paradigm shift may occur if the construction and use of family pedigrees could be combined with classical genetic analyses in a high-throughput manner. The goals of this study were to rapidly predict family pedigrees from an electronic health record (EHR) system with no human intervention and demonstrate its potential relevance in genetic research and clinical utility. Using basic demographic data in an EHR with nearly 2 million patients, we developed an algorithm that identified familial relationships among 579,561 individuals that were linked to 173,368 pedigrees. Manual validation demonstrated that over 98% of the relationships were accurately defined. To demonstrate the utility of EHR-linked families, we extracted diagnostic data for 19 congenital diseases (presumed genetic disorders) and 9 non-genetic phenotypes as controls, to determine if genetic relationships in EHR-linked pedigrees can predict disease risk. In essence, we created family histories for 28 clinical phenotypes using nothing more than readily available data in an EHR system. When comparing the predictive value of genetic relationship in the two disease categories, genetic relationship was strongly associated with congenital phenotypes (p=0.031). Of the 19 congenital disorders, genetic relatedness was predictive (p<0.0001) for 10 diseases. The strongest association was with "congenital anomalies of limbs" (p<1.0E-15). With the anticipated widespread application of genomic medicine, in combination with methods capable of predicting family pedigrees linked to extensive longitudinal phenotypic data in an EHR, a quantum leap may occur in how human genetic research is conducted for the advancement of precision medicine.

1788F

Decoding *cis*-regulatory function directly from DNA sequence. *K.A. Barr[1], J. Reinitz[2].* 1) Genetics Genomics and Systems Biology, University of Chicago, Chicago, IL; 2) Ecology and Evolution, University of Chicago, Chicago, IL.

In the last several years there have been major efforts to catalog regulatory genetic elements in both humans and model organisms, as well as to record differences in gene expression between individuals and tissues. However, the extent of sequence variation among individuals and vast number of cell types makes it impossible to record expression under every possible condition. To make matters more complicated, gene regulation is a function of large genetic loci that can span megabases and harbor numerous genetic variants that preclude the identification of functional variation. These issues may be addressed by mechanistic models of gene regulation that have the potential to calculate expression levels resulting from previously unobserved sequences or cell types, and to identify the functional regulatory variants in large loci and the means through with they act. In this work we present a thermodynamics-based model that is able to decode the regulatory function of the *even-skipped* (*eve*) genetic locus in *Drosophila melanogaster*. This locus has been studied for nearly three decades and serves as a model system for studies of gene regulation. Using confocal microscopy, our lab has quantified both transcription factor levels and mRNA levels at single-nucleus resolution along the anterior-posterior axis of developing embryos. These data amount to a set of quantitative single cell assays of transcription input and output in a native tissue context, providing an extraordinarily precise testbed for theoretical models. When our model is trained on this data – taking DNA sequence, binding preferences of transcription factors (TFs) in the form of position-weight matrices, and levels of these TFs as input – we are able to directly compute transcription levels. Our model identifies four enhancers within the *eve* locus and describes their binding site structure with single-nucleotide precision. We can identify which TFs are responsible for each expression feature of *eve* and predict the effects of mutations in *cis* and *trans*. The model code includes a dynamic-programming algorithm that scales linearly with the number of binding sites identified, which will allow it to be applied to new organisms and larger datasets.

1789W

Mixed-layered network modeling for gene expression across individuals and tissue types. *S. Yang[1], D. Pe'er[2,3], I. Pe'er[1,3].* 1) Computer Science, Columbia University, New York, NY; 2) Biological Sciences, Columbia University, New York, NY; 3) Systems Biology, Columbia University, New York, NY.

Expression quantitative loci (eQTLs) have been extensively studied, as they provide an attractive functional interpretation to sequence variation. Yet, many traditional eQTL studies had been mainly focused on *cis*- analysis of each transcript, without drawing on the complete picture of the regulatory repertoire. Furthermore, such studies standardly remove genomewide effects from data as a preprocessing step, interpreting them as artifactual batch effects rather than genetic or other biological signal. In this work, we devise a mixed-layered regression model, considering polygenic multi-loci eQTL regulation of gene expression. The model includes several layers of variables, where successive layers are connected through logistic functions whose parameters are inferred through regression with sparsity regularization. A key feature of our model is allowing *trans*-SNPs to indirectly affect a particular target gene through an intermediate hidden layer, which we call *cell environment variables*. We encourage these variables to affect gene expression in a cell-type-specific fashion. Variables in this hidden layer are designed to implicitly describe unmeasured quantities in the sampled cells, such as chromatin state, activity of a cellular pathway, level of a small molecule, or unknown mechanisms, all of which affect expression in a cell-type-specific manner. The flexibility of the model allows additional features: genomewide effects due to individual genotypes (population stratification) and samples (experimental, environmental and other sources) can be modeled directly in a non-linear fashion, which we show is required by real data; prior assumptions regarding locality and annotations of *cis*-eQTLs can be incorporated; and multiple tissues can be modeled as mixtures of cell types whose characteristic expression profiles are related according to a known hierarchy. We solve the model using stochastic gradient descent with backpropagation. Naïve implementation of such an inference procedure is computationally prohibitive at genomewide, transcriptomewide scale. We mitigate this challenge by considering only a representing subset of all SNPs, and further accelerate the computation through GPU. Results on simulated datasets allow estimation of the required number of iterations for accurate inference and thus demonstrate the model's applicability to the scale of real-data in less than a day of runtime. We test our modeling on the GTEx dataset, and present the results.

1790T

The eMERGE Network: Continuing the legacy of genomic discovery to enrich precision medicine. *M.A. Basford[1], P. Harris[1], M. Brilliant[2], J. Denny[1], B. Funke[3,4], S. Gabriel[3,4], A. Gharavi[5], R. Gibbs[6], H. Hakonarson[7], I. Holm[8], J. Harley[9], G. Hripcsak[5], G.P. Jarvik[10], E. Karlson[11], I. Kullo[12], E.B. Larson[10], R. Li[13], S. Murphy[11], H. Rehm[3,4], M.D. Ritchie[14], D. Roden[1], M. Smith[15], J. Smoller[11], S. Thibodeau[12], S. Weiss[11], C. Weng[5], K. Wiley[13], M.S. Williams[14], R. Chisholm[15], The Electronic Medical Records and Genomics (eMERGE) Network.* 1) Vanderbilt University Medical Center, Nashville, TN; 2) Marshfield Clinic Research Foundation, Marshfield, WI; 3) Partners Healthcare Laboratory for Molecular Medicine; Cambridge, MA; 4) The Broad Institute of Harvard and MIT, Cambridge, MA; 5) Columbia University, New York, NY; 6) Baylor College of Medicine Human Genome Sequencing Center (HGSC); Houston, TX; 7) Children's Hospital of Philadelphia, Philadelphia, PA; 8) Boston Children's Hospital, Boston, MA; 9) Cincinnati Children's Hospital Medical Center Cincinnati, OH; 10) Group Health Research Institute with the University of Washington and the Fred Hutchinson Cancer Research Center, Seattle, WA; 11) Harvard Medical School, Boston, MA;; 12) Mayo Clinic, Rochester, MN; 13) National Human Genome Research Institute (NHGRI), Bethesda, MD; 14) Geisinger Health System, Danville, PA; 15) Northwestern University, Chicago, IL.

The eMERGE Network is a national consortium, funded by NHGRI, NIH since 2007 to conduct genomic discovery and clinical implementation research by combining biorepositories with electronic health record (EHR) systems. Primary areas of focus include developing consistent approaches to gene and variant interpretation, establishing and improving processes for integration of genomic data into EHRs, evaluating methods for returning results, measurement of implementation and outcomes, genomic discovery, and advancing the science of electronic phenotype development. In previous cycles, the Network created a merged set of approximately 80,000 participants with genome-wide genotyping data that have been imputed using the 1000 Genomes reference. An additional 9015 samples were sequenced using the PGRNseq platform targeting 82 pharmacogenes with variant data available through SPHINX (emergeSphinx.org). Tools include an online platform for sharing clinical decision support artifacts, CDSKB.org, and a portal for developing, annotating and sharing electronic phenotypes, PheKB.org. Forty-seven computed phenotypes have been validated and implemented across the Network with 42 more in development. A large-scale survey of over 90,000 patients' attitudes regarding consent and data sharing was completed to help inform rule making for biobanks. One ultimate goal is to return genomic testing results to patients in a clinical care setting to enable precision medicine. During the next 4 years, the Network is producing targeted sequencing results on a projected 25,000 subjects with extensive EHR data to understand how returning CLIA-grade genomic results impacts clinical care, and to provide a robust dataset for discovery research, especially for assessing the contributions of rare variants in actionable genes. The Network designed a sequencing platform, eMERGEseq, containing the 56 ACMG genes, 53 other genes, and 1552 SNPs, including ancestry markers. Sequencing and clinical reporting is harmonized to the extent possible across two sequencing centers. The creation of these large sets of genotyped and sequenced patients with EHR phenotpyes will expand opportunities for discovery and implementation of genomic varants in healthcare.

1791F

Whole genome reference panel of two thousand Japanese individuals and bioinformatics in ToMMo. *M. Nagasaki.* Tohoku University, Sendai, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) located at the north-eastern Japan is now developing a biobank (the final goal is 150,000 volunteers with millions samples) that combines medical and genome information during the process of rebuilding the community medical system and supporting health to foster the reconstruction from the Great East Japan Earthquake on Mar/11/2011. One of the missions of ToMMo is to reveal a fine genetic architecture of Japanese population to tackle the further genome wide associate study analysis by combining the knowledge, which is daily accumulated in the ToMMo prospective genome cohort project, e.g. questionnaire data, physiological data, medical treatment records and other omics data from serum, plasma and immortalized lymphocytes. We have created the reference panel with one thousand samples to cover MAF > 0.5% variants including short insertion, deletion, and large structural variants in Japanese for constructing the Japanese whole-genome reference panel (1KJPN). Currently, we have constructed the second reference panel with two thousand samples (2KJPN). In our panel, to minimize the biases caused by the different equipment, protocol and bioinformatics analysis, we performed whole genome sequencing of thousands samples with 30x high coverage using the HiSeq 2500 rapid run mode and analyzed by the same bioinformatics pipeline. We presents our bioinformatics analyses, e.g. HLA-typing and CNVs, preliminary findings while constructing 2KJPN, and another effort to construct Japanese reference genome.

1792W

Expanding the scope of the NHGRI-EBI genome-wide association study (GWAS) Catalog. *J.A.L. MacArthur[1], T. Burdett[1], M. Cerezo[1], F. Cunningham[1], P.N. Hall[2], E. Hastings[1], L.A. Hindorff[2], H.A. Junkins[2], A. McMahon[1], A. Milano[1], J. Morales[1], Z.M. Pendlington[1], D. Welter[1], H. Parkinson[1].* 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom; 2) National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD.

The GWAS Catalog (www.ebi.ac.uk/gwas) is a manually curated, publicly available resource of all published GWAS and association results. As of May 2016 the Catalog includes 2,437 studies and over 24,000 SNP-trait associations. Studies and associations must meet the Catalog's strict eligibility criteria (www.ebi.ac.uk/gwas/docs/methods) to be included; studies must include an array-based GWAS and analysis of 100,000+ SNPs, while SNP-trait associations must have a p-value $<1\times10^{-5}$. GWAS study design, genotyping technologies and user needs have evolved since the inception of the Catalog. As part of ongoing assessments of user needs, we carried out a survey which indicated demand to extend the scope of the Catalog to include targeted/non-genome-wide arrays, all SNP-trait associations and follow-on studies. Here we assess the feasibility of including these in the Catalog. Association studies using large-scale targeted/non-genome-wide arrays, including the Metabochip, Immunochip and Exome arrays, are currently not included in the Catalog. Literature searching indicates that to date at least 100 published studies of this type are eligible for inclusion and would enrich the data on immunological and cardiometabolic phenotypes. The inclusion of all SNP-trait associations, regardless of p-value, would create a more comprehensive and representative Catalog, allowing users to more accurately compare and combine data from multiple studies. Follow-on studies would provide additional information to help users refine the region of association tagged in the GWAS. Increasing the scope will require the development of efficient methods to curate an increased number of studies and associations, along with the infrastructure to cope with handling and presenting data of increased volume. A pilot for the inclusion of targeted array studies has allowed us to estimate the resources and infrastructure developments required and we are currently applying for additional funds to include these data. We have also implemented a pilot of author submission to investigate whether authors can provide structured information directly to the Catalog, reducing the burden on curators. Our user submission tools are available and we actively encourage testing by authors. Increased responsiveness to user needs, community engagement and user involvement will make this resource more comprehensive and representative. This will increase the utility of the Catalog ensuring continued increase in usage.

1793T

A graph diffusion framework for multi-scale reproducibility and differential analysis of 3D chromatin contact maps. *O. Ursu[1], N. Boley[1], M. Taranova[1], R. Wang[2], A. Kundaje[1,3].* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Department of Statistics, Stanford University, Stanford, Stanford, CA; 3) Department of Computer Science, Stanford University, Stanford, CA.

Deciphering the three-dimensional organization of the genome is critical to understanding the role of long-range chromatin contacts in gene regulation and disease. Chromosome conformation capture techniques such as HiC, Capture-C and ChIA-PET are commonly used to experimentally measure chromatin 3D contact maps. However, the hierarchical, multi-scale organization of genomes into large-scale compartments, topologically-associated domains, looping interactions and dynamic contacts makes comparisons of contact maps particularly challenging. Here, we overcome this with a statistical framework that leverages graph diffusion to obtain 1) a novel measure of reproducibility and 2) a statistical test for dynamics in contact maps at multiple scales and with high power. We represent contact maps as graphs and use random walks on the graph to smooth the data while maintaining high resolution of contacts and boundaries of domains. This is critical as contact maps are sparse, leading to apparent changes in contacts that are the result of sampling noise and dropout, a phenomenon which is alleviated with our smoothing scheme. We develop a novel multi-scale concordance measure to assess reproducibility of contacts for random walks of increasing length. We calibrate our reproducibility scores on simulated data, and benchmark them on a variety of HiC datasets, recapitulating differences between technical replicates, biological replicates, and different cell types. Our framework generalizes seamlessly to contact maps from other assays such as ChIA-PET and CaptureC. Finally, we derive a statistical test for differential analysis of contact maps. We identify dynamics at multiple scales simultaneously, where the scales are defined naturally as clusters of genomic regions with similar changes in contact profiles. In addition to considering multiple scales, this test aggregates information across multiple neighboring genomic regions, substantially increasing statistical power to detect differences, as well as producing more robust results compared to existing methods. We quantify changes in contact maps between cell types, as well as quantify allele-specific contact differences. These are supported by orthogonal evidence from changes in histone marks, chromatin accessibility and regulatory factors defining boundary elements, suggesting that our method identifies statistically significant and biologically meaningful differences in 3D genome organization across different contexts.

1794F

Whole-genome characterization in pedigreed non-human primates using Genotyping-By-Sequencing (GBS) and imputation. *B.N. Bimber[1], M. Raboin[1], J. Letaw[1], K. Nevonen[1], J.E Spindel[2], S.R. McCouch[2], R. Cervera-Juanes[1], E. Spindel[1], L. Carbone[1], B. Ferguson[1], A. Vinson[1].* 1) Oregon National Primate Research Center, Beaverton, Oregon, and the Oregon Health & Science University, Portland, Oregon; 2) 2Section of Plant Breeding and Genetics, School of Integrative Plant Sciences, Cornell University, Ithaca, New York.

Rhesus macaques are widely used in biomedical research, but the application of genomic information in this species to better understand human disease is still undeveloped. Whole-genome sequence (WGS) data in pedigreed macaque colonies could provide substantial experimental power, but the collection of WGS data in large cohorts remains a formidable expense. Here, we describe a cost-effective approach that selects the most informative macaques in a pedigree for whole-genome sequencing, followed by low cost Genotyping-By-Sequencing (GBS) on the remaining animals to generate high accuracy, sparse genotype data. The dense, high-accuracy variants from the WGS animals are imputed into the remaining GBS individuals, providing genome-wide genotypes at greatly reduced cost. We developed GBS for the macaque genome using a single digest with *PstI*, followed by sequencing to 30X coverage, producing high-confidence genotypes over >22,000 macaque SNVs. To validate the utility of GBS data for genome-wide imputation, we focused on a 16-member pedigree, initially limited to chromosome 19. Using the approach of "Genotype Imputation Given Inheritance" (GIGI), we imputed alleles at 5,010 markers on chromosome 19, using 833 sparse markers from GBS. We evaluated multiple factors on imputation accuracy, including: 1) the strategy for selecting the individual to receive GBS, 2) the ratio of subjects with WGS to those with GBS, and 3) the density of the framework marker panel. We found that the GIGI-Pick algorithm for selection of WGS subjects out-performed other common approaches, and that a ratio of 1 WGS subject per 3-5 GBS subjects produces an optimal tradeoff between cost and number of imputed genotypes. The density of the framework marker panel had little effect on accuracy. Using these optimized conditions, we imputed an expanded set of ~7.6 million variants across all 20 macaque autosomes, achieving >99% accuracy per chromosome. This approach makes feasible the collection of accurate, dense genome-wide sequence data in large pedigreed macaque cohorts without the need for expensive WGS data on all individuals.

1795W

Efficient population-scale variant annotation and analysis through big data. *D. Gordon, H. Kuck, B. Kelly, J. Fitch, P. White.* The Genomics Institute at Nationwide Children's Hospital, Columbus, OH.

The genomics community must adopt new big data technologies to enable population-scale genomic analyses and accelerate knowledge discovery in human health. Advances in technology are empowering population-scale genome sequencing of thousands of individuals monthly, creating large volumes of variant data. Traditional methods for variant annotation and analysis are limited in scale, lack real-time functionality and impose significant bottlenecks on the process of discovery and clinical diagnosis. Facing these challenges of scale and performance requires a novel architecture, for which we have selected the Hadoop ecosystem as a foundation. Hadoop's MapReduce algorithm allows data to be seamlessly split across nodes within a computing cluster, and provides a high performant solution on commodity hardware. In addition to the MapReduce architecture, we leverage HBase, Hadoop's implementation of the BigTable storage system. HBase is a horizontally scalable data store, meaning that as the scale of genomics data increases, the system can be "scaled out" by adding more nodes to the computing cluster. Building upon our back-end storage and processing system able to process massive amounts of data, we have implemented a web-based data mining and analysis solution. Through this interface, users can submit raw variant data for annotation, manage genes of interest and pedigrees, and perform near-real-time queries of variants. As the scale and throughput of variant data continues to increase, so too does the challenge of keeping up with the flow of data and making it available in a way meaningful to researchers and clinicians. By creating a scalable, approachable, and affordable solution to this challenge, our solution significantly improve complex data mining on stored variant data and equips investigators with powerful and efficient tools for efficient querying thereby accelerating genome research and its application to human health.

1796T

Strategic interrogation of the exome for metabolic and glycomic variants in the setting of undiagnosed disease. *C.J. Adams[1], D.R. Johnson[1], T.S. Frisby[1], K.R. Chao[1], M. Davids[1], L.A. Wolfe[1], H. Ho[1], C.L. Hoppel[2], W.A. Gahl[1], D.R. Adams[1].* 1) Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD; 2) Case Western Reserve University, Cleveland, OH.

The NIH Undiagnosed Diseases Program (UDP) evaluates patients with illnesses that remain undiagnosed despite extensive medical investigation. Exome or genome analyses are undertaken on UDP patients with the objective of identifying DNA sequence variants that may contribute to the causation of the presenting illness. For any given case, multiple such candidates may be found, with varying levels of bioinformatic evidence. Ancillary data sets, including metabolic and glycomic testing, may augment variant prioritization by revealing biochemical pathway associations. For instance, variants in a gene with a predicted biochemical function but no disease association can be prioritized if metabolic testing reveals abnormalities consistent with dysfunction of the given gene. Such associations may be detected even when metabolic or glycomic disease was not present in the original differential diagnosis, as clearly-apparent metabolic or glycomic diseases are likely to have been identified prior to presentation to the UDP. Here we present in-progress examples of cases where synergy between metabolic and molecular data from UDP cases has prompted follow up investigation. For instance, a UDP patient presenting with developmental regression, seizures, CNS and neuromuscular findings was found to have an abnormal metabolic profile. The pattern of abnormal metabolites did not clearly match any known metabolic disease. However, several acylcarnitine profile abnormalities supported the presence of a defect of fatty acid oxidation. This finding was used to support the prioritization of compound heterozygous variants in Acyl-CoA Synthetase Medium-Chain Family Member 5 (*ACSM5*), a fatty acid synthetase gene with a presumptive but unproven role in fatty acid activation for the metabolism of medium-chain fatty acids. We present a framework for conducting the analysis of sequence variants given additional metabolic data, family-based controls, and extensive, standardized clinical phenotyping. This research is supported by the Intramural Research Program of the National Human Genome Research Institute and the Common Fund of the National Institutes of Health.

1797F

Imputation of high-resolution *KIR* types from SNP variation data. *S. Leslie[1,2,3], D. Vukcevic[1,2,3], P. Norman[4].* 1) School of Mathematics and Statistics, University of Melbourne, Parkville, Vic 3010, Australia; 2) School of BioSciences, University of Melbourne, Parkville, Vic 3010, Australia; 3) Data Science, Murdoch Childrens Research Institute, Parkville, Vic 3052, Australia; 4) Department of Structural Biology, Stanford University, California, USA.

Large population studies of immune system genes are essential for characterizing their role in diseases. Of key interest are genes encoding the killer-cell immunoglobulin-like receptors (*KIR*s), which have known and hypothesised roles in autoimmune diseases, resistance to viruses, reproductive conditions and cancer. These genes are highly polymorphic, making typing expensive and time-consuming. Consequently, despite their importance, *KIR*s have been little studied in large cohorts. This parallels the case of the human leukocyte antigen (*HLA*) genes, whose proteins are known to interact with *KIR* and have been implicated in many autoimmune conditions. *KIR*s are highly variable in terms of gene arrangement and copy number; haplotypes can comprise 4–20 *KIR*s. Further, individual *KIR* genes are highly polymorphic, with over 600 alleles currently described. Statistical imputation methods developed for other complex loci (e.g. *HLA*), based on single nucleotide polymorphism (SNP) data, provide an inexpensive high-throughput alternative to direct laboratory typing of these loci and have enabled important findings for many diseases. KIR*IMP, the first imputation technique designed specifically to impute *KIR* copy number variation and haplotypes defined by *KIR* copy number (called *gene content*), enables accurate typing of *KIR* copy number and gene content from SNP data. One may think of this as low-resolution typing. Here we extend KIR*IMP to high-resolution *KIR* typing: overlaying the allelic polymorphisms for each *KIR* gene copy present. To develop our method, KIR*IMP:02, we have assembled a reference dataset of over 2000 individuals with high-resolution *KIR*-types, derived from sequence, for three *KIR* genes (*3DL1*, *3DL2* and *2DL4*). These samples also have dense genotyping data across the *KIR* region. To assess the accuracy of KIR*IMP:02 we use cross-validation within the reference panel, as well as validating using an independent panel of approximately 100 samples with relevant SNP and *KIR* types. We present results on the accuracy of KIR*IMP:02 and its utility for association analyses. KIR-IMP:02 allows for the first time the augmentation of large existing (e.g. UK BioBank) and novel genetic data sets with high-resolution *KIR* types, enabling detailed investigation of the role of *KIR*s in human disease.

1798W

Comparison of algorithms for error correction in Ion Torrent PGM data: Application to hepatitis B virus. *L. Song[1], W. Huang[1], J. Kang[1], Y. Huang[2], H. Ren[1], K. Ding[1].* 1) Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Institute for Viral Hepatitis, Department of Infectious Diseases, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, China; 2) Center for hepatobillary and pancreatic diseases, Beijing Tsinghua Changgung Hospital, Medical Center, Tsinghua University, Beijing, China.

Ion Torrent™ Personal Genome Machine (PGM) technology is a long-read, low-cost and high-speed next-generation sequencing platform, especially with applications to microbial sequencing. However, there is a relatively higher insertion and deletion (indel) error rate in PGM data, and a systematic assessment of error-correction algorithms in PGM data applications to virus (e.g., hepatitis B virus (HBV)) has yet to be fully performed. In 19 quality trimmed PGM data for HBV reverse transcriptase (RT) region, we noted a total error rate of 0.004877 ± 0.001214 and deletion errors were obviously present in the homopolymers. Comparison of algorithms indicated different performances for error-correction, leading to different sensitivity and specificity of variants called from post-corrected alignments. Simulated data showed that error rate and sequencing depth affected the performances of error-correction algorithms. In conclusion, choice of algorithms for error-correction on Ion Torrent PGM data needs full considerations. .

1799T

Automated mosaicism detection. *M.G. Gordon[1], B.N. Pusey[1], A.P. Liebendorfer[1], V.L. Arthur[2], M.O. Ombrello[2], T.C. Markello[1].* 1) Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD; 2) Translational Genetics and Genomics Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Introduction: Mosaicism is an important finding in both congenital genetic syndromes and Knudsen "two hit hypothesis" conditions including cancer. The percent and location of mosaicism can currently be detected using b-allele frequency data from SNP chips (Markello et al. MGM (2012) 105:665.). The intrinsic vectorial nature of that method suggested a direct route to extend and improve this tool. **Methods:** This program automates our previously described method of detecting mosaicism. It directly models the degree of bimodal separation in the heterozygous band of the b-allele frequency that we previously have shown to reflect the degree of mosaicism in populations of cells. We converted our use of the cumulative distribution function and non-linear regression searches to vector operations that run on graphics processers. The increased throughput allows greater sensitivity and specificity as well as automation of the entire process. This has allowed us to include new refinements to the original method, include automatic edge detection for the mosaic region start and stop SNP positions and correct experimental artifacts such as the use of the Manhattan distance in commercial SNP data. **Results:** Analysis of two different Illumina chips from three data sets containing over 2,000 chips demonstrate that 1% of chips include a mosaic event. Mosaic regions as small as 1Mb and as low as 4% are automatically detected and reported. We have found one example of mosaicism recurrent in 2 generations presumably due to an inherited fragile site. One unique anonymized control case had the same low level 15% mosaicism in 7 different chromosomes. Genome wide chip analysis, can exclude mosaicism in a matter of seconds; large numbers of chips can be processed in minutes using off-the-shelf commercial video cards or even faster using GPU nodes on large computer clusters. **Conclusions:** We have developed an extension to our published method that can efficiently, economically, and automatically detect and quantitate mosaicism regions across the entire genome. Noisy b-allele frequency data from saliva or Formalin-Fixed Paraffin-Embedded samples, or exome, and or genome sequence read count b-allele frequency will have dynamic range limitations that will limit the sensitivity and specificity of this technique. We are in the process of developing an outward-facing website for users to submit genome-wide b-allele frequency data for processing and plan to launch by 2017.

1800F

Integrative variation discovery in whole exome data through two-tiered mapping and *de novo* **assembly.** *S. Tian, H. Yan, M. Kalmbach, S. Slager.* Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905.

Whole exome sequencing (WES) is widely used for the detection of causal variants. Current variant discovery largely adopts mapping-based pipelines, which generally work well in ordinary genomic regions but are much less sensitive in "difficult regions", i.e., the regions with low complexity, segmental duplications, or high sequence divergence from reference genome. Several de novo assemblers have been developed for variant discovery. They tend to be more sensitive to INDELs, SNP clusters, and structural variation. However, these de novo assemblers are mostly tailored to whole genome sequencing (WGS) data. In principle, the mapping-based and de novo assembly-based approaches complement each other and their combination should allow more complete variant detection. Here, we present the first pipeline combining the two variant discovery strategies. The pipeline supports parallel processing of multiple WES data and individual chromosomes in computer clusters, using mapping-based methods, chromosome-level de novo assembly, or both. The mapping-based module is designed for sensitive detection of SNPs and small INDELs, using internally assessed tools. It comprises a two-tiered mapping with BWA and GSNAP, followed by a joint calling using GATK UnifiedGenotyper, HaplotypeCaller (HC) and Platypus. The de novo assembly module, which was built on top of the String Graph Assembler framework to achieve haplotype assembly, outperformed all five other de novo assemblers tested in this study. To demonstrate our strategy, we focused on the ~4-Mb human leukocyte antigen (HLA) region on chromosome 6p21; although this region is associated with over 100 diseases, it is a complex region with up to 10%-15% divergence across haplotypes and enriched for segmental duplications. We demonstrated the superiority of this pipeline in identifying SNPs, INDELs and complex variants from NA12878. The "true" call set was complied from high-confidence calls in 14 datasets and three lists from 250-bp paired-end WGS data. In detecting known SNPs from the two 150-bp datasets, it showed 98.5% and 98.6-99.1% sensitivity in the HLA and non-HLA regions, respectively, versus 87.6-87.9% and 97.0-97.4% by the widely used BWA-MEM+GATK HC; comparable results were obtained for the three 100-bp datasets. Moreover, the pipeline can detect large INDELs up to 1 kb and provide haplotype information. Thus, our pipeline should be particularly useful for cataloguing genetic variants genome-wide.

1801W

CDS-KB: Shared knowledge to support last-mile clinical decision support implementation for precision medicine programs. *J. Denny[1], J. Kirby[1], L. Rasmussen[2], P. Dexter[3], S. Ellis[4], A. Hartzler[5], C. Overby[6], P. Caraballo[7], M. Levy[1], P. Peissig[8], B. Phelp[1], J. Starren[2], M. Williams[9], P. Harris[1], J. Peterson[1].* 1) Vanderbilt University Medical Center, Nashville, TN; 2) Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Indiana University School of Medicine, Indianapolis, IN; 4) Icahn School of Medicine at Mount Sinai, New York, NY; 5) Group Health Research Institute, Seattle, WA; 6) University of Maryland School of Medicine, Baltimore, MA; 7) Mayo Clinic, Rochester, MN; 8) Marshfield Clinic Research Foundation, Marshfield, WI; 9) Geisinger Health System, Danville, PA.

Clinical decision support (CDS) based on genomic knowledge and genomic data is vital for creating and maintaining a precision medicine program. Several genomic medicine initiatives have targeted improving standardized rule sets for CDS implementation[1,2], and while standards assist in the transportability of CDS rules and knowledge, development of CDS often requires significant local customization. CDS customization may be required to fit into existing EHR systems infrastructure, to adapt hospital specific workflows or to speak to content knowledge of the target audience. The scale and complexity of dense genomic data further highlights the difficulty in recreating a program design, CDS rule set, and EHR interfaces for health systems considering precision medicine programs[3]. In early 2015, representatives of the Electronic Medical Records and Genomics (eMERGE) and Implementing Genomics in Practice (IGNITE) Networks targeted the gap of last-mile implementation issues by developing a lightweight repository to disseminate unstructured Genomic Medicine implementation artifacts. The Clinical Decision Support KnowledgeBase (CDS-KB) hosted at cdskb.org was created to promote sharing implementation experiences at different sites. Since CDS-KB's creation, 55 artifacts have been shared from twelve institutions ranging from workflow documents, architecture diagrams, to template text. Artifacts range from CDS specific drug, gene, or disease rules to general system documents. A public webinar series is a key component of the CDS-KB "Learning" section and facilitates exchanges and discussions between implementers across institutions. A repository of implementation artifacts is a potentially valuable resource for those health systems designing and implementing a new genomic medicine program. Shared design and implementation documentation may lower the last mile implementation barriers and help increase the quality and adoption of genomic CDS. 1 Relling M, Klein T. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* 2011;**89**:464–7. doi:10.1038/clpt.2010.279 2 DIGITizE: Displaying and Integrating Genetic Information Through the EHR : Health and Medicine Division. http://www.nationalacademies.org/hmd/Activities/Research/GenomicBasedResearch/Innovation-Collaboratives/EHR.aspx (accessed 1 Jun2016). 3 Herr T. Practical considerations in genomic decision support: the eMERGE experience. *J Pathol Inform*.

1802T

Pharmacogenomics Clinical Annotation Tool (PharmCAT). *T.E. Klein[1], M. Whirl-Carrillo[1], R.M. Whaley[1], M. Woon[1], K. Sangkuhl[1], H.M. Dunnenberger[2], P.E. Empey[3], A.T. Frase[4], R.R. Freimuth[5], A. Gaedigk[6], A. Gordon[7], C. Haidar[8], J.K. Hicks[9], J.M. Hoffman[8], M.T. Lee[10], N. Miller[11], S.D. Mooney[12], T.N. Person[13], J.F. Peterson[14], M.V. Relling[8], S.A. Scott[15], G. Twist[11], A. Verma[13], M.S. Williams[10], C. Wu[16], W. Yang[8], M.D. Ritchie[4,13].* 1) Dept Genetics, Stanford Univ, Stanford, CA; 2) Center for Molecular Medicine, NorthShore University Health System, Evanston IL; 3) Department of Pharmacy and Therapeutics, School of Pharmacy, University of Pittsburgh; 4) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 5) Department of Health Sciences Research, Mayo Clinic, Rochester MN; 6) Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy-Kansas City, Kansas City, MO; 7) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 8) St. Jude Children's Research Hospital, Memphis, TN; 9) DeBartolo Family Personalized Medicine Institute, H. Lee Moffitt Cancer Center, Tampa, FL; 10) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 11) Center for Pediatric Genomic Medicine, Children's Mercy, Kansas City, MO; 12) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 13) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 14) Vanderbilt University Medical Center, Nashville, TN; 15) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029; 16) Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N Torrey Pines Rd, La Jolla, CA 92037.

Pharmacogenomics (PGx) decision support and return of results is an active area of genomic medicine implementation at many health care organizations and academic medical centers. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has established guidelines surrounding gene-drug pairs that can and should lead to prescribing modifications based on genetic variant(s). One of the challenges in implementing PGx is extracting genomic variants and assigning haplotypes (including star-alleles) from genetic data derived from sequencing and genotyping technologies in order to apply the prescribing recommendations of CPIC guidelines. In a collaboration between the PGRN Statistical Analysis Resource (P-STAR), The Pharmacogenomics Knowledgebase (PharmGKB), the Clinical Genome Resource (ClinGen), and CPIC, we are developing a software tool to extract all variants from CPIC level-A genes with the exception of *G6PD* and *HLA*, from a genetic dataset resulting from sequencing or genotyping technologies (represented as a .vcf), interpret the variant alleles, infer diplotypes, and generate an interpretation report based on CPIC guidelines. The CPIC pipeline report can then be used to inform prescribing decisions. We assembled a focus group of thought leaders in PGx to brainstorm the issues and to design the software pipeline. We hosted a one-week Hackathon at the PharmGKB at Stanford University to bring together computer programmers with scientific curators to implement the first version of this tool. Through this process, we have uncovered many of the challenges surrounding PGx implementation. For example, the inference of diplotypes is challenging for several CPIC level-A genes. This software pipeline will be made available under the Mozilla Public License (MPL 2.0) and disseminated in Github for the scientific and clinical community to test, explore, and improve. PharmCAT will provide a solution that will enable sites implementing PGx a way to more consistently interpret genomic results and link those results to published clinical guidelines. Furthermore, we are assembling (and will be maintaining) the translation tables that underlie the tool, which will significantly reduce the effort required to implement PGx clinically and ensure more uniform interpretations of PGx knowledge. As precision medicine continues to move into clinical practice, implementation workflows for PGx, like PharmCAT, would enable standardized and consistent implementation of PGx genes.

1803F

Identifying structural variants of *CYP2D6* **with massively parallel sequencing.** *S. Lee, M.M. Wheeler, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

Cytochromes P450 (CYPs) are an important class of enzymes that metabolize thousands of endogenous and exogenous chemicals. CYP2D6, a member of the CYP family, is responsible for the activation and elimination of approximately 25% of prescription drugs, including many antidepressants, opioids, and chemotherapeutics. The abundance and efficiency of CYP2D6 enzyme varies considerably between individuals, raising the individual risk of decreased drug efficacy or increased side effects. Genotyping of the *CYP2D6* gene is performed in clinical settings to prevent adverse drug reactions. However, accurate genotyping of *CYP2D6* is challenging because it is extremely polymorphic. There are more than 100 known alleles for *CYP2D6* and these are characterized by single nucleotide variants (SNVs), insertion-deletions (indels), and complex structural variants (SVs). The SVs are facilitated by the high sequence identity between *CYP2D6* and its neighboring, non-functional paralog, *CYP2D7*. Although several methods (e.g., qPCR, FISH, and aCGH) can detect the SVs in *CYP2D6*, they are time-consuming and largely dependent on known DNA variants. With the advent of low-cost and high fidelity DNA sequencing, it is now possible to genotype *CYP2D6* in a high throughput and unbiased manner. To achieve this, we have developed a *CYP2D6* genotyping strategy that uses next generation sequencing and a customized analysis pipeline. Our analysis pipeline accounts for presence of paralog genes and computes paralog-spcific copy numbers based on the read depth. We validated the performance of this pipeline using 32 trios from the HapMap project sequenced by a targeted capture sequencing platform. Using this data set, we studied the inheritance of SVs from parent to offspring and compared these results to previous genotyping efforts that employed different analytic strategies. Within each trio, we were able to detect expected inheritance patterns of *CYP2D6* alleles. In addition, our results show high concordance between our genotype calls and those reported previously. We are currently analyzing a larger whole-genome data set to show that our method can be generalized to different types of next generation sequencing data and further validate the accuracy of this approach. Overall, our customized analysis enhances SV detection in *CYP2D6* and adds to our understanding of the genetic variation in this clinically relevant gene..

1804W

Target Gene Explorer: A durable and dynamic workbench connecting genetics and drug discovery. *M.P. Reeve[1], J. Wierzbowski[1], A. Kirby[1], S. Ripke[2,3], H. Huang[2,3], J. Heymont[1], M. Daly[2,3], J. Hutz[1].* 1) Human Biology & Data Sciences, Eisai, Andover, Ma; 2) Analytic and Translational Genetics Unit, MGH, Boston, MA; 3) Broad Institute, Cambridge, MA.

While genetic association studies have generated an avalanche of potential clues to the causal mechanisms of many human diseases, actionable therapeutic hypotheses have been slow to emerge from these studies owing to the challenges in the statistical and functional interpretation of such results. Many disparate genome browsers and databases provide access to unfiltered genome annotations and genetic results, but there is currently no way to maintain group curation of relevant information integrated with proprietary data. We introduce an electronic workbench to facilitate the comprehensive organization of results from GWAS and variant-based functional assays for the purposes of assisting therapeutic target evaluation and creation of durable institutional or public knowledgebases.   With a gene (potential drug target) as the focal point, the software identifies candidate published associations, introducing an adjudication system for relevance that allows the user to review and store an assessment of the finding. Qualitative assessments are pre-computed through heuristics designed to assess statistical robustness and variants' known or predicted functional effects on genes in the region. The process relies on LD patterns across global populations derived from reference panels and derives 'credible result sets' for association (directly presented in papers or based on LD surrounding index SNPs reported in GWAS publications).   Public resources of relevance to target evaluation (e.g., variation patterns in ExAC, expression patterns in GTEx, established links and mutations relevant to rare disease) are presented and user annotations enabled to create durable and shareable conclusions. Target-specific results from exome sequencing and functional assays – whether proprietary internal data or public findings – are captured and tracked alongside reference variation patterns in the gene. The software will prompt users when new data or publications of potential relevance to the target validity are available, and facilitate the transfer and persistence of knowledge within a large group by providing a permanent record of the adjudication of such 'clues' and allowing re-review of underlying data at a later time if needed. Implementation is such that private instances can be shared across groups or sites as needed - potential users include pharma (for tracking knowledge of targets) and disease consortiums for collating results and interpretations.

1805T

RNA sequencing analysis identifies genes involved in the genistein induced bone differentiation in mouse osteoblastic cell line MC3T3-E1. *Y.S. Cho[1], M. Kim[2], J. Lim[1], J.-H. Lee[1], K.-M. Lee[2], C.W. Nho[2].* 1) Biomedical Science, Hallym University, Chuncheon, Gangwon-do, South Korea; 2) Natural Products Research Center, KIST Gangneung Institute of Natural Products.

Genistein, structurally similar to 17β-estradiol, is a phyto-estrogen that may be beneficial for postmenopausal women as a substitute for the loss of endogenous estrogens. Although genistein is ingested in the diet and also widely consumed as dietary supplements, its underlying molecular mechanism that replaces endogenous estrogens in postmenopausal women is not fully understood. To gain insight into the effect of genistein on the bone differentiation, we tried to detect the differentially expressed genes by the treatment of genistein. To meet this objective, we conducted RNA sequencing analysis in genistein treated and untreated mouse osteoblastic cell line MC3T3-E1. After adaptor trimming and removing reads with poor quality (Q < 20), TopHat and Bowtie programs were employed to map RNA-seq reads to mouse reference genome version mm10. Genome-wide gene expression in the cells was normalized by total-count normalization (FPKM) and differentially expressed genes in genistein treated cells were identified by Cufflinks-Cuffdiff algorithm. Our analysis shows 50 up-regulated (including *CCL7*, *MMP13*, and *WISP2*) and 78 down-regulated genes (including *CAMK2B*, *CCL5*, and *CXCL9*) in genistein treated MC3T3-E1 cells (t-test p < 0.05 and 1.5-fold change). These results were validated by the real-time qPCR experiments. Pathway analysis demonstrates that these differentially expressed genes are highly involved in cytokine-cytokine receptor interaction, implying the role of genistein in this biological function.

1806F

Accounting for technical batch effects in single-cell RNA sequencing analysis. *C. Jia[1], M. Li[1], N.R. Zhang[2].* 1) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA; 2) Statistics Department, the Wharton School, University of Pennsylvania, Philadelphia, PA.

Recent technological breakthroughs have made it possible to measure RNA expression at the single-cell level, thus paving the way for exploring expression heterogeneity among individual cells. Current single-cell RNA sequencing (scRNA-seq) protocols are complex and introduce technical biases that vary across cells, and can lead to severe type I error inflation without proper adjustment. Compared to bulk RNA sequencing, in scRNA-seq the reverse transcription and preamplification steps could lead to dropout events and amplification bias, the former describing the scenario in which a transcript expressed in the cell is lost during library preparation and is thus undetectable at any sequencing depth. Due to the high prevalence of dropout events and amplification bias in scRNA-seq, it's crucial to account for them in data analysis. In order to adjust for these cell-to-cell technical differences, we propose a statistical framework, TASC (Toolkit for Analysis of Single Cell RNA-seq). TASC uses an empirical Bayes approach to reliably model the cell-specific dropout rates and amplification efficiency by use of external RNA spike-ins. Subsequently, TASC incorporates the estimated technical parameters, which reflect cell-to-cell batch effects, into a hierarchical mixture model to estimate the biological variance of a gene and detect differentially expressed (DE) genes. More importantly, TASC is able to adjust for additional covariates to further eliminate confounding that originates from cell size and cell cycle differences. In our simulations based on real scRNA-seq data, TASC displays superior sensitivity and specificity in detection of DE genes. The higher sensitivity of TASC is more pronounced when fold change is moderate; for example, when fold change is 1.75, at the 0.0001 significance level, the average power of TASC is 25%, 37%, and 428% higher than DESeq2, MAST and SCDE, respectively. Moreover, TASC is robust in the presence of batch effects. In contrast, DESeq2, MAST and SCDE yield highly inflated type I error rates. In real data analyses, TASC discovers the largest number of DE genes, and its performance is consistent across various sample sizes. TASC is programmed to be highly efficient computationally, taking advantage of multithreaded parallelization. A typical run on a dataset of 100 cells and 4,000 genes takes approximately 45 minutes using 24 cores. We believe that TASC will provide a robust platform for researchers to leverage the power of scRNA-seq.

1807W

Fleximer: Accurate quantification of RNA-seq via variable length k-mers. *C. Ju, R. Li, Z. Wu, Z. Yang, W. Wang.* Computer Science, University of California, Los Angeles, Los Angeles, CA.

The advent of RNA-Sequencing (RNA-Seq) has made it possible to quantify transcript expression in a large scale simultaneously. This technology generates small fragments of each transcript sequence, known as sequencing reads. As the first step of the data analysis towards expression quantification, most of the existing methods align these reads to the reference genome or transcriptome to establish their origins. However, this read alignment process is computationally costly. Recently, a series of methods have been proposed to perform a lightweight quantification analysis in an alignment-free or pseudo-alignment manner. These methods utilize the k-mers idea, which are short consecutive sequences representing the signatures of each transcript, to estimate the relative abundance from RNA-Seq reads. Current k-mers based approaches make uses of a set of fixed size k-mers; however, the true signatures that can distinguish one transcript from another may not exist in a fixed size. In the work, we demonstrate the importance of k-mer selection in transcript abundance estimation. We propose a novel method, Fleximer, to efficiently discover and select an optimal set of k-mers with variable sizes. Using both simulated and real datasets, the preliminary results have demonstrated that Fleximer is capable to achieve higher prediction accuracy with a smaller number of k-mers. These k-mers with flexible lengths are able to cover more reads in the datasets, and thus substantially reduce the percentage errors in transcript abundance estimation.

1808T

Association between telomere length and SNP array probe intensities in a cohort of 100,000 subjects. *M. Kvale[1], K. Lapham[1], J. Lin[1], E. Blackburn[3], P. Kwok[1], C. Schaefer[2], N. Risch[1].* 1) Institute for Human Genetics, UCSF, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Salk Institute, La Jolla, CA.

The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort contains data on telomere length and Affymetrix Axiom SNP array genotypes for each of 100,000 subjects. In a GWAS of relative telomere length and genotypes in this cohort, a large number of associations were found significant at the p<5e-8 level, and several with extreme statistical significance. To evaluate the veracity of these genetic associations, an investigation was undertaken of genotype call quality, one aspect of which was the underlying SNP array probe intensities. In a GWAS of relative telomere length with the log probe intensity contrast (log(A)-log(B)), thousands of associations significant at the p<5e-8 level were found, revealing an excess correlation of telomere length with probeset intensity beyond what would be expected from purely genetic effects. The correlation of telomere and genotype assay data beyond the putative genetic associations breaks a fundamental GWAS assumption of the independence of phenotype and genotype assays. To reduce the effects of probe intensity artifacts while maintaining genotype information, we developed a novel intra-cluster regression analysis for the GWAS analysis. By this approach, we were able to effectively eliminate the large number of artifactual GWAS associations due to probe intensity correlation that were not due to genotype, while retaining robust SNP genotype associations. This led to the elimination of 115 artifactually significant SNPs. Our final GWAS identified 25 genome-wide significant SNP association regions, 21 of which have been previously reported and 4 of which are novel.

1809F

Identification of differentially expressed genes in biased expression data based on RNA-Seq. *J. Sun[1], M. Tang[2], K. Shimizu[1], K. Kadota[1].* 1) The University of Tokyo, Tokyo, Japan; 2) Gertrude H. Sergievsky Center, Columbia University, New York, USA.

Identification of differentially expressed genes (DEGs) under different experimental conditions is one of the fundamental tasks in RNA-Seq study. The goal of DEG identification is to determine candidate genes for further analysis as well as to provide evidences to reveal biological process. edgeR and DESeq2 are two well-known software packages for DEG identification. Generally, DEG identification consists of two steps: normalization and hypothesis test, and improvement of any step leads to increase accuracy and specificity in DEG identification. We have proposed a robust normalization strategy, called DEG elimination strategy (DEGES), to improve the normalization step. Overexpressed gene in one of the compared conditions (i.e., biased expression) is known to compromise normalization. We introduced procedures to remove bad influences affected by genes with biased expression into normalization steps. Furthermore, we implemented DEGES-based normalization into TCC, an R/Bioconductor package that aims for DEG identification. According to simulation studies, we found that TCC outperformed edgeR and DESeq2 in two-condition comparison as well as ANOVA-liked multi-condition comparison. We believe that identifying of DEGs using TCC contributes higher reproducibility given that TCC provides high accuracy and specificity in DEG identification. The latest version of TCC is available on Bioconductor repository (http://www.bioconductor.org/packages/release/bioc/html/TCC.html).

1810W

DNASTAR software for accurate variant detection and interpretation using a variant annotation database. *M. Keyser[1], K. Maxfield[1], T. Schwei[1], T. Durfee[1], D. Nash[1], S. Baldwin[1], R. Nelson[1], K. Dullea[1], E. Edlund[1], A. Pollack-Berti[1], J. Stieren[1], F. Blattner[1,2,3].* 1) DNASTAR, Inc., Madison, WI; 2) University of Wisconsin, Department of Genetics, Madison, Wisconsin, USA; 3) Scarab Genomics LLC, Madison, Wisconsin, USA.

DNASTAR offers an integrated software suite for assembly and analysis of next-generation sequence data from all major sequencing platforms, supporting key workflows in a desktop computing environment. The SeqMan NGen program provides extremely accurate alignment and variant calling, while the integrated Variant Annotation Database makes interpreting the results feasible via automated import of variant annotations sourced from dbNSFP, 1000 Genomes Project and the Exome Variant Server. Here we present an example of how the software can be used to quickly identify variants in cancer causing genes from a whole exome data set. By utilizing this DNASTAR workflow, users can easily align large NGS data sets and quickly identify interesting variants using live, interactive views and comprehensive filtering tools. Alignment and analysis is scalable from individuals to large groups of individuals so that variant and gene sets of interest can be efficiently isolated and identified no matter the project size.

1811T

Models of single-sample and multisample SNP calling. *A. Pourshafeie[1], S. Shringarpure[2], C.D. Bustamante[2].* 1) Department of Physics, Stanford University, Stanford, CA; 2) Department of Biomedical Data Sciences , Stanford University, Stanford , CA.

Variant calling is an important step in the analysis of Next Generation Sequencing (NGS) data, since it affects the results of downstream analyses. Single-sample and multi-sample calling are the main methods for SNP discovery from genomes of many individuals. While single-sample calling is efficient and parallelizable, it is limited to extracting information from a genome of a single individual. Multi-sample calling was developed to address this shortcoming by sharing information across many samples. While many empirical observations of the characteristics of these callers exists, there are no models that describe the general behavior of these callers with varying coverage, sample size and minor allele frequency (MAF). We developed Bayesian models for single-sample and multi-sample SNP calling that characterizes the performance of these methods. We use our models to show the dependence of these methods on factors like coverage, sample size, and MAF. Using the SNP quality Phred score as a metric, we show that multi-sample quality score has a more direct dependence on SNP MAF than the single-sample quality score. We further explore our model using simulations and variant calls on a dataset with hundreds of individuals sequenced to 30x coverage. We compare our predictions for Phred quality score and concordance score to the real data. In particular, our model discusses the tradeoffs between coverage, number of individuals and error rates and explains the following previously reported empirical observations: 1) For low-coverage data, multi-sample calling has higher power than single-sample calling. 2) For high-coverage data, both methods have high power. 3) In large datasets, multi-sample calling has less power than single-sample calling for identifying variants with very low frequency (singletons etc.).

1812F

IOBIO Dev Kit: Resources for making genomic, real-time web applications and services. *C. Miller, A. Ward, Y. Qiao, T. DiSera, G. Marth.* Human Genetics, University of Utah, Salt Lake City, UT.

IOBIO (http://iobio.io) is an open-source web-based genomic platform enabling real-time analysis of large, remotely-stored, distributed datasets, with analysis algorithms operating as web servers on data-streams in standardized data formats. Several IOBIO apps have been previously released and can be found at http://iobio.io/applications.html. Here we introduce the IOBIO Dev Kit (https://developer.iobio.io), which provides libraries, examples, and documentation to build your own IOBIO apps and services. The dev kit consists of three main libraries: iobio.js is a client-side javascript library that streamlines creating and executing iobio commands, while hiding complicated web-socket connection code; iobio.viz is a d3-based, streaming genomic visualization library that consumes iobio data streams; and minion is a server library that wraps (with only a few lines of code) standard command line tools and converts them into iobio web services that can be mixed and matched with all other iobio web services. As an instructive example, we have rebuilt our bam.iobio.io app (http://bam.iobio.io) using dev kit components with all code found here: https://github.com/chmille4/bam.iobio.io. Using the IOBIO Dev Kit, developers will be able to create new genomic analysis apps on the web at a fraction of the time it currently takes to build desktop applications: IOBIO will be no less than an ecosystem for genomic analysis.

1813W

Bayesian nonparametric isoform discovery and quantification of many RNA-seq samples. *D. Aguiar[1], L. Cheng[2], B. Dumitrascu[3], F. Mordelet[4], B.E. Engelhardt[1,5].* 1) Department of Computer Science, Princeton University, Princeton, NJ; 2) Department of Electrical Engineering, Princeton University, Princeton, NJ; 3) Lewis Sigler Institute, Princeton University, Princeton, NJ; 4) Institute for Genome Science and Policy, Duke University, Durham, NC; 5) Center for Statistics and Machine Learning, Princeton University, Princeton, NJ.

The human genome contains approximately 22,000 protein-coding genes, each composed of exons that may be spliced in multiple ways, leading to possibly many distinct RNA *isoforms*. This *alternative splicing* encourages molecular diversity and plays an important role in disease etiology. The goal of isoform inference methods is to characterize the exon composition of isoforms for each gene across multiple samples and to quantify sample-specific proportions of each isoform using mapped RNA-seq reads. However, isoforms are difficult to characterize from short-read RNA-seq data because they are often composed of different combinations of identical exonic subsequences of transcribed DNA. Many tools exist for isoform reconstruction and quantification but they have a number of drawbacks making their application ill-equipped for multisample RNA-seq data: (1) many quantification methods assume a high-resolution isoform reference, which is unreliable, in particular, for non-model organism genomes, rare tissues, or disease samples; (2) most methods consider a single sample in isolation, ignoring the sharing of isoforms across samples; and (3) many methods make technology dependent assumptions by controlling for specific biases that do not generalize across technologies or samples. To address these drawbacks, we develop BIISQ, a **B**ayesian nonparametric method for **I**soform discovery and **I**ndividual **S**pecific **Q**uantification from RNA-seq data. BIISQ assumes neither knowledge of the existing isoforms nor sample-specific covariates, but instead postulates an infinite admixture model to probabilistically assign the aligned reads into their isoform of origin while inferring isoform compositions across samples. Joint inference across many samples enables the sharing of statistical strength to better identify low abundance isoforms and allowing each nucleotide base to have an independent frequency allows BIISQ to account for technical biases. We develop efficient posterior estimation based on stochastic variational inference that enables the discovery of novel isoforms and characterization of sample-specific and global isoform distributions quickly. We validate BIISQ using RNA-seq simulations and long sequence read technologies and demonstrate superior precision and recall compared with several existing isoform reconstruction methods. Lastly, we apply BIISQ to a large-scale RNA-seq study, GEUVADIS, for validation and to identify novel isoforms and isoforms specific to populations.

1814T

LeafCutter: Annotation-free quantification and prediction of RNA splicing. *D.A. Knowles, Y.I. Li, J.K. Pritchard.* Stanford University, Stanford, CA.

We recently developed LeafCutter to study sample and population variation in intron splicing. LeafCutter identifies variable intron splicing events from short-read RNA-seq data and finds alternative splicing events of high complexity. Instead of focusing on exon inclusion as do most splicing quantification tools, we quantified RNA splicing by modelling what is removed in each splicing event. This approach obviates the need for transcript annotations and overcomes the challenges in estimating relative isoform or exon usage in complex splicing events.Using LeafCutter, we identified over three times as many genetic variants that impact RNA splicing (i.e. splicing QTLs, or sQTLs) compared to contemporary methods. Notably, we find that sQTLs are highly enriched among variants that are also associated with diseases including multiple sclerosis.Motivated by the connection between RNA splicing and disease, we next built a machine learning model to predict the effects of genetic variants on RNA splicing from DNA sequence alone. Inspired by the process by which the core spliceosome selects introns to be excised during pre-mRNA splicing, we model intron selection dependent on the relative strength of adjacent splice sites determined by sequence features. We learn a mapping from sequence to splice-site strength in a tissue-specific manner and predict alternative intron excision in a panel of tissues from GTEx. We outperform the state-of-the-art in predicting the effect of common and rare genetic variants in individuals with matched whole genome and RNA sequencing. Our model therefore opens the door to improved personal exome and genome interpretation, in particular for rare diseases where relevant tissues are difficult to sample. .

1815F

Annotation sources database management and genome annotations software workflow in Ion Reporter. *G. Altun, Y. Lou.* Thermo Fisher Scientific, South San Francisco, CA.

Ion Reporter™ software is a cloud-based software that can analyze next generation sequencing research data on local servers as well as on the cloud. Ion Reporter software not only finds data that is statistically significant for a given experiment, it also provides metadata such as annotations. The annotation workflow of Ion Reporter software has gained immense popularity in the past two years. The reasoning behind this is because the Ion Reporter annotation pipeline is very comprehensive and consists of significant number of pre-built private and public genomic and clinical research databases. The information from these databases are processed, optimized and stored in the Ion Reporter internal database. The Ion Reporter's internal database currently stores millions of records from multiple public and private datasets. These records are searched by the software within seconds to annotate statistically significant biological data such as SNP (Single-Nucleotide Polymorphism) mutations and CNVs (Copy-Number Variations). This presentation will provide more details Ion Reporter software features and capabilities that provide biologically meaningful results and help decipher the genomic data.

1816W

Functional validation of human protein-truncating genetic variants. *I.M. Armean[1,2], K.J. Karczewski[1,2], J.L. Marshall[1,2], B.B. Cummings[1,2], E. Minikel[1,2], D. Birnbaum[1,2], P. Singh[1,2], B. Weisburd[1,2], M. Lek[1,2], M.J. Daly[1,2], A. Palotie[1,2], S. Kathiresan[2,3], D.G. MacArthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.

Knockout model organisms have been used for decades to study the function of genes. The identification of so-called "human knockouts", individuals with both copies of a gene inactivated by null mutations, in different populations across the globe offers a unique opportunity for investigating gene function. Homozygous protein-truncating variants (PTV) are typically rare but are more often observed in populations with historical bottlenecks or high levels of consanguineous mating, which skew the frequency spectrum and induce homozygosity, respectively. In addition to providing information about the function of human genes, human knockouts can provide valuable information about the potential efficacy and toxicity of therapeutic inhibition of specific biological targets. Here we outline a pipeline for the functional validation of predicted protein-truncating genetic variants using both public databases and newly generated data. We first identify human knockouts from samples including over 60,000 exomes assembled by the Exome Aggregation Consortium (ExAC). Secondly, we describe approaches to filtering candidate PTVs using both an *in silico* pipeline, LOFTEE, and publicly available RNA-seq data from over 1,000 genome-sequenced individuals. Finally, we propose an approach to direct functional validation of candidate gene-disrupting PTVs using participant cell lines or CRISPR-engineered equivalents differentiated into relevant tissues. Finally, we describe a schema for the aggregation of these variants into a database of LoF variants, dbLoF, providing a resource for pharmaceutical development, transplant biology and understanding of rare Mendelian diseases.

1817T

Simple and accurate adjustment for winner's curse. *S. Bacanu[1], T.B. Bigdeli[1], V. Vladimirov[1], A.H. Fanous[1,2], K.S. Kendler[1].* 1) VIPBG, Virginia Commonwealth University, Richmond, VA; 2) VA Medical Center, Washington, DC.

Genome-wide SNP and sequencing scans, henceforth denoted as genome scans, have discovered/will discover numerous genetic variants significantly associated with various phenotypes. However, while significant signals explain only a small fraction of the variation in most traits, variants with smaller signals in aggregate explain a much larger part. To extract the signals of all magnitudes from a genome scan, e.g. by estimating their unbiased values, our group recently proposed a very simple method for winner's curse adjustment, which we called **F**DR **I**nverse **Q**uantile **T**ransformation (FIQT). The essence of this method was to perform the multiple testing adjustment on the p-value scale (i.e. FDR) and quantile transform it on the desired scale, e.g. mean Z-score, ORs or regression coefficients. Compared to competing algorithms, FIQT was more accurate in predicting the true values of effect sizes and, for large fraction of scenarios, its prediction was sensibly better. However, the method likely did not achieve its full accuracy potential due to its treatment of individual markers as independent. To remedy this shortcoming, we extended FIQT (eFIQT) such that it now takes into account the linkage disequilibrium (LD) between adjacent markers. At the price of a reasonable increase in computational power, eFIQT offers a meaningful increase in the already class-leading accuracy of FIQT.

1818F

Clinical Interpret, an application giving improved accuracy of interpretive analysis of variant and clinical data and providing novel e-diagnostics for future and deeper understanding of disease and its application in a clinical context. *L. Bean[1], J. Shovelton[2], L. Parkes[2], G. Alexandrou[2], D. Luerrssen[2], J. Holdstock[2], K. Ghandi[1], M. Hegde[1].* 1) Human Genetics, Emory University, Decatur, GA; 2) Oxford Gene Technology, Oxfordshire, UK.

Next-generation sequencing (NGS) promises to be the lynchpin for the delivery of precision and personalised medicine. Although NGS is quickly becoming the centre- piece assay in both clinical and research settings, this technology is not able to interrogate all clinically relevant areas of the genome. Genetic data management systems do not, however, incorporate data from all biological assays mainly due to varying data formats, variability in and complexity of interpretive analysis and difficulties of data management. Our solution addresses these difficulties, and aims to increase the ease and accuracy of interpretive analysis of genomic data for further and deeper understanding of disease and the application of genomics in a clinical and research context. Clinical Interpret is a clinician-friendly GUI-driven interpretation package for whole-genome scale genetic data. To this end, Clinical Interpret has been designed to import data from any number of popular tools including molecular, biochemical and cytogenetic assays to allow the user to perform a comprehensive analysis on all genetic evidence for a given case. It provides functionality for variant annotation by attaching meaningful information to variants such as gene name, predicted effect on the protein etc. Critically, information that may be clinically relevant, such as previously reported disease association segregation data, and population data is made visible to the end-user via its intuitive user interface. The software supports commonly used and custom variant filtering workflows thereby allowing users to employ analysis algorithms that are locked to ensure all analyses are reproducible, immutable and auditable. Integration with existing databases for genetic disease is a core part of the functionality. In addition, the software provides a mechanism to flag, store and collate information in a local database. The simple preparation of reports for clinical use is another key aspect that will enhances clinical interpretation for NGS data analysis. Furthermore, with all data being stored in an easily accessible medium Clinical Interpret provides a powerful tool for clinical research activities that complement the clinical diagnostic applications. In summary, Clinical Interpret provides a bridge between the biologists running the experiements, in clinical and research setting, the bioinformaticians processing the data and the clinicians delivering a clinical interpretation.

1819W

A sequencing deliverable: Theoretical sensitivity to heterozygous SNPs. *K. Bergin, D. Benjamin, J. Bloom, M. Costello, J. Rose, C. Tolonen, Y. Farjoun.* The Broad Institute of MIT and Harvard, Cambridge, MA.

Sequencing experiments can be characterized by the quality, quantity, and uniformity of the resulting reads. However, in many experiments, the ultimate metrics of interest are the sensitivity and specificity to detect genetic variants. Here we derive and implement a theoretical measure of sensitivity to heterozygous Single Nucleotide Polymorphisms (SNPs) within a widely used software package (Picard). In particular, we estimate the probability of calling a true heterozygous site with confidence greater than a threshold. The confidence is defined as the likelihood ratio between the heterozygous model and the homozygous-reference model for the data. Our estimate is based on the distributions of read-depth and base-quality within the region of interest. This metric is shown (on over 2500 libraries) to be a good predictor of the empirical sensitivity (found by variant calling with HaplotypeCaller) for various data types and qualities on a well-characterized sample (NA12878). In practice, this metric provides a more comprehensive view of the quality of sequencing data and a more actionable deliverable than coverage and base quality on their own..

1820T

Linear mixed models for quantitative traits in health-system scale data. *M. Bilow, P. Zhou, Y.-T. Chen, J. Cong, E. Eskin.* UCLA, Los Angeles, CA.

Modern health systems and numerous national and international projects have ascertained genetic information on tens or hundreds of thousands of individuals. Linear mixed models are a standard tool for determining association between genetic data and quantitative traits; however, performing linear mixed model analysis of quantitative traits becomes computationally intractable when the number of individuals rises into the tens of thousands. The most computationally-intenstive step of performing linear mixed model analysis involves performing a singular value decomposition; this process scales quadratically with the number of individuals. This significant time complexity combined with the challenge of manipulating input data makes performing mixed-model analysis at scale very challenging. Using Apache Spark and the MLib librarires, we demonstrate a performant and easy-to-apply method capable of performing end-to-end linear mixed model analysis at scale.

1821F

Phasing whole genome sequencing data in pedigrees utilizing lineage specific alleles. *A. Blackburn, J. Blangero, H.H.H. Göring.* South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.

Phasing genotype data is useful for enabling haplotype specific genetic analyses, identification of compound heterozygosity, improving genotype error checking, imputation of missing genotypes, and it can be used to reduce the multiple testing burden inherent to the whole genome sequencing approach by applying segment based tests. Phasing in pedigrees using whole genome sequence (WGS) data is an attractive proposition, but current methods are limited by pedigree size, by the distance of relationships, by computation time, and/or by robustness to missing individuals. Within pedigrees, alleles carried by a single founding chromosome, which we call lineage specific alleles, are highly informative for identifying DNA segments that are identical-by-descent. Here we present software designed to phase WGS genotypes for whole chromosomes within pedigrees based on this rationale. Our method is capable of phasing any size pedigree, from trios and nuclear families to large multigenerational pedigrees across a range of levels of non-sequenced individuals, using only data inherent to the pedigree dataset itself. We have compared the performance of our software to AlphaPhase and SHAPEIT (utilizing the duoHMM option) using simulated data. WGS genotypes were simulated using male X-chromosomes from the 1000 genomes project to represent chromosomes among pedigree founders and gene dropping. Our software consistently produces the lowest switch errors rates in these simulations across a variety of scenarios. The percentage of heterozygous genotypes our software phases varies based on pedigree size and the percentage of individuals that are sequenced within the pedigree, whereas SHAPEIT consistently makes a phasing call for every heterozygous genotype, albeit with a somewhat higher (though still very accurate) switch error rate. In large pedigrees, filling in the gaps of the phasing results generated by our software with the results from SHAPEIT performs better than either method alone, and should be considered the current standard bearer for phasing WGS data in pedigrees.

1822W

Findings from the Fourth Critical Assessment of Genome Interpretation (CAGI), a community experiment to evaluate phenotype prediction. *S.E. Brenner[1], R.A. Hoskins[1], J. Moult[2], CAGI Participants.* 1) University of California, Berkeley, CA; 2) IBBR, University of Maryland, Rockville, MD.

The Critical Assessment of Genome Interpretation (CAGI, 'kā-jē) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. CAGI participants are provided genetic variants and make predictions of resulting phenotype. These predictions are evaluated against experimental characterizations by independent assessors. The fourth CAGI experiment recently concluded. It included 11 challenges which reflected: non-synonymous variants and their biochemical impact measured by targeted assays; noncoding regulatory variants and their impact on gene expression; research exomes for prediction of complex traits; personal genomes and trait profiles; and clinical sequences and associated referring indications. There were notable discoveries throughout the CAGI experiment, and general themes emerged. The independent assessment found that top missense prediction methods are highly statistically significant, but individual variant accuracy is limited. Moreover, missense methods tend to correlate better with each other than with experiments (for reasons that may reflect the predictive methods and the assays themselves). Structure-based missense methods excel in a few cases, while evolutionary-based methods have more consistent performance. Bespoke approaches often enhance performance. On the clinical studies, predictors were able to identify causal variants that were overlooked by the clinical laboratory, and it appears that physicians may not always order the most relevant genetic test for their patients. The results showed that predicting complex traits from exomes is fraught. Beyond this, creating a genetic study that provides a reliable gold standard is remarkably difficult. However, there were notable improvements in the ability to match genomes to trait profiles. Results from the recent CAGI experiment will be presented, exploring these topics and discussing how the field has advanced. Clinical implications will be discussed. Complete information about CAGI may be found at https://genomeinterpretation.org.

1823T

Fast genotype phasing in very large cohort. *R.C. Brown, G. Lunter.* Wellcome Trust Centre of Human Genetics, Oxford, Oxfordshire, United Kingdom.

Title: Fast genotype phasing in very large cohorts Haplotype phasing is an essential first step in genome-wide association studies and many other genetic analyses. The Li and Stephens model accurately describes the structure of genetic data in large populations, and underpins most existing algorithms for haplotype phasing in large populations. However, the full implementation of the Li and Stephens model as an HMM scales poorly, and existing algorithm rely on clever approximations and heuristic narrowing of the search space to achieve acceptable run-times. We present a haplotype phasing method that takes advantage of a novel and fast algorithm for calculating the exact maximum likelihood solution to the Li and Stephens model over large populations that we recently introduced [ref bioRxiv]. The algorithm is built on the positional Burrows-Wheeler transform to implement efficient haplotype searching on large-scale populationd data. On simulated data, the algorithm's run-time scales linearly with the number of polymorphic sites, but is independent of the cohort size, a highly desirable characteristic for very large cohorts. We present results on an initial implementation, which has allowed experimentation of cohorts in the tens of thousands, with the method expected to scale well to even larger sizes given its desirable complexity properties. Real world haplotypes obtained from the 1000 Genomes project, combined into new individuals were used initially to test the accuracy of the method. A section of chromosome 10, subset at sites common to those on an Affymetrix snp chip were extracted for experiment. Results obtained show runtimes and switch error rates that are comparable to the popular phasing tool SHAPEIT2. We expect that our approach will give superior performance for larger cohorts because of the favourable runtime behaviour.

1824F

Conform-gt: Software for harmonizing genetic marker data in VCF files. *B.L. Browning.* Medicine, Med Genetics, University of Washington, Seattle, WA.

Many genotype data sets cannot be easily combined because the chromosome strand or the VCF reference allele differs between data sets for some markers. Naïve use of inconsistent genotype data creates errors in downstream analyses such as genotype imputation or genetic association testing. The Conform-gt program is an open source software program that automatically re-orders marker alleles and flips marker chromosome strand when necessary so that marker and genotype data in a target VCF file are consistent with a reference VCF file. The software can match reference and target markers using either chromosome position or marker identifier. Matching by marker identifier is particularly useful when reference and target data are defined with respect to different genome builds. Once markers are matched, marker alleles for each target VCF record are updated to match the reference data using information from allele labels and statistical tests of allele frequency and allele correlation. The output log file reports harmonization data for each target marker, including whether the target marker could be matched to a reference marker, whether an input target marker was defined using the same chromosome strand as the reference marker, and the information that was used to harmonize the reference and target chromosome strand. The conform-gt program is freely available for download from https://faculty.washington.edu/browning/conform-gt.html.

1825W

Predicting the translational consequences of reading-frame changes induced by stop-lost and frame-shift variants. *M. Butkiewicz, J.L. Haines, W.S. Bush.* Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Annotation and prediction of the functional consequences of DNA sequence variants is gaining importance for human genomic studies. Variants that create a "Loss-of-Function" (LOF) are of special interest given their disruptive aspect towards protein sequence, and are presumed more likely to be pathogenic. At the sequence level, LOF typically implies the insertion of a stop-codon, disruption of a splice site, or insertions/deletions (InDels) leading to a derangement of the translational reading frame. Although affected transcripts may be subject to nonsense-mediated decay (NMD), recent studies show that a substantial portion of transcripts actually evade NMD, therefore the resulting effects on protein function are unclear. Current in silico annotation approaches, such as SNPEff or the ENSEMBL Variant Effect Predictor (VEP), only report the impact of a variant in the context of altered DNA sequence; information about changes in the resulting transcript, reading frame, or translated protein product is typically not provided. To alleviate these limitations, we introduce COCOS, the Codon Consequence Scanner, a plugin for VEP that provides new information about the altered reading frame and the translated protein sequence changes resulting from LOF variants. The new amino acid sequence stemming from variants that produce an altered reading frame, such as stop-lost variants and InDels, is generated until a subsequent stop-codon is reached. We applied COCOS to data from the 1000 Genomes project (phase 3), evaluating 1,283 stop-lost single nucleotide variants and over 2.5 million small InDels. 2244 variants (4,451 transcripts) with viable translation sequence alterations were identified. While these variants were previously predicted to be LOF, approximately 36% (815) of variants exhibit a benevolent character, only introducing 10 or less altered amino acids. Conversely, approximately 13% (291) of variants introduced a protein with at least 100 altered amino acids, with 7% (21/291) of these variants located in genes classified as tolerant to variation and associated with a pathological background; e.g. SRCAP (prostate cancer), AHNAK (Systemic lupus erythematosus), and SH3PXD2A (Alzheimer's disease). In conclusion, information provided by COCOS allows for a more detailed analysis of variant consequences with respect to the sequence of gene transcripts, and thus will guide the scientific community to place these sequence alterations into their biological context.

1826T

Copy number estimation from targeted and shallow sequencing. *Z. Che, A. O'Hara, R. Keshavan, W. Sherman, V. Wasnikar, S. Shams.* BioDiscovery, Inc., El Segundo, CA.

Targeted panel sequencing has been a popular method to achieve high depth of coverage for certain regions of interest at an affordable cost compared to whole genome sequencing. Shallow whole genome sequencing, where average read-depth can be as low as 0.1x, provides a cost savings-approach for identification of large copy number variant (CNV) events; it has been utilized in various application areas, including oncology and Pre-Implantation Genomic Screening (PGS). Depending on genomic coverage (whole genome vs. whole exome vs. targeted panel) and sample type (constitutional vs cancer), different algorithmic approaches may prove to be more or less ideal for estimating copy number from next generation sequencing (NGS) results. Some approaches require matched normal samples (e.g. ngCGH) while others use the characteristics of the sample itself for normalization (e.g. Wisecondor and QDNASeq). Other methods use a pool of reference samples (e.g. Pooled Reference, xHMM, and CONFIER). Here, we introduce a new approach (dCNVSeq) that extends the Pooled Reference algorithm to function with shallow, as well as targeted, sequencing data by introducing a novel dynamic binning approach. The approach uses a Hidden Markov Model to segment the genome into areas forming the "backbone" using the off-target reads and additional areas where targeted reads are present. dCNVSeq uses coarse binning in the backbone area providing base line copy number as well as detection of large CNV events and fine binning in targeted areas to provide high resolution CNV detection in targeted regions. We present a comparison of copy number estimation results from ngCGH, QDNASeq and dCNVSeq using WES, targeted panels, and shallow sequencing data.

1827F

Accurate identification of disease-specific non-coding risk variants based on multi-omics profiles. *L. Chen[1], P. Jin[2], Z. Q[3,4].* 1) Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; 3) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA; 4) Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA 30322, USA.

The majority of variants identified by Genome-wide association studies (GWASs) fall outside of the protein-coding regions. Understanding the cryptic link between non-coding sequence variants and pathophysiology of complex diseases is a fundamental challenge. To overcome the lack of annotations in the non-coding regions, various recent computational methods have been developed to identify non-coding risk variants using genome-wide genomic and epigenomic profiling data. A common feature of these methods is that they do not distinguish risk variants associated with different diseases. Since different biological mechanisms are believed to contribute to the etiologies of different diseases, it is desirable to characterize the impact of a non-coding variant in a disease-specific manner. In this work, we describe DIVAN, a data-driven machine learning approach that aims to identify disease-specific risk variants. Using 1,806 epigenomic profiles across cell types and factors, along with other static genomic features, we adopt a novel feature selection based ensemble-learning framework to achieve this goal. To evaluate the predictive performance, we first conduct comprehensive tests across 45 different diseases belonging to 13 different classes and compare DIVAN with multiple existing methods. We observe that the proposed method accurately discriminates (AUC from 0.65 to 0.88 with median 0.74 in cross validation) disease-specific risk variants from benign ones in non-coding regions and outperforms other competing methods. We further demonstrate DIVAN achieves better predictive performance than other competing methods in an independent testing dataset including 36 diseases. These results suggest that the DIVAN has the potential to identify novel variants in a disease-specific manner, which will be important in interpreting personal genome sequencing data. We use number of reads instead of peaks around the variants as feature in DIVAN in order to accommodate the scarcity of known risk variants for each disease. Our major finding is that, H3K9me3, a histone mark associated with repressed chromatin and depleted in most diseases, shows up as the most informative features in distinguishing risk variants from benign ones in 34 out of 45 diseases studied. Enriched in most diseases, open chromatin, defined by DNaseI and FAIRE, comes in second. Overall, histone marks contribute more to informative features than transcription factors and open chromatin in DIVAN. .

1828W

In silico variant effect prediction for whole exome analysis: Clinical caveats. *S. Chervitz, J. Anderson, D. Glazer, S. Galasinski, K.M. Chen, J. West, R. Chen.* Personalis, Inc., Menlo Park, CA.

Computational tools such as SIFT, PolyPhen and others provide in silico prediction of the pathogenicity of missense sequence variants based on sequence information. Some of these tools, such as SIFT and PolyPhen, have been in use since the early 2000's, before the completion of the human genome. Current versions of SIFT & PolyPhen2 have sensitivity ~70%, but they suffer from low specificity (~15%, Flanagan et al 2010). If ~1% of missense variants are truly damaging in a typical exome, then, given the 20% false positive rate for SIFT and PolyPhen2, such tools would incorrectly flag 19 of 20 missense variants as damaging, for a positive predictive value of only 5%. The clinical utility of these tools, given this performance is questionable at best. Efforts to improve the performance of in silico tools are on-going (Masica & Karchin 2016), but given their wide use in the clinical community, we need to recognize and understand the limitations of these pathogenicity prediction tools. Here, we assess the reliability of SIFT, PolyPhen2 and other variant effect predictions on a set of variants observed in clinical whole exome samples processed by Personalis, Inc. We examine the correlation between the damaging/tolerated calls of the tools compared to a reference set of clinical significance values assessed by genetic counselors using a variety of newer resources. We highlight the risk of misinterpretation of results from commonly used in silico tools, particularly for variants of uncertain significance (VUS), and underscore the need of the clinicians to understand the limitations of such tools when assessing variant results from whole exome analyses. The growing amount of publicly available data from individual whole exome and whole genome sequences is increasingly making the use of in silico tools obsolete for variant classification. Resources from such new data include allele frequencies in control and patient populations and case observations. Growing use of transgenic animal models as well as the improvement and wider usage of in vitro functional studies will also increase the number of useful resources available for classification. **References**: Flanagan et al. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers 14(5):730. Masica DL, Karchin R (2016) Towards Increasing the Clinical Relevance of In Silico Methods to Predict Pathogenic Missense Variants. PLoS Comput Biol 12(5): e100472 .

1829T

LocusZoom.js: Web-based plugin for interactive analysis of genome and phenome wide association studies. *C.P. Clark, M. Flickinger, R. Welch, P. VandeHaar, D. Taliun, M. Boehnke, G. Abecasis.* Biostatistics, University of Michigan, Ann Arbor, MI.

Genome-wide association studies (GWAS) have revealed hundreds of loci associated with common human genetic diseases and traits. Phenome-wide association studies (PheWAS), wherein each genetic variant is analyzed against a broad range of phenotypes, provide a different cross-sectional view over the same genotype data sets. In order to facilitate understanding these association results, data visualizations must be generated that combine association results with relevant context such as local linkage disequilibrium (LD), recombination patterns, and the positions of genes in the region. The original LocusZoom, a web-based plotting tool for fast visual display of GWAS results, targeted this problem by allowing users to upload data, specify region(s) of interest, and generate static visualizations. Since its inception in 2010 it has generated over 500,000 plots for use in presentations, publications, and other research settings. We have developed a new implementation of LocusZoom targeted at interactively visualizing both GWAS and PheWAS results: LocusZoom.js. This extensible open source JavaScript framework generates plots supporting many new features such as real time rearrangement, elastic scaling, simultaneous display of multiple analyses, dynamic inclusion and removal of analyses, and interactivity to afford deeper study (for example: conditional analysis). Plots can be generated by loci within the genome, providing panning and zooming capabilities to explore nearby loci, or plots can be generated using any other value for the domain: for example, plotting by grouped phenotypes results in a PheWAS. Publication-quality static snapshots of interactive plots can be easily exported. LocusZoom.js was designed as a general module for generating interactive visualizations of genetic data for a wide variety of use cases. It has been implemented in the Accelerating Medicines Partnership (AMP) Type 2 Diabetes Knowledge Portal for multiple phenotype analysis and PheWeb, a Michigan Genomics Initiative resource for phenome-wide association studies. Source code and user documentation can be accessed at github.com/statgen/locuszoom. Functional versions of original LocusZoom and new LocusZoom.js can both be accessed at http://locuszoom.sph.umich.edu.

1830F

mini-mitoCalc: **An ultra-fast program to estimate mitochondrial DNA copy number from whole-genome sequences.** *J. Ding[1], Y. Qian[1], C. Sidore[2], R. Nagaraja[1], F. Cucca[2], G.R. Abecasis[3], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy; 3) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

The mitochondrial cellular power plant is partially encoded in its own constituent circular genome of ~16kb. In humans, every cell has 100-10,000 copies of the mitochondrial DNA (mtDNA). The copy number is tightly regulated in various tissues, and is both a critical determinant of the level of mitochondrial function and a potential biomarker for diseases. With the collection of whole-genome sequencing data in various large-scale genetic studies, we and others have shown that the mtDNA copy number per cell can be directly estimated from whole-genome sequencing. The computation is based on the rationale that average sequencing coverage should be proportional to underlying DNA copy number for autosomal and mitochondrial DNA. Because the mitochondrial genome is comparatively quite small, most computing time is spent calculating the average autosomal DNA coverage across all ~3 billion bases. That makes investigation of the tens of thousands of samples already available in the public domain very slow. Here we present "*mini-mitoCalc*", which is >100 times faster than conventional programs including our published program "*mitoCalc*". It takes advantage of the indexing of sequencing alignment files, focusing on a selected subset of the nuclear genome to estimate autosomal DNA coverage accurately (correlation > 0.999 with the full genome estimate). Consequently, a computer cluster with 50 CPUs can now finish analyzing low-pass sequencing (a 4X average nuclear DNA coverage) of 10,000 samples in 3 hours rather than the 17 days required by the original program. As a result, analyses can be conducted efficiently to test for association of mtDNA copy number with quantitative traits or to identify any variants that regulate mtDNA copy number using large-scale consortium data.

1831W

Combined prioritization of genetic variants using the Variant Effect Scoring Tool (VEST). *C. Douville[1], D.L. Masica[1], P.D. Stenson[2], D.N. Cooper[2], D.M. Gygax[3], R. Kim[3], M. Ryan[3], H.K. Carter[4], R. Karchin[1,5].* 1) Biomedical Engineering and Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD; 2) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 3) In Silico Solutions, Fairfax, VA; 4) Division of Medical Genetics, University of California, San Diego, CA; 5) Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD.

The average genome contains tens of thousands of single nucleotide variants (SNVs) as well as hundreds of insertion/deletion (indel) variants that alter protein sequence. Computational tools for most variant types have been developed to assess their potential impact on human health. Most computational tools, however, can only prioritize one type of genetic variant. Here, we present the Variant Effect Scoring Tool (VEST) [1,2], a method that jointly prioritizes all non-silent variant types: missense, in-frame and frameshift indels, splice site, nonsense and nonstop. VEST is trained on pathogenic variants from the Human Gene Mutation Database [3] and polymorphisms from the Exome Sequencing Project [4]. It uses a mix of predictive features that are shared between variant consequence types and those that are unique to a single variant consequence type. Scores for different consequence types are integrated using characteristic analytical null distributions, so that p-values are comparable across types. To our knowledge, the only other method for joint prioritization of multiple variant consequence types is Combined Annotation Dependent Depletion (CADD) [5]. We use a hold-out validation protocol to compare VEST and CADD. For a variant set dominated by missense VEST ROC AUC=0.9 and CADD ROC AUC=0.88. For a variant set where missense and indels have equal portions VEST ROC AUC=0.91 and CADD ROC AUC=0.74. VEST 3.0 is available at http://cravat.us and https://hub.docker.com/r/karchinlab/cravat-mupit/. [1] Carter, Hannah, et al. "Identifying Mendelian disease genes with the variant effect scoring tool." *BMC genomics* 14.3 (2013): 1. [2] Douville, Christopher, et al. "Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST□Indel)." *Human mutation* 37.1 (2016): 28-35. [3] Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* **133**:1–9. [4] Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**:216–220. [5] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**:310–315.

1832T

ClinGen's gene and variant curation interface suite: Centralized and consistent evaluation of the clinical relevance of genes and variants. *S. Dwight[1], M. Wright[1], K. Dalton[1], K. Liu[1], M. Choi[1], F. Tanaka[1], J. Zhen[1], D. Azzariti[2], L. Babb[2], C. Bizon[3], M. Distefano[2], S. Harrison[2], R. Patel[4], P. Pawliczek[4], E. Rooney Riggs[5], N. Strande[6], A. Buchanan[5], R. Ghosh[4], J. Berg[6], C. Bustamante[1], C. Martin[5], A. Milosavljevic[4], S. Plon[4], H. Rehm[2], J.M. Cherry[1], the ClinGen Consortium.* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Partners HealthCare Laboratory for Molecular Medicine, Cambridge, MA; 3) Renaissance Computing Institute, University of North Carolina at Chapel Hill, NC; 4) Baylor College of Medicine, Houston, TX; 5) Geisinger Health System, Danville, PA; 6) University of North Carolina at Chapel Hill, NC.

Critical to the field of clinical genomics is an understanding of the strength of correlation between genotype and phenotype. The central mission of ClinGen (clinicalgenome.org) is to define the "clinical relevance of genes and variants for use in precision medicine and research." Inherent to this mission is the ability to access and evaluate evidence for a gene or variant's role in disease in a consistent, efficient manner. As such ClinGen is building a suite of interconnected interface tools, one for gene curation and the other for variant curation. These tools are currently accessible to ClinGen curators and are also being designed and developed for use by the broader community. The ClinGen gene and variant curation interfaces have been developed according to the following important specifications: 1) gene curation will follow the Clinical Validity Classifications framework established by ClinGen's Gene Curation Work Group while variant curation will follow the ACMG Standards and Guidelines, 2) both interfaces will centralize evidence from relevant resources in order to enable efficient, consistent curation, 3) curated evidence and evidence retrieved from external resources will be shared between the gene and variant curation tools, 4) the interfaces will be designed to guide curators through the curation process, 5) controlled vocabularies and ontologies will be used in order to promote the capture of discrete evidence, facilitate connections, and promote consistency, 8) the ClinGen Allele Registry will be used to normalize variants and provide an identifier for ClinVar submission, 7) external and curated evidence will be viewable by all curators, while curated evidence can only be edited by its creator, 8) the interfaces will support expert review of provisional classifications and interpretations through access to all evidence, 9) contextual help and documentation will be included to assist the curator within the interface, and 10) JSON-LD will be used to promote rich relationships within these data. The ClinGen gene curation interface is currently in production and actively supports ClinGen curation while an early version of the ClinGen variant curation interface is in production and undergoing rapid and iterative update cycles. The rich evidence captured by both tools exists in a structured format that allows it to be accessible on ClinGen's public portal (clinicalgenome.org).

1833F

***Ella*: Decision support and structured data in clinical variant classification.** *M. Eike, S.T. Seljebotn, E. Nafstad, T. Håndstad, H. Fontenelle, D.J. Honneffer, E. Severinsen, T. Grünfeld, D.E. Undlien.* Dept of Medical Genetics, Oslo University Hospital, Oslo, Norway.

We have made a software tool named *ella* to aid and improve the quality and reproducibility of clinical interpretations of genetic variants, a core task for genetic laboratories. Clinical classification is usually done according to procedures that detail how different types of evidence should be assessed and combined, and many have adopted the revised ACMG guidelines for variant interpretation (Richards et al., 2015). The conclusion, however, still very much depends on the experience of the interpreter and their subjective judgement. With the increased volume from high-throughput sequencing, this quickly becomes a daunting task. Moreover, the documentation often involves a high degree of unstructured data in the form of free text comments and/or literature references, which makes automated decision support and quality control difficult to achieve. In response to these challenges, we have created *ella*, a web-based software tool that guides the interpreter through a series of steps where different types of information are presented for assessment. A major step involves evaluating literature references through a dynamic questionnaire, which translates key evidence from the reference into structured data. Together with a set of configurable rules, this enables the software to automatically suggest evidence categories based on ACMG guidelines for both various types of annotation and evaluated literature, as well as a clinical classification. As a result, the reasoning behind each evaluation becomes more standardised, transparent, and reproducible, which increases trust in the classifications and facilitates their sharing and future reuse. A main goal has also been to simplify the decision making process for the user, and *ella* has been carefully designed in close collaboration with lab engineers at our department. The current version of the tool is particularly useful for large-volume sample analyses with re-occurring variants and defined phenotypes (e.g. genetic tests for inheritable cancer), but we are also working to include more open-ended setups for low-volume, less clearly defined phenotypes. *Ella* is currently being tested at our department and will be used in production later this year. This presentation will demonstrate the logic behind the software, with specific examples.

1834W

ExpressionLncr: A pipeline for leveraging latent gene expression data in lncRNA studies. *G. Ellis[1], L. Akhabir[1], C. Brown[2], D. Daley[1].* 1) Centre for Heart Lung Innovation, University of British Columbia, Vancouver, British Columbia, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

Background: Investigating the functionality of lncRNA and other RNA in regulating the genome is an active and growing area of research. Databases of non-coding RNAs such as NONCODE and LNCipedia contain 141k and 119k human annotated lncRNAs to date, respectively—necessitating informatics tools to sift through and investigate functionality en masse. A wealth of functional genomics data is already deposited in NCBI GEO but, likewise, harnessing these 70 thousand experiments on 1.8 million samples in an effective manner can be challenging for those without informatics resources. **Hypothesis:** Existing functional genomics data can be leveraged to predict functionality of lncRNAs. **Methods:** We have created a program called ExpressionLncr(na) to harness this latent information. It is a pipeline to investigate the potential expression of lncRNAs by leveraging existing NCBI GEO gene expression information. The software sources lncRNAs from lncRNA databases such as NONCODE or LNCipedia and is not restricted to lncRNAs, allowing user-specified chromosome features. Features are restricted to reference organisms with expression probe array annotation information in Ensembl. The tool computes matches for positional overlap between Ensembl expression probes and lncRNAs. Summary results from GEO DataSets relevant to these overlapping features are used to calculate presence or absence of expression at each lncRNA. Positive links between probe expression data and lncRNA position may suggest possible functionality of the lncRNA worth further investigation. ExpressionLncr is written as a collection of command-line scripts with an optional graphical interface. The software will be available in two formats, both freely available: as a set of tools for the popular Galaxy bioinformatics web platform, and as an operating system independent desktop application. Future work is planned to extend the pipeline to RNA-seq information. **Summary:** ExpressionLncr is a bioinformatics pipeline to investigate the functionality of lncRNAs and other chromosomal features by computing positional overlap between lncRNA databases and existing gene expression probe information in NCBI GEO. Exploiting this latent information should help investigators interested in non-coding RNAs in planning new studies as well as prioritising candidate non-coding RNAs for molecular biology experiments.

1835T

A framework for benchmarking aligners and variant callers. *K. Ghose[1], G. Rakocevic[3], J. Spencer[2], W. Lee[1], D. Kural[1].* 1) Seven Bridges Genomics, Cambridge, MA; 2) Seven Bridges Genomics, London, UK; 3) Seven Bridges Genomics, Belgrade, Serbia.

There is a huge abundance of tools available for almost every step in any given genomics data processing pipeline. While this proliferation of tools can be beneficial, it also poses problems for researchers trying to construct a pipeline for their data analysis. Researchers are faced with the daunting task of determining which tools are trustworthy, give good results for their domain of interest and fit their accuracy, cost and time trade offs. While developing a new breed of genomic processing pipeline (a graph reference based aligner and variant caller) Seven bridges engineers had to also develop a framework for efficiently testing and benchmarking the performance of the new pipeline, comparing performance between different pipeline versions as well as against existing genomic pipelines. The variant calling benchmarking component is compliant with the GA4GH specifications for benchmarking variant callers.The benchmarking framework enables tool developers and users to run multiple pipelines against the same simulated or standard truth data sets with various parameter combinations and receive reports at multiple levels of detail. The reports start with leaderboard summaries, that give a high level overview of the field of tools being compared, which can then be inspected at increasing levels of detail, sufficient to debug tool performance at the level of individual variant calls and read alignment.While components of the framework can be run manually and locally, it is much more convenient to use the automation component to run and manage the huge volume of benchmark reports and raw data on the Seven Bridges Cloud Computing Platform. This has the added benefit of yielding accurate estimates of pipeline run time and cost when executed on the same platform.Though the framework has been developed to benchmark alignment and variant calling pipelines, the architecture has been designed to be flexible to be able to accommodate different kinds of genomics pipelines, with appropriate plugins for data analysis and display tailored to those classes of pipelines.

1836F

Towards the internet of DNA: Large-scale multi-institutional case-solving improves the power of and access to sequence-based rare disease diagnostics. *J. Gulcher[1], Y. Meng[1], K. Meirelles[1], H. Dai[1], Q. Zhao[1], T. Yu[2], S. Ennis[3], H. Wang[4], W. Zhou[4], P. Fang[1].* 1) WuXi NextCODE, Cambridge, MA; 2) Boston Children's Hospital, Boston, MA; 3) University College Dublin, Dublin, Ireland; 4) Children's Hospital of Fudan University, Shanghai, China.

Economical, high quality exome and whole-genome sequencing provides the raw data with which to diagnose and better treat an increasing proportion of rare disease cases. The challenge is interpreting that data efficiently and effectively, leveraging the wealth of large sequence databases, reference sets, and clinical genetics expertise worldwide. Doing so will benefit patients, who are, by definition, geographically scattered from others with similar rare disorders and who have variable levels of access to clinical genomics expertise. We will present our approach to meeting the challenge of data access and expert analysis through multi-institutional case-solving. We use a computationally efficient standard data architecture (a Genomically Ordered Relational database or GORdb) to store whole-exome or whole-genome sequence data. This includes aligned reads, variant data, and phenotype data, all identity-encrypted, at two sites: on local servers in China and on a HIPAA-compliant secure cloud in the United States. These data are made accessible to multiple clinical institutions in the US (WuXi NextCODE in Cambridge and Boston Children's Hospital), Ireland (UCD), and China (Fudan Children's Hospital) using a standard, validated decision-support user interface (Clinical Sequence Analyzer, or CSA).   With the ability to query and visualize raw sequence data at a distance, clinicians, clinical geneticists and bioinformaticians in Shanghai, Dublin, and Boston can collaborate on dozens of cases per week. All can work simultaneously from the same data, harmonized with the same set of public reference data, to identify known, novel and de novo pathogenic and likely pathogenic variants; instantly visualize the aligned reads for confirmation; search sequence databases at other institutions for the same variants or other variants in the same genes; and create diagnostic reports from their consensus findings. Final reports are signed off by certified medical professionals at the patients' institutions.   All of this is done without moving the raw data itself. This makes it much easier to share data and interpretation, and in a systematic and validated manner. For diagnostics this offers a scalable way to overcome major logistical and compliance barriers to using the full power of the data and the latest knowledge. It also makes possible the linking together of large collections of sequence data for discovery.

1837W

Development of a text mining algorithm to extract socioeconomic data from electronic health records for precision medicine research. *B. Hollister[1], E. Farber-Eger[1], N. Restrepo[2,3], D. Crawford[2,3], A. Non[4], M. Aldrich[5,6].* 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 4) Department of Anthropology, University of California, San Diego, La Jolla, CA; 5) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 6) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Socioeconomic status (SES) is a fundamental contributor to health, particularly when analyzing racial disparities in disease. SES data are rarely included in genetic studies and translational precision medicine research, due in part to the difficulty of collecting these data when studies were not originally designed for that purpose. The emergence of large clinic-based biobanks linked to electronic health records (EHRs) provides research access to large patient populations with longitudinal phenotypic data captured in structured fields as billing codes, procedure codes, and prescriptions. SES data however, are often not explicitly recorded in structured fields, but rather recorded in the free text of clinical notes and communications. The content and completeness of these data vary widely by practitioner. To enable gene-environment studies that consider SES as an exposure, we sought to extract SES variables from racial/ethnic minority patients (n=11,197) in BioVU, the Vanderbilt University Medical Center biorepository linked to de-identified EHRs. We developed several measures of SES using information available within the de-identified EHR, including broad categories of occupation, education, insurance status, and homelessness. 200 patients were randomly selected for manual review to develop an algorithm for extracting SES information from de-identified medical records. The algorithm consists of 15 categories of information, with 830 unique search terms. An additional 50 records were then randomly selected for algorithm evaluation. These 50 records were reviewed by two independent reviewers, with Kappa statistics ranging from 0.47 for occupation to 0.92 for retirement, and percent positive agreements ranging from 80% (Medicaid) to 93.3% (retirement). The manual review results from the 50 records were then compared to the results from the algorithm, resulting in positive predictive values of 35% (education), 63.6% (retirement), 87.5% (unemployment), 81.8% (Medicaid), and 33.3% (homelessness), suggesting some categories of SES data are easier to extract than others. The SES data extraction approach developed here will enable future EHR-based genetic studies to integrate SES information into statistical analyses. Ultimately, incorporation of measures of SES into genetic studies will help elucidate the impact of the social environment on disease risk and outcomes.

1838T

Accurately detecting genome-wide changes in gene expression or epigenetic modifications with a novel and efficient permutation-based method. *W. Huang, L. Li.* Biostatistics and Computational Biology Branch, NIEHS, NIH, Research Triangle Park, NC.

Cost-effective sequencing-based assays have become dominant methods for studying gene expression and epigenetic regulations, particularly for identifying genetic/epigenetic changes responsible for phenotypic differences. There are many exists methods for analyzing these sequencing data though only a few of methods widely used mainly due to their easy-to-use. Sequencing technologies and their applications are, however, still advancing quickly. Some existing tools are no longer sufficient for new and more advanced sequencing applications. For example, many early tools, designed for data with few replicates, cannot take full advantage of sequencing data with more replicates; and some of these tools rely on some unrealistic assumptions (e.g., Poisson model). Many out-of-date methods have either an inflated false positive rate or low statistical power to identify true difference. Taking advantage of increasing number of biological replicates from more recent sequencing data, and parallel computing power from high-performance computing (HPC) clusters, we introduce a novel balanced permutation method in parallel computing to identify differentially expression or epigenetic changes from sequencing data. Our semi-parametric method uses the balanced permutation to estimate distribution under true-null model, and then uses the estimated distribution to calculate the statistical significance of observed data. Our balanced permutation method can deal with cases that the numbers of replicates between two groups are different, and can automatically switch from systemic permutations to random permutation when possible permutations is larger than a defined threshold. We did the performance evaluation of our method with both simulated and real RNA-seq data. The results showed that our method, in comparison with traditional permutation methods, not only more powerful in detecting changes, but also more efficient in computation time. Our method is implemented in R, and will be freely available to the public once it is released.

1839F

Imputing a large multiethnic cohort of 200,000 individuals: The Department of Veterans Affairs Million Veterans Program (MVP). *J. Huang[1], S. Pyarajan[1,9], H. Hunter-Zinck[1], N. Sun[6,7], Y. Shi[1], M. Li[1,8], J. Gelernter[6,7], P. Loh[10], S. Das[11], C. Pan[2,3], E. Hauser[4,5], H. Zhao[6,7], T. Assimes[2,3], P. Tsao[2,3], J. Gaziano[1,9], C. O'Donnell[1,9].* 1) VA Boston Healthcare System, Boston, MA; 2) VA Palo Alto Health Care System, Palo Alto, CA; 3) Stanford University School of Medicine, Palo Alto, CA; 4) Durham VA Medical Center, Durham, NC; 5) Duke University, Durham, NC; 6) VA Connecticut Healthcare System, West Haven, CT; 7) Yale School of Public Health, New Haven, CT; 8) University of Utah, Salt Lake City, UT; 9) Harvard Medical School, Boston, MA; 10) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 11) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Introduction Statistical inference of missing genotypes (imputation) has been critically essential for the discovery of thousands of complex trait loci in meta-analyses of genome-wide association studies (GWAS). Although imputation on sample sizes exceeding 100,000 has been reported (UK Biobank), imputation in the Million Veteran Program (MVP) is unique in at least two aspects. First, the MVP Axiom array contains ~300,000 variants with MAF <1%, a much bigger proportion than that in UK Biobank. Second, MVP participants consist of diverse ethnic ancestries (~30% samples of non-European ancestry). **Methods** We used EAGLE2 and Minimac3, while UK Biobank used SHAPEIT3 and IMPUTE3. We run pre-phasing on full chromosomes and imputation in chunks (5Mb with 250kb flanking region). We first evaluated the applicability of commonly used genotype filtering criteria (i.e., MAF>1%, Missingness <5%; HWE P>1E-09) to the MVP genotype data and designed a preferred strategy on keeping/excluding samples and variants for imputation. We then used Pseudo-GWAS approaches to design a preferred pre-phasing strategy on a large dataset with a diverse ethnical background. We compared our MVP Axiom array data with a few other commonly used arrays including Axiom UK Biobank, Illumina Ominiexpress, and Illumina Omni5. Finally, we evaluated the computational resources needed for imputing a dataset with ~200,000 samples, within a local cluster that has stringent regulations on data security. **Results and Conclusions** Using EAGLE2 and Minimac 3, we were able to complete pre-phasing and imputation in ~14 days. The imputed data is comprised of 540 chunks of 5Mb across 22 autosomes, including 49,859,576 variants for 192,538 Samples. Overall, 32,204,361 and 12,802,159 have imputation $r2 > 0.3$ and > 0.8 respectively. However, we found that the overall r2 does not correlate well with empirical r2, therefore, we recommend to calcualte r2 based on subset of samples with the same ethnical background. We found that our Axiom MVP panel has comparable performance to the Axiom UK Biobank array, measured by imputation accuracy and genomic coverage. Genotyping and imputatoin for the next ~200,000 samples are expected to complete soon, and we will present most up-to-date results.

1840W

Maximizing sensitivity using ensemble genotyping approaches for whole-genome sequencing. *K. Hwang[1], I. Lee[2], X. Liu[3], H. Li[1], D. Won[1], S. Kong[3].* 1) School of Computer Science and Engineering, Soongsil University, Seoul, South Korea; 2) Samsung Genome Institute, Samsung Medical Center, Seoul, South Korea; 3) Computational Health Informatics Program, Boston Children's Hospital, Boston, MA.

Analytical validity and reproducibility between genome analysis pipelines remain questionable. There is an immediate need for a comprehensive evaluation of genome analysis pipelines. Here we present approaches to improve sensitivity of genotype calls based on a comparative analysis of 70 pipelines (7 aligners x 10 variant callers). Using two whole-genome sequences from a European (NA12878) and an African (NA19240) descents, we highlight the sources of discordant variant calls and compare the results with high-confidence variants from the Genome in a Bottle Consortium (GIAB) and the 1000 Genomes Project (1KGP) data. Surprisingly, rare variants with potential high impacts – i.e., interesting variants potentially associated with phenotypes – were enriched with discordant results, which could result in false positives and false negatives. For NA12878, 92% of common SNVs (MAF ≥ 5%) were identified by more than 56 pipelines (80% of the 70) while only 23% of rare SNVs (MAF < 0.5%) were concordantly called. Similar tendencies were observed for NA19240 and for INDELs from both individuals. Sequence context such as repetitive sequences and frequency of INDELs were associated with discordant calls between pipelines. Interestingly, high-confidence variants were concordantly called by the 70 pipelines when focusing on disease genes with high fixation indices between EUR and AFR (96% of 344,791 NA12878 GIAB SNVs and 92% of 545,698 NA19240 1KGP SNVs). However, significant proportions of high-confidence SNVs in such disease genes (33 for NA12878 and 2,658 for NA19240) were detected by less than 20% of the pipelines, suggesting that the use of one pipeline could be suboptimal for identifying possible phenotype-associated variants for further evaluation and validation. For these SNVs, the highest sensitivity was achieved by applying eight pipelines for NA12878 and 47 pipelines for NA19240, when the order of applying pipelines was optimized by a greedy algorithm. We demonstrate that discordance, including a small set of high-confidence variants with potential disease-association, between whole-genome genotyping pipelines is affected by multiple factors such as allele frequency, repetitive sequences, and the frequency of INDELs. Moreover, the minimum number of pipelines needed for maximizing sensitivity varies by sequencing data, suggesting that caution should be exercised in the identification of variants with clinical implications in genomic sequences from minorities.

1841T

Hyperlink Management System for automated integration of biological databases by use of data IDs. *T. Imanishi.* Dept. Molecular Life Sciences, Tokai University School of Medicine, Isehara, Kanagawa, Japan.

Hyperlink Management System (**HMS**; http://biodb.jp/) is a system for automatically updating and maintaining hyperlinks among major public databases in the field of life science. Everyday, we create corresponding tables of data IDs that are used in major databases for human genes and proteins, and provide a CGI-program that returns correct and up-to-date URLs of corresponding data in many databases. By using the CGI program on this server, users can easily create hyperlinks to major databases in their own databases and homepages. Furthermore, on this web site, we provide useful tools such as "All Database Search", "ID converter system", "Downloading ID tables", and "ID Resolver". Based on HMS, we also developed and released BioDBScan, a free web-based application that alerts you when any data about genes, proteins, chemicals or drugs of your interest is newly released in world-renowned databases. As a recent advancement of HMS, we newly added several databases in the field of human proteome. We also added several databases concerning human genomic polymorphisms into HMS, and realized a comprehensive, recursive data search through a chain of hyperlinks, which can promote discoveries in the field of genome medicine. As a result, total number of databases integrated in HMS increased to 42 for human molecules, 19 for mouse molecules, and 17 for drugs and chemicals, and total number of data IDs covered exceeds 250 millions. We will continue to integrate more databases in the field of life science that are distributed in the world, and in the near future we plan to install new functions such as enrichment analysis tools.

1842F

Assessing the impact of gene ontology changes over time with GOTrack.

M. Jacobson[1,2], A.E. Sedeño-Cortés[1], P. Pavlidis[1,2]. 1) Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada; 2) Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada.

   The Gene Ontology (GO) is widely used to annotate gene function, and is considered a gold standard for genomics data annotation and interpretation. It has become more popular to use "enrichment" analysis was a means for interpreting complex genetics findings. Given its prominence and value as a bioinformatics resource, it is important to understand how changes to GO and annotations over time impact users. While some studies of GO over time have been presented, routine assessment by end users remains challenging. Here we present GOTrack, a user-friendly web-based system for analyzing the historical stability of GO annotation data in multiple model organisms. The web site allows exploration of GO annotations at the gene and term level with multiple metrics and visualizations. Users can easily perform a "historical enrichment" analysis where over-representation analysis can be run on a gene hit list of interest using GO annotations from any era, with measurements of stability over time provided. Using the data and algorithms presented we investigate the impact of changes to GO on the stability and robustness of gene set enrichment results in a corpus of hit lists. GOTrack will help the research community assess the impact of changes to GO on their research and to focus on the most robust enrichment results.

1843W

Better interpretable models for proteomics data analysis using rule based mining.

F. Jayrannejad, T. O.F. Conrad. Department of Mathematics and Computer Science, Free University Berlin, Berlin, Berlin, Germany.

   Recent advances in -omics technology has yield in large data e.g. in the area of ms based proteomics. Due to noisy measurement and large size of this data, analyzing it has become a challenging issue. One of the main objectives in the analysis is identification of relevant patterns which can be used for classification. So, a method is required to find easily interpretable models from this data.An association rule (AR) mining algorithm mines patterns present in the data set frequently. The proposed method is to use AR mining algorithms to find appropriate classifying features. Many other techniques, e.g. machine learning methods, have been proposed to solve this problem. As mentioned before, one of the purposes of proteomics studies is to find features which can be used in diagnosis of diseases and provide an easily interpretable model. Many of the machine learning approaches e.g. neural networks do not provide easily interpretable models for the end users. On the other hand, most diseases are connected to a group of features (presence of a peak at a certain m/z value). ARs show this inter-dependency in an easily interpretable way. In addition, there are some groups of features which presence of only one of them might cause the occurrence of the phenotype. Therefore, a disjunctive AR mining algorithm, TitanicOR,is used to find this features in each class. In some cases, e.g. classifying two different types of blood cancer, there might be some features connected to a phenotype shared between two classes, here blood cancer. These features are not discriminating. In order to eliminate them, the idea of jumping emerging patterns is applied, to select the features that have a high frequency in one class and are not present in the other class at all. So, the method suggested in the current study starts with the peaks-lists from the spectras. Afterwards, the disjunctive emerging patterns will be mined using an adaptation of TitanicOR. Later, a collective likelihood based classification algorithm- PCL- will be used as classifier.Experiments of the proposed method on the real data available publicly and artificial data sets shows this method outperforms in terms of number of the features reported, reporting the inter-dependency between the features and diagnostic power of every selected feature individually, in comparison with state-of-the-arts method, SVM(liblinear) which has outperformed other classifies (Liu et. All, 2008).

1844T

Visualizing correlated causal variants. *R.D. Johnson[1], G. Kichaev[2], B. Pasaniuc[2,3,4].* 1) Dept of Mathematics, University of California, Los Angeles, Los Angeles, CA, USA; 2) Bioinformatics IDP, University of California Los Angeles, Los Angeles, CA, USA; 3) Dept of Human Genetics, Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 4) Dept of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA.

   Integrative methods that leverage disparate data sources have become important strategies in genetics research. For example, large-scale GWAS and functional annotation data derived from the ENCODE/Roadmap project have been successfully assembled to improve power in GWAS, fine-mapping, and eQTL studies. However, while the statistical engines that drive these methods have been well-established, principled ways of visualizing the resulting outputs is currently lacking. Here, we describe PAINTOR CAN-viz, a fine-mapping tool that visually summarizes an integrative fine-mapping experiment. The tool provides visual representation of the local correlation structure (LD), the functional annotations used, as well as association statistics and posterior probabilities for each SNP. Through PAINTER CAN-viz, we can visualize which SNPs form the credible set, which helps to identify the more significant SNPs. Additionally, the tool allows users specify smaller intervals of the genome to further examine areas of interest. Examining this output enables rapid identification of variants of interest for follow-up functional studies, providing insight into the driving mechanisms underlying a GWAS association. This command line tool is implemented in Python and we make freely available to the research community [https://github.com/ruthjohnson95/PAINTOR_CAN-vis].

1845F

PEMapper and PECaller – A more efficient whole genome sequence analysis pipeline. *H.R. Johnston, V. Patel, P. Chopra, M. Zwick, D.J. Cutler.* Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

   PEMapper and PECaller represent a marked improvement in efficiency and performance of whole genome mapping and genotype calling. PEMapper, while similar to BWA and Bowtie in approach, leverages the high-memory environment of modern computing clusters to allow the mapping processes to happen in parallel by significantly decreasing the number of read/write events to storage drives. This significantly increases the throughput of whole genome mapping by, for the first time, allowing a parallelized cluster to map multiple genomes without creating massive read/write bottlenecks in the process. PECaller represents an innovative approach to variant calling, in that it leverages both SNV and indel information simultaneously. This, paradoxically, allows PECaller to better distinguish between a true SNV and a true indel. Additionally, variant calling is done simultaneously in all samples of a given experiment. This allows PECaller to better differentiate between loci that are homozygous for the reference allele versus those that are 'missing data'. It also permits the imposition of a population genetics prior on the data as well as sophisticated models of read error. These combine to allow PECaller to better eliminate false-positives while still calling true heterozygotes effectively. This pipeline is dramatically more efficient than GATK in terms of resources used (storage space, computational time, network utilization, and efficiency in a computer clustered environment). Additionally, the data generated by PEMapper/PECaller is of the highest quality. We find that the number of false-positive variants called is likely to be less 0.1% (1 in 1000) per sample; the number of false-negative variants missed (true variants not called) is likely less than 3%, and the overall genotyping accuracy at known SNPs is somewhere between 99.7% and 99.9%, depending on the underlying accuracy of Illumina 2.5M arrays used as a comparison standard. When run on a novel experimental data set, PEMapper and PECaller perform exceptionally well. Over 400 samples were run through the pipeline, and the data quality is very high. Samples have theta between .0008 and .0009 across the board, as expected and the Ts/Tv ratio is greater than 2. Theta in the exome region is ~.00045, and the exome Ts/Tv ratio is greater than 3. Coupled with the accuracy information attained in the comparison to Illumina 2.5 array data, it is clear that this pipeline is both efficient and accurate.

1846W

High accuracy HLA alleles imputation system for Affymetrix Axiom Japonica Array™ and its application on Steven-Johnson Syndrome. *S.S. Khor[1], X.W. Zheng[2], H. Sawai[1], N. Nishida[1,3], Y. Hitomi[1], M. Ueta[4], R. Shingaki[5], S. Iwata[5], S. Kinoshita[4], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine,, University of Tokyo, Tokyo, Tokyo, Japan; 2) Department of Biostatistics, University of Washington, Seattle, WA, United States; 3) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 4) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 5) Toshiba Corporation, Life Science Div, Life Science Business Dept, Tokyo, Japan.

Human leukocyte antigen (HLA) or the Major histocompatibility complex (MHC) has been reported to be associated with various diseases and traits. Due to the complexity of the linkage disequilibrium (LD) structure of the MHC region, many genome-wide association studies (GWAS) failed to identify the primary causal variants or associated HLA alleles due to high cost of HLA typing. Statistically imputation of HLA alleles has becoming a routine in fine-mapping of GWAS studies. However, HLA imputations are highly dependent on the matching of HLA imputation reference and target population (eg: European reference performed poorly on Japanese GWAS). We have previously reported the reference panel for Japanese using Affymetrix 6.0, Affymetrix Axiom ASI and Ilumina HumanOmni2.5 reaching HLA imputation accuracies of 95.2-99.5% across the six HLA genes (-A, -B, -C, -DRB1, -DQB1, -DPB1). Recently, ToMMo (Tohoku Medical Megabank Organization), Toshiba healthcare together with Affymetrix Axiom® have developed Japonica Array™ specially customized for Japanese population. Japonica Array™ is shown to provide better SNP imputation performance in comparison to other commercially available genotyping platform. Based on our HLA imputation system (HIBAG), Japonica Array™ performed consistently well across the six HLA genes and performed exceptionally well in HLA-B compared with the other 3 genotyping platforms. Moreover, the Japonica Array™ could impute the risk *HLA* alleles (*HLA-A*02:06* and *HLA-B*44:03*) for cold medicine related Japanese Steven-Johnson Syndrome (SJS) with 100% accuracy in comparison with Affymetrix Axiom ASI with an accuracy of 59/59 and 28/29 in 117 patients who carry the risk *HLA* alleles respectively. Therefore, the new Japonica HLA imputation could provide a nearly 100% accuracy imputation for common alleles in comparison with other commercial genotyping platforms.

1847T

JGA: Japanese Genotype–phenotype Archive. *Y. Kodama[1], J. Mashima[1], M.T. Minowa[2], M. Kawashima[2], K. Miyazaki[2], N. Mitsuhashi[2], E. Kaminuma[1], O. Ogasawara[1], K. Okubo[1], Y. Nakamura[1], T. Takagi[1].* 1) DDBJ Center, National Institute of Genetics, Mishima, Shizuoka, Japan; 2) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan.

This poster focuses on the Japanese Genotype-phenotype Archive (JGA, http://trace.ddbj.nig.ac.jp/jga) in collaboration with our partner institute, the National Bioscience Database Center (NBDC, http://humandbs.biosciencedbc.jp) of the Japan Science and Technology Agency. JGA is the Japanese counterpart of the dbGaP (database of genotypes and phenotypes) and EGA (the European Genome-phenome Archive) at NCBI and EBI, respectively, serving as one of centralized repositories providing authorized access to individual-level phenotype and genotype data. To exploit personal genomic data while respecting the privacy and informed consents of study participants, it is essential to establish a centralized repository for data management and a policy for data usage. In the US, the National Institutes of Health (NIH) have established guidelines for sharing human subject data. Likewise, NBDC in Japan established guidelines and policies for sharing human-derived data. The Data Access Committee (DAC) at NBDC reviews and makes decisions about data submission to JGA and usage requests from researchers. JGA accepts data that are de-identified by submitters. Acceptable data types include raw data from array-based and next-generation sequencing platforms, clinical images, and phenotype data associated with data samples. Processed and analyzed data such as alignments, assemblies and variations are also acceptable. Upon submission, the JGA team will archive the original data files in encrypted form in the database. Information in JGA is organized in a hierarchical JGA data model based on that of EGA. JGA assigns stable, unique identifiers prefixed by 'JGA' to studies and subsets of information from those studies. Once access has been granted by DAC, datasets with access permission can be downloaded with secure downloading software. Information about available studies (summary level data) can be accessed freely on the JGA (https://ddbj.nig.ac.jp/jga/viewer/view/studies) and NBDC websites. As of 27 May 2016, 28 studies are available at JGA. In the poster, we present contents, system and recent developments of JGA.

1848F

Reducing genotoxicity risk in gene therapy - Identification of tumor insertion sites using next gen sequencing. *H. Kuck[1], P. Westmoreland[2], V. Nadella[1], H. Zhong[1], D. McCarty[2], P. White[1].* 1) The Institute for Genomic Medicine at Nationwide Children's Hospital, Columbus, OH; 2) The Center for Gene Therapy, The Research Institute at Nationwide Children's Hospital, Columbus, OH.

Gene therapy utilizing recombinant adeno-associated virus (rAAV) vectors has been a safe and effective treatment in mouse models of human disease and in clinical applications. Although the risk is much lower for integration into chromosomal DNA, potentially promoting tumor growth, than with retroviral vectors, there have been reports of genotoxicity limited to specific vectors associated with excess hepatocellular carcinomas (HCC) in mice. In order to understand the potential mechanisms of genotoxicity and identify insertion patterns of integration events that could promote tumor formation, we identified insertion sites from tumors in tumor prone C3H/HeJ mice treated with self-complementary AAV (scAAV) vectors, either conventionally expressing a gene, or designed to promote insertional activation.   To identify insertion sites we followed Agilent's SureSelect QXT NGS target enrichment workflow to enable us to enrich genomic DNA fragments containing scAAV. Genomic DNA extracted from the tumors was fragmented, adaptor-ligated and PCR amplified. Custom SureSelect Probes of 120bp designed using Agilent's SureDesign service to target the scAAV vectors were used for library capture. The target enriched DNA library was further PCR amplified using dual indexing primers and sequenced on Hiseq 4000 using Paired end 2X 150bp chemistry. For bioinoformatic analysis of the resulting sequence data we developed a custom application to identify insertions sites. Read pairs were aligned to the mouse and vector using BWA-MEM and SAMBLASTER was used to find split and discordant reads. Potential insertion sites are identified and quantified by analyzing split and discordant read pairs that align to both mouse and vector. These sites are ranked by the strength of supporting evidence: insertion sites where multiple reads have junctions at the same location with high quality alignment to both sequences most likely identify a true insertion site. Insertion sites will be validated by PCR amplification of the identified vector-chromosome junctions and Sanger sequencing.   Our results indicate repeated patterns of vector insertions associated with specific proto-oncogenes depending on which vector was administered. This suggests new possibilities in vector design to minimize the risk of the most common proto-oncogene interactions.

1849W

Interpretable, integrative deep learning models of *in-vivo* **transcription factor binding reveal the context-specific heterogeneity of regulatory DNA.** *A. Kundaje[1,2], N. Boley[1], J. Israeli[3], A. Shrikumar[2], P. Greenside[4].* 1) Dept. of Genetics, Stanford University, Stanford, CA 94305; 2) Dept. of Computer Science, Stanford University, Stanford, CA 94305; 3) Biophysics Program, Stanford University, Stanford, CA, 94305; 4) Biomedical Informatics Program, Stanford University, Stanford, CA 94305.

We develop the first multi-task, multi-modal convolutional neural networks with biophysically motivated deep architectures that integrate both haploid or diploid raw DNA sequence and chromatin accessibility profiles (DNase-seq or ATAC-seq) to model and predict in-vivo binding sites of 100s of TFs (using ChIP-seq data from the ENCODE project) across diverse cell types with high accuracy. Our integrative models provide very significant improvements over other state-of-the-art methods including recently published deep learning TF binding models especially when evaluated in terms of precision and recall of predictions across the whole genome in cell types not used in training. Using calibrated models, we provide the largest resource of high-resolution, high-confidence predicted in-vivo TF binding profiles of 100s of TFs in 250 cell types with DNase-seq data but no TF ChIP-seq data. Next, we develop DeepLIFT (Deep Learning Importance Feature Tracker), a novel interpretation engine for extracting predictive and biological meaningful patterns from deep neural networks (DNNs) for diverse genomic data types. Since DNNs learn inherently distributed representations, we find that multiple convolutional filters often cooperatively represent distinct regulatory features such as TF binding preferences and hence caution against overinterpreting individual filters. Filter nullification and 'in-silico mutagenesis' are commonly used methods to score the relative importance of specific inputs such as individual bases in input DNA sequences. These approaches are not only computationally expensive but can often provide misleading results when the inputs contains redundant signals that potentially buffer each other. DeepLIFT is the first method that can integrate the combined effects of multiple cooperating filters and compute importance scores accounting for redundant patterns. We apply DeepLIFT on our TF binding models to obtain novel representations of TF sequence affinity models that highlight the importance of flanking nucleotide sequences; infer high resolution point binding events; discover novel regulatory sequence grammars involving homodimer and heterodimeric co-factor binding with spatial constraints; learn de-novo TF DNase footprints and unravel the heterogeneity of TF binding sites.

1850T

Giggle: A scalable and fast search engine for large-scale multi-omics data integration. *R.M. Layer, B.S. Pedersen, T.L. Di Sera, A.R. Quinlan.* University of Utah, Salt Lake City, UT.

While the analytic benefit of integrating genetic variation with functional genomic annotations is clear (see Pickrell, Am. J. Hum. Genet, 2014), our current capacity to consider these data is far outstripped by the pace at which annotations are accumulating. There are thousands of annotations for the human genome, ranging from gene annotations to deep catalogs of genetic variation. Large-scale functional studies have created an even richer resource: 1,600 annotation tracks have been produced by ENCODE, 2,800 by Roadmap Epigenomics, and 8,500 by GTEx. Leveraging this wealth of information requires a scalable mechanism to quickly identify and quantify overlaps across large and diverse sets of annotations. Existing tools such as TABIX and BEDTOOLS were designed to investigate a single annotation set, not thousands. We developed GIGGLE, a novel and highly scalable genomic interval searching strategy that indexes the genome features from many (thousands or more) different annotation files and formats. GIGGLE is based on a B+tree, where the bounds of an interval serve as keys to the path and file offset of the original record. Single-interval GIGGLE searches return a list of overlapping genomic intervals per annotation set, which on their own can be used for rapid summary and prioritization statistics. Multi-interval GIGGLE searches can also give a list of tracks ordered by an estimate of significance based on either intersection or proximity. This estimate uses a Fisher's exact test; we demonstrate that this metric is well correlated with permutation tests, yet is substantially faster. These summaries are returned nearly instantly, thereby enabling extremely fast (seconds not minutes) data exploration. A single-interval search across GTEx (31 files, 1.7M intervals) requires 0.076s (12.9X faster than TABIX, 540.8X faster than BEDTOOLS), and a search of 1597 intervals across Roadmap Epigenomics (1905 files, 55M intervals) takes a mere 0.92s (66.8X faster than TABIX, 94.3X faster than BEDTOOLS).GIGGLE (https://github.com/ryanlayer/giggle) can be used as a C API, via Python bindings, and as both a command line and web-based tool to search a curated set of published and experimental data. We present the use of GIGGLE as a large-scale public search engine with an index across major projects like ENCODE, Roadmap Epigenomics, and GTEx.

1851F

Integrative post-GWAS analysis unifies cell type-specific and phenotype-centric functional annotations and provides insights into the genetic basis of complex traits. *Q. Lu[1], R. Powles[2], U. Shaham[3], Q. Wang[2], Y. Hu[1], B. Li[1], S. Abdallah[4], D. Ou[4], C. Jin[5], Y. Lu[6], H. Zhao[1,2,7].* 1) Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA; 2) Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; 3) Department of Statistics, Yale University, New Haven, CT, USA; 4) Yale University School of Medicine, New Haven, CT, USA; 5) Yale College, New Haven, CT, USA; 6) Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA; 7) VA Cooperative Studies Program Coordinating Center, West Haven, CT, USA.

Many genetic variants influence complex traits through perturbing regulatory elements that are specific to disease-relevant cell types. Accurate prediction of cell type-specific functionality in the human genome is a critical step towards understanding the genetic basis of complex traits. In this project, we first predicted functional DNA elements for 127 distinct tissue and cell types through integrative analysis of various histone marks, DHSs, DNA methylation, and RNA-seq data. These integrative annotations, coupled with GWAS summary data, enabled us to stratify trait heritability by cell type. In total, we identified significantly enriched cell types for more than 40 complex diseases and traits. Compared with any individual annotation, our integrative approach substantially increased power to identify disease-relevant cell types. Our analysis highlights strong enrichment in immune cells for neurological disorders. Primary monocytes were significantly enriched for both Alzheimer's disease (AD) and Parkinson's disease (PD) signals ($P=2.3\times10^{-5}$ and 2.3×10^{-6}). Brain anterior caudate was highly significant for schizophrenia ($P=1.7\times10^{-11}$), bipolar disorder ($P=1.6\times10^{-4}$), cognitive performance ($P=8.7\times10^{-6}$), education years ($P=8.7\times10^{-10}$), and body mass index ($P=8.1\times10^{-8}$). Liver was consistently enriched for multiple lipid traits, and surprisingly, for AD ($P=4.4\times10^{-4}$) as well. Signal enrichment in liver remained significant after removal of *APOE* region, suggesting a potential role of the metabolic system in AD etiology. Finally, integrative annotations that accurately identify cell type-specific functionality, in conjunction with enrichment analysis well supported by the literature, open the door to phenotype-centric pathogenicity prediction for non-coding DNA variants. Functional variants in trait-relevant cell types should be given higher priority in their potential to influence the trait. Based on this idea, we have developed a deep learning approach to predicting trait-specific pathogenicity using cell type enrichment results. Among 7 recently validated regulatory variants for coronary artery disease, 4 ranked in the top 5% of all variants based on our integrative phenotype-centric pathogenicity prediction ($P=1.9\times10^{-4}$), while CADD identified none. Taken together, our approach unifies cell type-specific genome annotation and phenotype-centric pathogenicity prediction, and provides insights into the genetic basis of many complex diseases and traits.

1852W

The SSV evaluation system: A tool to prioritize short structural variants for studies of possible regulatory and causal variants. *M.W. Lutz[1], R. Saul[2], D.K. Burns[3], A.D. Roses[1,3], O. Chiba-Falek[1,4].* 1) Department of Neurology, Duke University School of Medicine, Durham, NC 27710; 2) Polymorphic DNA Technologies, Alameda, CA 94501; 3) Zinfandel Pharmaceuticals, Chapel Hill, NC 27517; 4) Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27710.

Short Structural Variants (SSVs) are short genomic variants (<50 bp) other than SNPs, and include short deletions, short insertions, insertion/deletions (In-del), mixed (cluster that contains multiple classes), multiple nucleotide polymorphism (MNP), and microsatellites or simple sequence repeats/short tandem repeats (SSRs/STRs). Recently it has been suggested that SSVs contribute to many human complex traits. However, high-throughput analysis of this class of variants presents numerous technical challenges. In order to facilitate the discovery and assessment of SSVs, we have developed a prototype bioinformatics tool, the "SSV evaluation system", which is a searchable, annotated database of SSVs in the human genome, with associated customizable scoring software that is designed to evaluate and prioritize SSVs that are most likely to have significant biological effects and impact on disease risk. The SSVs database currently contains annotations for over 7.5 million (M) SSVs including, 1.3M Simple Sequence Repeats, 2.7M insertions and 3.2M deletions. Annotation includes data describing the genomic location and type of the SSV, variability indicators, clustering index, gene context, tissue specific regulatory tracks and epigenomic signals, conservation, position within a linkage disequilibrium (LD) block, and genome-wide association studies (GWAS) signals. This new bioinformatics tool is a component in a broad strategy that we have been using to discover potentially important SSVs within candidate genomic regions that have been reported in the literature or identified in GWAS, with the goal to prioritize candidate functional/causal SSVs for in-depth analyses. The SSV evaluation system allows the investigators to focus the follow-up genotyping and sequencing efforts as well as validation and functional experiments on a relatively small list of SSVs that are more likely to be functional and causal. We have been using the SSV evaluation system to discover new candidate causal variants related to complex neurological diseases of the aging brain such as, Synucleinopathies, late-onset Alzheimer Disease (LOAD) and Amyotrophic Lateral Sclerosis (ALS). We have found the SSV evaluation system to be a powerful tool to guide genetic investigations aiming to uncover SSVs that underlie human complex diseases including idiopathic neurodegenerative disease in aging.

1853T

Optimizing pediatric health care outcomes through genomics studies and innovative infrastructure involving effective and efficient research organization/healthcare provider integration. *F. Mentch[1], H. Qiu[1], J. Connolly[1], C. Kim[1], R. Chiavacci[1], R. Pellegrino[1], L. Hermannsson[1], P.M.A. Sleiman[1,2], H. Hakonarson[1,2].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

The Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia (CHOP), founded in 2006, is one of the largest genetics/genomics research programs and has the world's largest collection of pediatric samples linked to extensive phenotype and genotype information. Being embedded in CHOP and as a member of the University of Pennsylvania research network, the CAG has access to world-class clinical and scientific expertise in a multitude of disease and research areas, recruits participants from a large and disease/condition-enriched patient population, and works with motivated families that routinely volunteer parental samples and grant recontact opportunity for deep phenotyping and access to extensive pedigree information. CAG has contributed multiple bioinformatics tools to the community, has established the Philadelphia Neurodevelopmental Cohort with a recall of nearly 10,000 previously recruited patients for extensive phenotyping in a joint CHOP/Penn effort, has contributed to enhanced sample handling technology in collaboration with industry leaders in robotics hardware, and is working with another innovative company on a methodology to overcome limitations imposed by short read DNA technology. The CAG has numerous on-going collaborations with CHOP/UPENN researchers and other national and international researchers and is actively engaged in translating our genetic discoveries to therapeutics in collaboration with four drug discovery companies. CAG is an active participant in the NHGRI sponsored Electronic Medical Records and Genomics (eMERGE) Network. We report the unique organization and infrastructure that enabled rapid investigation and novel discoveries in over 60 disease phenotypes involving more than 100,000 children yielding over 500 publications in both complex and Mendelian disease areas, and resulting in on-going clinical trials in pediatric oncology and ADHD and planned clinical trials in autism, schizophrenia and several rare disease disorders. We describe the recruiting strategies, sample preparation, tracking and management methodologies and database organization and access methods that in combination have accelerated CAG's genomics discovery activities and the translational work for new and improved diagnostics and therapies for others to follow.

1854F

JustOrthologs: A fast, accurate, and user-friendly ortholog-identification algorithm. *J.B. Miller, B.D. Pickett, P.G. Ridge.* Biology, Brigham Young University, Provo, UT.

Orthologous gene identification is fundamental to all aspects of biology. Specifically, ortholog identification between humans and other organisms can provide functional insights for genes of unknown function. Currently, most ortholog identification algorithms require all-versus-all Basic Local Alignment Search Tool (BLAST) comparisons, which are time consuming and memory intensive. We developed a novel algorithm, JustOrthologs, which exploits a previously unidentified bias in coding sequence (CDS) length. By requiring the same CDS lengths between two genes, JustOrthologs decreases ortholog identification runtime by 90%, while obtaining comparable precision and recall scores to OrthoMCL, OMA, and OrthoFinder. Using this technique, we compared 308 complete eukaryotic genomes and confirmed gene annotations for 1,447,473 genes, clustered 370,606 genes in previously unreported groups, and identified 7,182 potentially mislabeled genes across 464,982 ortholog groups.

1855W

Phasing variants of duplicated genes using whole genome sequencing data. *T. Mimori[1], N. Nariai[2], K. Kojima[1], Y. Sato[1], Y. Kawai[1], Y. Yamaguchi-Kabata[1], M. Nagasaki[1].* 1) Tohoku Medical Megabank Organaization, Tohoku University, Sendai, Miyagi, Japan; 2) University of California, San Diego.

Genomic studies based on population-scale whole-genome sequencing data has revealed the diversity and fine-scale picture of human genomes. The variation of individual genomes has been described and cataloged in form of variant types, such as SNPs, indels, short tandem repeats, inversions and copy number variations (CNVs). The genotypes and phases of these variants have been analysed for building reference panel, which has been a foundation of genetic association studies via genotype imputation. Although the diversity of duplicated genes is also an important source of phenotypic difference, variants and haplotype structures inside of duplicated genes are not fully investigated, which could be partly due to the difficulty and complexity of such analysis. We proposed a computational apporach CNValloc to reveal phase of variants inside of duplicated genes using high-throughput sequencing data of multiple individuals. In addition to performance evaluation via synthetic data, we present a utility of CNValloc for highly variated genes within and among populations such as AMY1 gene.

1856T

Sparse polygenic modeling approaches are highly effective for prioritizing causal eQTL variants. *N. Nariai[1], C. DeBoever[2], H. Li[1], K. Frazer[1,3].* 1) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 2) Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA; 3) Department of Pediatrics, University of California, San Diego, La Jolla, CA.

   Genotyping data combined with RNA-seq data of tissue samples from many individuals can identify expression quantitative trait loci (eQTL) variants. However, there are several drawbacks with the standard single-variant association analysis used for eQTL studies: 1) combinatorial effects from multiple causal eQTL variants cannot be estimated; 2) variants that show the strongest associations are not necessarily causal, but can be just in linkage disequilibrium (LD) with the causal ones; and 3) association signals are more significant for common variants than low-frequency variants, given the same effect size. To address these issues, we assess several sparse polygenic modeling approaches for prioritizing causal eQTL variants. We evaluate Lasso, Elastic Net, and Bayesian Sparse Linear Mixed Models (BSLMM) in identifying eQTL variants utilizing simulated gene expression data but real genotype data of 503 unrelated individuals to maintain LD structure among tested variants. We find that Elastic Net and BSLMM outperform the standard single-variant association analysis in identifying more causal eQTL variants at the same false discovery rate across a range of sample sizes (100 or 503), number of causal variants (1, 5, 10, or 100) and heritability (20% or 60%). We apply BSLMM to RNA-seq data and whole-genome sequencing data of 131 iPSC lines for eQTL variant prioritization. Among the top 10,000 ranked eQTL variants, multiple variants (ranging from 2 to 29) are jointly associated with expression levels for ~ 67% of genes with one or more eQTL variants, and 13.9% of the eQTL variants are low-frequency (MAF 1-5%); whereas only 7.6% of the top eQTL variants identified by single-variant association analysis are low-frequency. To determine if the prioritized eQTL variants are likely to be causal, we calculate their enrichment in stem cell regulatory regions. We observe a higher enrichment in iPSC DNase I hypersensitive sites for eQTL variants prioritized by BSLMM, compared with those identified by single-variant association analysis. Additionally, the top ranked eQTL variants are highly enriched in H1 hESC Nanog and Oct4 binding sites, which play important roles in reprogramming and pluripotency. We show that BSLMM is highly effective for prioritizing causal eQTL variants, with the ability to jointly estimate effect sizes of multiple variants and simultaneously take LD structure into account, regardless of minor allele frequencies, even from a limited number of samples.

1857F

Benchmarking of mutation calling methods with NGS data. *M. Nomura, Genome Evaluation Group in Center for iPS Cell Research and Application, Kyoto University.* Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan.

   With genomic sequences generated by next generation sequencers, we have checked iPS cells and differentiated cells for clinical use from various points of view in genome, epigenome and transcriptome. Especially, we have been focusing on detecting genomic mutations, i.e., SNVs, Indels, CNVs and SVs, in iPS cells manufactured in the iPS cell stock project. For our purpose, mutation calling algorithms must handle the following issue that stems from unique characters of iPSC clones: iPSC clone is composed of homogeneous cells but small portions of heterogeneous cells may exist. Therefore, we pay attention to the sensitivity of algorithms but the specificity is also important to reduce false positives for validating candidates of mutations by other methods. To achieve both high sensitivity and specificity, mutation calling methods that integrate existing mutation callers known as ensemble approach, e.g., SomaticSeq[1] and MetaSV[2], have been proposed recently. Because performance of calling methods can be altered depending on sequencing conditions, e.g., read length and read depth, and because they have been usually evaluated with standard genomes such as NA12878 and cancer genomes, we have to reevaluate their performance concerning nature of our data. We have started making synthetic data sets that reflect characteristics of our data and applying existing calling methods including MuTect2[3] to the synthetic data as well as synthetic data generated in ICGC-TCGA DREAM Genome Mutation Calling Challenge[4] to evaluate their performances. In this poster, we will present our very preliminary results.   References 1. An ensemble approach to accurately detect somatic mutations using SomaticSeq, Genome Biol., 16 : 197 (2015). 2. MetaSV: an accurate integrative structural-variant caller for next generation sequencing, Bioinformatics, 31 : 2741 (2015). 3. https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_cancer_m2_MuTect2.php 4. https://www.synapse.org/#!Synapse:syn312572.

1858W

Interactive analytics for very large scale genomic data. *C. Pan[1,2,3], N. Deflaux[4], G. McInnes[3], M. Snyder[2,3], J. Bingham[4], S. Datta[3], P. Tsao[1,5,6].* 1) Palo Alto Veterans Institute for Research, Palo Alto, CA; 2) Department of Genetics, Stanford University, CA; 3) Stanford Center for Genomics and Personalized Medicine, Stanford University, CA; 4) Google Inc, Mountain View, CA; 5) Division of Cardiovascular Medicine, Stanford University, CA; 6) VA Palo Alto Health Care System, CA.

Large-scale genomic sequencing is now widely used to decipher questions in diverse realms such as biological function, human diseases, evolution, ecosystems, and agriculture. With the quantity and diversity these data harbor, a robust and scalable data handling and analysis solution is desired. Here we present interactive analytics using public cloud infrastructure such as Google Cloud Platform and distributed computing database Dremel, as developed according to the standards of Global Alliance for Genomics and Health, to perform complex information representation, comprehensive quality controls, and biological information retrieval in large volumes of genomic data. We demonstrate in a dataset of 500 deeply sequenced (50x) human genomes that (1) the nested and repeated features of our database schema help to preserve rich information of alignment and variant calling while achieving a near lossless compression, (2) a comprehensive set of population genomics metrics can be implemented in Dremel using query languages for monitoring sample-level and variant-level data quality, and (3) the query solutions we developed can quickly help researchers navigate the statistics of genotypes and search for biological and medical information by referencing database records. Such computing paradigms of combining public cloud and distributed computing database can provide orders of magnitude faster turnaround for common analyses, transforming long-running batch jobs submitted via a Linux shell into questions that can be asked from a web browser in seconds.

1859T

Kmer-SSR: An efficient tool for perfect SSR detection in genetic sequences. *B. Pickett, J. Miller, P. Ridge.* Department of Biology, Brigham Young University, Provo, UT.

Simple Sequence Repeats (SSRs) exhibit great inter- and intra-specific variability in the number of repeats. This attribute makes them useful markers in forensics, population genetics, and other applications. We developed a new algorithm for SSR detection that is conceptually based in k-mer decomposition. Our algorithm could be implemented as part of any tool that uses k-mer decomposition; e.g., sequence alignment tools using De Bruijn graphs. Our stand-alone Python implementation is fast, 100% accurate and complete, and provides fine-tuned user control while maintaining simplicity for the average user. Conversely, existing softwares are plagued with one or more deficiencies in these areas.

1860F

Mutation visualization on 3D protein structures in the context of known pathogenic mutations and other data. *P.M. Ponomarenko.* Bioinformatics, ICG SB RAS, Novosibirsk, Russian Federation.

   Next Generation Sequencing translates genetic information into bits of data with rapid cost decline. Doctors have access to it in a form of tables with thousands and/or millions of lines. Using databases, filters and sometimes machine learning they look for mutations that may be pathogenic to help with diagnosis. At this stage data continue to be in a form of tables with numbers and words like BRCA1, NM_007294.3:c.107C>A p.Ser36Tyr, HET, DP=15, SIFT=0, PolyPhen=1, Exon 3 out of 22, etc. This information is very useful but hard to build your intuition around and make a treatment decision even after you studied many publications about this gene. Is this mutation in exon 3 more likely to be pathogenic than in exon 4? Is position 455 of the same importance for protein function as position 460, some changes in which are known to be pathogenic? Is this particular aminoacid change can be pathogenic? There are many questions doctors and/or lab directors need to know before assigning treatment. These questions can be answered to some extent by computational algorithms like SIFT and PolyPhen scores mentioned above, but most of these algorithms rely on conservation scores and were developed for a broad range of diseases and genes. On the other hand doctors and researchers are trained to notice minor visual changes and abnormalities in patients and data like chest X-rays, CAT scans, etc. For that reason I think that visual representation of mutation on protein 3d structure with relevant data about positions of known pathogenic mutations, drug binding domains and other information may give doctors a new picture that is easy to understand and remember. To show mutations in the context of 3d structures, I created a database of homology models using ICM-Pro by Molsoft with different levels of quality for human proteins found in NCBI RefSeq when a base structure can be found in RCSB protein databank and using Blast and UniProtKB. Users can find relevant structure and pinpoint mutations of interest with single search term in HGVSp notation. Once homology model is found user can select different features for display online as 3d slides using ActiveICM like PDB structures used to build homology model, known and likely pathogenic mutations from HGMD and ClinVar as well as predicted drug binding pockets, active sites and other relevant data. These 3d slides can be downloaded in a form of icb file to share with colleagues for further research or put inside PowerPoint presentation.

1861W

Deriving genetic ancestry using deep learning on electronic health records. *K.M. Quinnies, R.S. Vanguri, N.P. Tatonetti.* Columbia University Medical Center, New York, NY.

   Genetic ancestry is a significant factor in disease risk and an essential factor for studies aiming to discover variants with clinical significance. With existing genetic data, clustering by ancestry is relatively simple and often achieved with principal components analysis (PCA). PCA reduces dimensionality while maintaining variability, making shared ancestry in complex populations easily identifiable. However, genetic data are not always readily available in clinical and research settings. On the other hand, clinical data, like those stored in electronic health records (EHR) are vast, but can be difficult to understand and do not directly capture ancestry information. We present a novel strategy to estimate genetic ancestry using deep learning on the electronic health records of 5,000 patients. Once trained, our model can then be used to predict the genetic ancestry of any patient in the EHR that meets the minimum data requirements (nearly 2 million patients in our institutional EHR). We call this inferred genetic ancestry the "clinical ancestry score." To compute this score, we begin by deriving principal components using whole exome sequencing data from the 1000 genomes project. The 1000 genomes project contains known populations which are sufficiently diverse (incl. caucasians, african americans, and multi-ethnic groups). Whole exome sequencing data from 2,700 individuals and genotyping data from 2,300 individuals which have been cross referenced with Columbia University Medical Center's Clinical Data Warehouse (CDW) is then projected onto the derived eigenvectors. We then formed a dataset of clinical variables, focusing on the most commonly collected variables (i.e. disease diagnoses, medical history, drug orders, and demographic data) and used the clinical variables to predict the two principal components which have the largest possible variance. We achieved a classification area under the receiver operator characteristic curve (AUROC) score of 0.75 (p = 0.0003). In the future, we will have a model that can be applied to the majority of patients in our EHR data (since it has been trained on commonly collected clinical variables). This model will predict the ancestry of the patient with error that can be estimated. We will be able to use this clinical ancestry score as another genetic proxy variable to select subgroups of patients with shared genetic etiologies.

1862T

A suite of programs for pre- and post-Imputation data checking. *N.W. Rayner[1,2,3], N.R. Robertson[1,2], A. Mahajan[1], M.I. McCarthy[1,4].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 3) Sanger Institute, Hinxton, Cambridge, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Rigours data quality control prior to imputation is vital to ensure high quality output and although the availability of public imputation servers has greatly reduced the complexity associated with performing an imputation run, they do not remove this requirement to verify the data both prior to and post imputation. To simplify this checking we have developed two programs that check the data at both the pre- and post-imputation stages. Both are written in Perl with some additional steps in the post-imputation program in Java. At the pre-imputation stage the program compares a Plink format bim and frequency file to the selected reference panel, 1000G phase 3 and HRC r1 and r1.1 are supported. The program checks all variants in the bim file that match the reference for a wide variety of potential issues, such as differences in chromosome and position, alleles, allele frequency (AF), strand and reference allele. Summary statistics are written to a log file and for each potential issue a list of variants affected is produced. These lists are used in Plink to update issues and remove variants. This validation is quick to run (~5 minutes using HRC reference panel v1.1) and updating ensures the data submitted to the imputation servers will not be rejected and that the output is of high quality. Post-imputation, the program takes the most common output formats from the current suite of imputation programs (Minimac and IMPUTE) or servers (hosted by Oxford, University of Michigan, and Sanger) and produces a wide range of charts on information score, AF and position to visually assess the quality of the imputation run, flagging such issues as missing regions in the output, as evidenced by two or more regions of 1MB or greater without SNPs, or regions with low information score. For ease of visualisation these charts are collated into a single html file with embedded images, viewable in any current web browser. The program can also be run on multiple imputed data sets, comprising any supported imputation format, and produces an additional html summary across all. Both programs are freely available to download and have been used extensively in single cohort studies, as well as for large consortia such as DIAMANTE where >35 genome-wide data sets were prepared using our pre-imputation program and >20 were checked post-imputation. Using the post-imputation program we were able to easily identify a wide range of issues such as AF differences and missing or duplicated regions.

1863F

Asynchrony: **A framework for asynchronous software tested with bioinformatics use cases.** *N.R. Robertson[1,2], N.W. Rayner[1,2], M.I. McCarthy[1,2].* 1) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK.

Modern computer processors are being released with an increasing number of independent cores; however standard synchronous programming techniques do not facilitate the use of more than one core. Once a program is developed it is necessary to manually parallelize the work by running multiple executions concurrently, adding burden and complexity for the end user. We have developed *Asynchrony,* an asynchronous software framework for Java, which enables the programmer to organize a program into a multitude of sub-tasks that can be executed in parallel. We have been able to evaluate and expand this framework by building a series of tools to support key bioinformatics use cases such as file processing, sorting, compression, and indexed file access. Our software is able to scale across available processors, reducing execution times and improving ease of use. By utilizing multiple cores, we can show significant performance benefits over standard compression tools such as *gzip* and *bgzip*, while closely following the performance profile of the parallelized *gzip* implementation, *pigz*. Using 8 cores, our software takes 50 seconds to gzip encode a 3 GB file of 147 million records from dbSNP; *bgzip* manages the same task in over 5 minutes. For indexed file access, we provide an alternative to the block compression index software *tabix*, a tool widely used by bioinformatics researchers to access files by genomic position. In our test dataset of 147 million records, *tabix* is able to locate 10K search terms in 5.89 seconds. While our software is targeted at a more general use case, using 8 cores it can achieve the same lookup in 4.56 seconds, a performance we hope to improve in future iterations. We have extensively tested these tools on large datasets. For the interim release of UK Biobank data (152,732 individuals, 847,441 markers) have used the toolset to transpose and compress 1.8 TB of individual-wise intensity data into 753 GB of marker-wise gzip indexed data. Intensity data can now be randomly accessed in milliseconds without laboriously parsing whole files to generate cluster plots. Our suite of tools is available for public use immediately. The framework itself will be made available in pre-release form.

1864W

Alternate-scaffold aware variant calling in whole genome sequencing.

P.N. Robinson[1,2], M. Jäger[1], M. Schubach[1], T. Zemojtel[1], K. Reinert[2]. 1) Institute for Medical Genetics, Charite-Universitätsmedizin, 13353 Berlin, Germany; 2) Institute for Bioinformatics, Freie Universität Berlin, 14195 Berlin, Germany.

The last two human genome assemblies have extended the previous "golden-path" (linear) paradigm of the human genome to a graph-based model in order to better represent regions that display too much variability in human populations to be adequately represented by a single haploid sequence. In this work, we explore challenges and opportunities presented by the new genome model for variant calling in whole-genome sequencing (WGS). We show that stretches of sequences that are largely but not entirely identical between the primary assembly and an alternate scaffold and can thereby result in multiple variant calls against regions of the primary assembly that correspond to the reference sequence of the alternate scaffolds. In WGS analysis, this results in characteristic and recognizable patterns of variant calls (termed alignable scaffold-discrepant position or ASDP). We present a heuristic algorithm that can be used to flag these ASDPs in VCF files and annotate the most likely scaffold genotype. In 121 in-house genomes, on average 51.1±3.8 of the 178 regions were found to best correspond to an alternate scaffold rather than the primary assembly sequence, and filtering these genomes with our algorithm lead to the identification of 7864 variant calls per genome as ASDPs (and thus potentially false positive calls of variants against the primary assembly), or 6.51% of all variants located in the 178 regions. We found that 376 of 791 GWAS hits located within one of the regions corresponded to ASDPs. Since the distribution of the alternate loci shows population-specific biases, this should be taken into account in future association studies. We implement our algorithm as a Java application called ASDPex, which is made available under a BSD2 license.

1865T

Hail: An open-source framework for scalable genetic data analysis.

C. Seed[1,2,3], A. Bloemendal[1,2,3], J. Bloom[1,2,3,4], J.I. Goldstein[1,2,3], T. Poterba[1,2,3], B.M. Neale[1,2,3]. 1) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 4) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

The widespread application of massively parallel sequencing for complex trait analysis offers unprecedented power to link genetics with disease risk. However, these projects pose substantial challenges of scale and complexity, making even trivial analytic tasks increasingly cumbersome. To address these challenges we are actively developing Hail, an open-source framework for scalable genetic data analysis.The foundation of Hail is infrastructure for representing and computing on genetic data. This infrastructure builds on open-source distributed computing frameworks including Hadoop and Spark. Hail achieves near-perfect scalability for many tasks and scales seamlessly to whole genome datasets of thousands of individuals. On top of this infrastructure, we have implemented a suite of standard tools and analysis modules including: data import/export, quality control (QC), analysis of population structure, and methods for performing both common and rare variant association. Simultaneously, we and other groups are using Hail to manage the engineering details of distributed computation in order to develop and deploy new methods at scale.In addition, Hail exposes a high-level domain-specific language (DSL) for manipulating genetic data and assembling pipelines. As an example, porting a rare-variant analysis from Python to Hail reduced the number of lines of code by ~10x and improved performance by ~100x. We aim to grow Hail into a scalable, reliable and expressive framework on which the genetics community develops, validates, and shares new analytic approaches on massive datasets to uncover the biology of disease.

1866F

Genotype data quality analysis at the next order of magnitude in the Department of Veterans Affairs' Million Veteran Program (MVP). *Y. Shi[1], T. Webster[2], H.S. Hunter-Zinck[1], M. Li[1,11,12], C. Dow[2], N. Sun[3,9], J. Huang[1], C. Pan[4,7], J. Gelernter[3,10], T.L. Assimes[4,7], P.S. Tsao[4,7], C.J. O'Donnell[1,6], E.R. Hauser[5,8], H. Zhao[3,9], S. Pyarajan[1,6] for MVP Genomic Working Group.* 1) VA Boston Healthcare System, Boston, MA; 2) Thermo Fisher Scientific, 3420 Central Expressway, Santa Clara, CA 95051; 3) VA Connecticut Healthcare System, West Haven, CT; 4) VA Palo Alto Health Care System, Palo Alto, CA; 5) Durham VA Medical Center, Durham, NC; 6) Harvard Medical School, Boston, MA; 7) Stanford University School of Medicine, Stanford, CA; 8) Duke University, Durham, NC; 9) Yale School of Public Health, New Haven, CT; 10) Yale School of Medicine, New Haven, CT; 11) University of Utah, Salt Lake City, UT; 12) Johns Hopkins University, Baltimore, MD.

The U.S. Department of Veterans Affairs (VA) has initiated the Million Veteran Program (MVP) to implement personalized medicine in the care of U.S. veterans. Integration of a patient's genetic information with the clinical record is an essential step in personalized medicine. Here, we describe genotypic data from the first 200,845 patients in the MVP. The MVP encompasses a biobank of samples of consented veterans as well as nationally consolidated clinical information on the same patients from the VA's electronic health record system. The first installment of data includes genotypes of 200,845 patients genotyped with a custom Affymetrix Axiom MVP_1.0 array of 668,513 genetic markers in two independent laboratories and analyzed in 45 experimental batches. The initial quality control was performed separately in each laboratory following Affymetrix Best Practices to remove low quality samples. Additional quality analysis was performed on samples that passed initial quality control and were delivered to the VA. After genotype re-calling within the VA, the call rate for samples is approximately 99%. After a stringent genotype cleaning procedure, we retain 92% of probesets and 97% of markers. Over all samples, the median fraction of missing genotype calls per sample is less than 4%. About 3% of the samples were intentional duplicates, and we conducted concordance analysis using these samples. Concordance between genotypes of intentionally duplicated samples is extremely high with nearly 100% concordance for all duplicate pairs when analyzing common markers (minor allele frequency (MAF) greater than 5%). Even for extremely low frequencies, markers with MAF of 0.005% to 0.01% are still exactly concordant for approximately 80% of duplicate pairs with at least one copy of the minor allele. We also explored population structure of the samples and found substantial diversity in genetic ancestry. Approximately 71% of our samples are of European ancestry but 18% of samples have inferred admixture with approximately 5% of the total sample with admixture from three or more continents. The completed quality control analysis of the first 200,845 samples demonstrates that the MVP is a high quality and diverse resource for researchers seeking to understand the complex interactions between genetics, the environment, and medicine.

1867W

Development of a long indel detection method using the realignment of the misaligned reads. *D. Shigemizu[1,2,5], F. Miya[1,2], A. Fujimoto[3], K.A. Borovich[2], S. Okuda[4], T. Tsunoda[1,2,5].* 1) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 2) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Department of Drug Discovery Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 4) Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 5) CREST, JST, Japan.

Insertions and deletions (indels) have been implicated in dozens of human diseases through radically altered gene functions by short frameshift indels as well as long indels. However, the accurate detection of these indels from next-generation sequencing data is still challenging, particularly for long indels (50bp), due to the short DNA sequencing reads. In this study, we developed a new method that detects these long indels using BWA soft-clipped fragments (unmatched fragments in partially mapped reads) and unmapped reads. We report on a comparison of our detections with those from GATK HaplotypeCaller, which was recently extended to detect long indels using *de novo* assembly, using whole exome sequencing data from a HapMap-JPT sample, NA18943. The number of long indels detected by our method and the GATK were 70 and 34, respectively, and approximately 20% of our detections (13/70) were in common with those of GATK. On the other hand, while all of the long indels in GATK were less than 121 base pairs, more than half of our detections (36/70) were longer, ranging from 121bp to 2,792bp. To show that our method is effective in longer indel detection, we further performed Sanger sequencing of 6 randomly selected indels from these 36, and all were confirmed. Though we will perform Sanger sequencing verification for all of the long indels and estimate the false positive and false negative rates of our method, we expect that this methodology will contribute to the discovery for long indels associated with human diseases in the near future.

1868T

Multi-omic data exploration using pathway and correlation analysis in Agilent's GeneSpring. *S. Shukradas[1], D. Roy Choudhury[1], B. Prasad[1], M. Rosenberg[1], D. Renu[2], P. Tata[2], A. Wandycz[1].* 1) Agilent Technologies, Santa Clara, CA; 2) Strand Life Sciences, Bangalore, India.

Scientists today face serious challenges as they try to analyze increasingly large and more complex sets of data, such as those generated by genomics, transcriptomics, proteomics, and metabolomics experiments. GeneSpring provides a platform for data analysis and integration enabling comparisons across a variety of data types allowing for identification of co-regulated genes, proteins and metabolites in an intuitive way. An integrated approach for analyzing multi-omic data can significantly alleviate bottlenecks and provide valuable information that is often missed in individual experiments. One successful method for translating diverse analytical data into biological understanding can be achieved by projecting and visualizing experimental data onto the curated biological pathways (WikiPathways/BioCyc/KEGG) or literature-derived networks. GeneSpring makes it easy to resolve nomenclature inconsistencies by providing a powerful built-in tool, Agilent BridgeDB. This resource automatically links an annotated metabolite, protein, or gene identifier in the experiment to the corresponding identifier used by the pathway database, resulting in better annotation coverage for downstream analysis. Complementing more traditional bioinformatics techniques, correlation tools and specialized visualizations allow investigators to identify co-regulated genes, metabolites, and proteins in an intuitive and easy-to-use manner. In this analysis, we demonstrate how GeneSpring can be used to study correlation of expression of transcriptomic and proteomic data to identify curated signaling pathways that might be deregulated specifically in Glioblastoma (GBM).

1869F

Comprehensive analysis of the spatial distribution of missense variants in protein structures reveals patterns predictive of pathogenicity. *R.M. Sivley[1,4], W.S. Bush[3,5], J.A. Capra[2,4].* 1) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Biological Sciences, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 4) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 5) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

The spatial distribution of genetic variation within protein structures is shaped by functional constraints and thus provides information about the importance of protein regions and the potential pathogenicity of protein alterations. We developed a statistical framework in which to model and compare spatial patterns of genetic variation across multiple distance scales. We then performed a comprehensive 3D analysis of protein-coding single-nucleotide variants (SNVs), including population-derived missense (ExAC, N=196,176) and synonymous (ExAC, N=122,963), pathogenic (ClinVar, N=4,827), and cancer (COSMIC, N=114,574) variants in 4,575 human proteins with solved structures in the Protein Data Bank. Population-derived missense and synonymous, pathogenic, and cancer variants have drastically different spatial distributions. Missense variants observed in the general population are found in nearly every protein in our dataset (N=4,523) and are typically dispersed throughout structure, while synonymous variants trend towards spatial randomness. In contrast, nearly 20% (88 of 453) of proteins with at least three pathogenic variants exhibit significant spatial clustering in 3D. This is substantially greater than the 0.3% (15 of 4,547) of proteins with significant clustering of cancer variants, in which there has been considerable recent interest. Indeed, we find that pathogenic variation is more clustered than recurrent cancer variation in two thirds (200 of 301) of proteins with sufficient occurrence of each. Motivated by the significant 3D spatial clustering of pathogenic variation and the general dispersion of putatively benign variation, we hypothesized that analysis of variants' spatial distributions could provide a new line of evidence to use in the prioritization of variants of unknown significance (VUS). We will present several examples that demonstrate the predictive performance of our approach compared to common methods for the prioritization of candidate mutations. These findings suggest that explicit modeling of the spatial distribution of protein-coding variation has great promise to complement existing strategies for VUS prioritization and pathogenicity prediction.

1870W

EGA: Towards distributed consented genetic and phenotypic data access. *J.D. Spalding[1], M. Alberich[2], J. Almeida-King[1], A. Carreño-Torres[2], P.A. Garcia[1], A. Gil[2], J. Kandasamy[1], G. Kerry[1], O.M. Llobet[2], A. Lloret-Villas[2], M. Moldes[2], A. Senf[1], S. de la Torre[2], S. Ur-Rehman[1], P. Flicek[1], A. Navarro[2,3,4], H. Parkinson[1], J. Rambla[2].* 1) European Genome-phenome Archive, European Bioinformatics Institute (EMBL-EBI), Cambridge, Cambridgeshire, CB10 1SD, UK; 2) Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain and Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Institute of Evolutionary Biology, Barcelona, Spain; 4) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

The European Genome-phenome Archive (EGA) is a controlled-access archive, which facilitates the secure long-term storage and controlled distribution of genetic and phenotypic data, for the purpose of biomedical research. Access to data is via application, and data delivery is encrypted to protect the participants' identities. As of April 2016, the EGA stores over 3.9 petabytes of data derived from over 700,000 unique samples in over 900 studies. The EGA is developing the technology to become a distributed and federated archive. This addresses both scalability issues, where the data can be hosted close to the compute, and Ethical, Legal, Social Implications (ELSI) issues, where data cannot leave a certain jurisdiction, while facilitating data access and discovery. For data access this includes implementing and extending the Elixir Authentication and Authorisation Infrastructure (AAI) to encompass the EGA use-case for greater Levels of Assurance (LoA) (e.g. step-up authentication). Accessibility to the raw data in EGA is improved by using AES encryption technology to facilitate random-access to indexed encrypted files, and EGA is developing a FUSE layer to allow native applications to access encrypted data directly. This allows EGA data to be accessed directly from a third-party application of platform, for example Galaxy. These technologies are available as a set of Docker containers, or a pre-built virtual machine, facilitating the implementation of new EGA nodes. To improve data discoverability, EGA has implemented the v0.3 3-tier Beacon (https://www.ega-archive.org/beacon/#/), improved integration with BioSamples - a sample archive for accessioning samples and preventing duplication (https://www.ebi.ac.uk/biosamples/), and meta-data annotation of both existing and new meta-data via Zooma (http://www.ebi.ac.uk/spot/zooma/) and the Ontology Lookup Service (http://www.ebi.ac.uk/ols/index). Our collections include reference data for rare and common diseases, including data derived from the UK10K project, WTCCC, Human Induced Pluripotent Stem Cells Initiative (HipSci), RD-Connect and the International Cancer Genome Consortium (ICGC), and improved meta-data curation allows enhanced queries across these datasets. The EGA is maintained by European Bioinformatics Institute (EMBL-EBI) and the Center for Genomic regulation (CRG) and is available at http://www.ega-archive.org.

1871T

The UCSC Genome Browser: New display modes to facilitate study of clinical data. *M.L. Speir, G.P. Barber, C.M. Lee, K.R. Rosenbloom, A.S. Hinrichs, C. Tyner, L. Guruvadoo, B.T. Lee, A.S. Zweig, D. Karolchik, R.M. Kuhn, W.J. Kent.* Genomics Institute, University of California Santa Cruz, Santa Cruz, CA.

The University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu) is a free, web-based tool that allows researchers to visualize and explore large, varied data sets from a number of sources including ClinVar, OMIM, GTEx, and GENCODE. In addition to providing a large repository of browseable data, we enable users to upload and visualize their data sets alongside those we offer. The Genome Browser also provides numerous other web-based tools for exploring and analyzing these data in different ways. One of our newest visualization tools, multi-region configuration, features the ability to display multiple, sometimes distant, regions of the genome alongside one another in the same Genome Browser image. The exon-only and gene-only modes of this feature let users remove both intronic and intergenic sequence or just intergenic sequence from the display, respectively, allowing them to focus in on the coding regions of the genome. These modes are based on a pre-selected gene set determined to be appropriate for that genome. The display will react to changes and filters applied to this gene set, removing regions from the display that are no longer covered by exons or genes. Another configuration option of the multi-region feature allows for user-defined regions that are then combined into a single Genome Browser image, permitting the visualization of multiple genes, variants, or other important genomic regions. These user-defined regions are specified using a remote BED file that contains basic position information. A final multi-region configuration option enables researchers to view the sequence of an alternative locus inserted into its corresponding region in the reference genome, rather than being displayed as an isolated chromosome. While multi-region configuration adds a great deal of customization and configuration to the Genome Browser display, the current implementation has limitations that we plan to address in future releases. These improvements include allowing gene set selection for both the exon- and gene-only modes and expanding the custom region mode to support additional ways to specify regions.

1872F

SurvivalGWAS_SV: Software for the analysis of genome-wide association studies of imputed genotypes with "time-to-event" outcomes. *H. Syed, A.L. Jorgensen, A.P. Morris.* Department of Biostatistics, Block F, Waterhouse building, University of Liverpool, Liverpool, United Kingdom.

   Methodology for the analysis of genome-wide association studies (GWAS) has focused primarily on binary (case-control) phenotypes and quantitative traits. Software implementing these methods, such as PLINK and SNPTEST, can efficiently handle the scale and complexity of genetic data from GWAS, allowing for imputed genotypes at millions of single nucleotide polymorphisms (SNPs). However, the key outcome of interest in many genetic studies, particularly those in pharmacogenetics, is often "time to event" data (such as death, disease remission or occurrence of an adverse drug reaction), which cannot adequately be modelled by existing software. To address this challenge, we have developed the SurvivalGWAS_SV software implemented using C# and run on LINUX operating systems. SurvivalGWAS_SV is able to handle large scale genome-wide imputed data, modelling time to event outcomes under an additive dosage model in the number of minor alleles at a SNP. Analysis can be performed using either Cox proportional hazards or Weibull regression models, the latter capable of analysing accelerated failure time and non-proportional data. The software also allows for multiple covariates and incorporation of SNP-covariate interaction effects. Running analysis of 1.5 million SNPs across 22 chromosomes for 1000 individuals will take approximately 6 hours to complete, dependent on number of computer cores used and processor speed. In conclusion, we introduce a new console application analysis tool for GWAS with time to event outcomes. SurvivalGWAS_SV can analyse large scale GWAS datasets efficiently and effectively, without incurring memory issues. With its particular relevance to pharmacogenetic GWAS, SurvivalGWAS_SV will enable discovery of genetic biomarkers of patient response to treatment for a range of complex human diseases, ultimately allowing personalisation of therapeutic intervention.

1873W

Genetic architect: Discovering genomic structure using learned neural architectures. *S. Targ[1], L. Deming[1], N. Sauder[2], D. Almeida[2], C.J. Ye[1].* 1) UCSF Institute for Human Genetics, San Francisco, CA; 2) Enlitic, San Francisco, CA.

   Deep learning demonstrates excellent performance on tasks in computer vision, text and many other fields, but has only recently been applied to challenges in genomics. While the genetic code allows us to annotate the 5% of the genome encoding proteins, we do not have a "grammar" for decoding the rest of the non-coding sequences important for gene regulation, evolution of species and susceptibility to diseases. The design and application of neural network architectures specific to functional genomic data is a promising approach for automating the annotation of genomic information into forms that humans can grasp. We develop a broadly applicable search algorithm to discover two models, AttentionNet and PromoterNet, which combine simple and cutting-edge neural network building blocks to learn patterns in genomic data and identify the most important sequence motifs predictive of TF binding and gene expression. Our learned models contain convolutional layers, residual blocks, an LSTM recurrent layer, and an attention-based dimensionality reducing step, followed by fully connected layers. Previous deep learning approaches to genome annotation use convolutional networks, which are ideal for detecting local features. However, more closely approximating the structure of genomic information would take into account that a real genome is a sequence, not a disjointed set, of local features – an input type on which recurrent architectures generally excel. We use AttentionNet to learn general structure and sequence features from 108 K562 ENCODE ChIP-seq datasets recovering known JASPAR motifs for 42/57 tested TFs and achieving an AUC of 0.933, significantly outperforming other models. We also develop a novel benchmark regression task of predicting lineage-specific gene expression from ImmGen Consortium microarray data using input genomic sequences +/- 1kb around TSSs. Our approach, PromoterNet, obtains an average Pearson r correlation of 0.6, outperforms existing models, and identifies both conserved TSS proximal core promoter elements and lineage-specific features in genes selected for maximal variance in expression. Our results show these learned architectures are well-suited to understand the regulatory structure of multimodal genomic data including ChIP-seq, RNA-seq, ATAC-seq and DNase-seq, genomic position, and chromatin conformation and capture effects of local patterns and long-range sequential dependencies on functional genomic outcomes.

1874T

Turning *big data* **into** *small data* **through crowdsourced curation: Integrating all types of medical and scientific knowledge.** *T.D. Taylor, N. Kumar.* Laboratory for Integrated Bioinformatics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan.

　Big data in the form of scientific media comes in many languages and formats: journal articles, books, images, videos, etc. Text media often includes additional embedded or associated information such as tables, figures and supplementary data. While there are many resources for browsing, searching and annotating some of this media, there is no single place to search them all at once, and generalized search engines do not allow for the comprehensive and precise searches researchers require. One could argue that any scientific media that is on the web is therefore connected, but much of it remains offline or is inaccessible and is therefore neither discoverable nor connected. To address these issues, we created **iCLiKVAL** (iclikval.riken.jp), a web-based tool that uses the power of crowdsourcing to accumulate annotation information for all scientific media found online, and potentially offline. Annotations in the form of key-relationship-value tuples (any language) added by users through a variety of methods can make vast amounts of unstructured data easier to comprehend and visualize by turning it into *small structured data*. This allows for richer data searches and discovery of novel connections by integrating all forms of scientific knowledge through common terminology. Users can create or join common interest groups to annotate related media together as a community. They can also create and edit their own controlled vocabulary lists, or import established vocabularies such as MeSH and GO terms, and then select which lists to use for auto-suggest terms in each of the form fields. Within user groups, vocabulary and bookmark lists can be shared. We also implemented a notification center, several customization options, and annotations page where users can view and edit all of their own annotations. Most of the pages, such as annotations, bookmarks, search history, reviews and vocabularies, are searchable, sortable and filterable, so users can quickly find what they need. Besides PubMed articles, we recently added annotation support for YouTube, Flickr, SoundCloud and anything with a DOI name, allowing for the inclusion of hundreds of millions of media objects. We also created a Chrome Browser extension that allows any non-password protected online media to be bookmarked and annotated (even offline), to facilitate the annotation process. The iCLiKVAL database is completely searchable, and all of the data is freely available to registered users via our API.

1875F

CRAVAT 4.2: Informatics tools for high-throughput analysis of exome variants. *C. Tokheim[1,5], C. Douville[1,5], D.L. Masica[1,5], R. Kim[2,5], R. Bhattacharya[3], N. Niknafs[1], D. Gygax[2], K. Moad[2], M. Ryan[2], R. Karchin[1,4].* 1) Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD; 2) In Silico Solutions, Fairfax, VA; 3) Department of Computer Science, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD; 4) Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, MD; 5) Authors contributed equally.

　Analyzing thousands or millions of exome variants causes a daunting prioritization challenge. CRAVAT is a high-throughput web-based resource for analysis of exomic variants (single base substitutions and indels). CRAVAT provides scoring, sorting, filtering, and interactive visualizations to assist with identification of important variants. Variants relevant to cancer are particularly emphasized. Variants are submitted in either the standard Variant Call Format (VCF) or a simple tab-delimited format; subsequently results are quickly emailed to users in spreadsheet and machine-friendly format. Summaries are provided on the variant- and gene-level, which include population Minor Allele Frequency (MAF), occurrence in COSMIC, mutation call quality, and frequency of gene variants. CRAVAT 4.2 is enhanced with new variant scoring, an interactive viewer, detection of variant hotspot regions, and expanded tool ecosystem. *In silico* pathogenicity scores are now provided for all non-silent mutation types in a unified statistical framework that allows one-pass sorting. The interactive viewer includes lollipop diagrams, population MAF widget, dynamic sorting and columns, integrated display of variants on protein structure, and a pathway enrichment viewer. The protein structure viewer is enhanced with a new algorithm that detects mutation hotspot regions in 31 human cancer types. Users with protected data can now use a docker version for analysis on local computers or in the cloud. The utility of CRAVAT is exhibited by a wide user base, which has submitted over 6,000 jobs annotating 355 million variants in the past year. CRAVAT provides streamlined analysis of exomic variants, using multi-faceted information for improved variant prioritization.

1876W

Homogenization of dbGaP data. *C. Tripathi, G. Ellis, L. Akhabir, D. Daley.* University of British Columbia, Vancouver, British Columbia, Canada.

   Introduction: The overarching goal of sharing data and collaborating is to decipher complex trait etiologies by genetic, epigenetic and genomic means. The database of Genotypes and Phenotypes (dbGaP) stores and distributes results of genotype-phenotype association studies. To harness the full potential of this valuable resource, integration of data from multiple studies is necessary. Studies submitted to dbGaP are from different time points and have thus used different genotyping chips and human reference genome versions (hrg). Consequentially, SNPs on these chips differ in aspects such as: number, identifiers (rs number), strand genotyped (+/-) and genomic coordinates. Such differences present challenges to data integration. To address the need to homogenize dbGaP data, we have designed a suite of web-based tools. We will demonstrate utility of the tools by taking example of one such tool: to compute overlap of SNPs between genotyping chips. The tool can be used choose the most appropriate chip for a hypothesis and to choose the custom content for chips.**Methods:** The following tools have been coded in Java to form **suite of new tools**: 1) Tool to estimate overlap of SNPs between genotyping chips 2) Tool to homogenize SNPs in studies by identifying SNPs common across all studies of interest 3) Tool that homogenizes variable names in a semi-automatic fashion.Tools already available are integrated to form **suite of existing tools**: 1) NCBI Remap tool to map coordinates of SNPs to latest version of human reference genome 2) NCBI Remap to update SNP rs numbers 3) A plink command to homogenize DNA strands to which alleles are mapped.**Results:** As an example of how the tool to compute SNP overlap works, we used it to estimate SNP overlap between the high density Illumina Genome Wide Association Studies (GWAS) Omni5 chip with the NHGRI GWAS catalog. We found that of the 4,415,047 SNPs on Omni5, the overlap with: directly genotyped SNPs of GWAS catalogue (8,696 SNPs) is 7,420; GWAS cancer (1,610 SNPs) is 1,237; GWAS cardiovascular (1,835 SNPs) is 1,276; GWAS alzheimer (313 SNPs) is 218; GWAS dementia (829 SNPs) is 550 and GWAS diabetes (807 SNPs) is 620.**Conclusion:** Our web-based tools will facilitate homogenization of dbGaP and similar data. On request, the tool to estimate SNP overlap has been distributed to Illumina and fellow research labs. On completion, all tools will be shared with collaborators and other researchers to catalyze data integration and utilization.

1877T

Automated genotypic imputation of PAGE II data using scientific workflows. *K. Vahi[1], S. Buyske[2], L. Zhou[3], T.C Matise[3], E. Deelman[1].* 1) Science Automation Technologies, USC Information Sciences Inst, Marina del Rey, CA; 2) Department of Statistics & Biostatistics, Rutgers University Hill Center, Piscataway, NJ; 3) Department of Genetics, Rutgers University Piscataway, NJ.

   The PAGE II (Population Architecture using Genomics and Epidemiology) study has genotyped 50,000 samples using MEGA, an Illumina high density consortium-built array array. PAGE is also imputing genotypes on an additional 50,000 subjects who were genotyped in 20 different GWAS studies. Imputation on a PAGE sub-study, using SHAPEIT and IMPUTE2 with the 1000 Genomes Project reference panel, has over 650 computational tasks.   Imputation can be parallelized by chromosome to reduce total processing time. Researchers often write custom batch scripts to setup complex computational pipelines. Such scripts are responsible for both making the data available to the batch jobs and the application specific logic that identifies the processing steps and pipeline structure. This approach results in complicated error recovery semantics, and often can only run on the computation resource for which it was developed. Keeping these concerns in mind, we modeled the PAGE production imputation pipeline as a computational workflow using the Pegasus Workflow Management System. Pegasus represents the application pipeline in an abstract form, agnostic of the resources available and the location of data and executables. It compiles these workflows into an executable form and discovers location of data and executables by querying catalogs. The executable workflow is executed on potentially remote and distributed resources, such as remote clusters and clouds. Pegasus has well defined error recovery semantics. Pegasus supports automatic job retries in case of job failures, workflow-level checkpointing and fallback to alternative data sources for staging data. When a fatal error is encountered, a rescue workflow containing only the remaining work is provided. Pegasus provides monitoring and debugging tools that enable users to easily identify what steps in the pipeline failed and the reason for failures.The imputation pipeline is currently in production and is setup to run on the Rutgers Genetics Cluster and has been applied to datasets from seven PAGE studies with minimal user intervention. When a new study dataset becomes available, it takes the researcher only a few minutes to launch the imputation workflow. The clean separation of application logic and underlying compute infrastructure, ensures that the workflow description generation code does not need to be altered when porting to a different computing environment such as Amazon EC2. The workflow will be shared publicly on GitHub.

1878F

Building high-quality, de novo genome assemblies by scaffolding next-generation sequencing data with BioNano's next-generation genome mapping. *J. Wang, A.W. Pang, E.T. Lam, W. Andrew, T. Ananthara- man, A. Hastie, W. Stedman, H. Sadowski, M. Saghbini, Z.Y. Zhu, M. Austin, M. Borodkin, H. Cao.* BioNano Genomics, San Diego, CA.

Combining next generation sequencing (NGS) data and next-generation mapping (NGM) data from BioNano Genomics's Irys System has recently been demonstrated as a promising workflow to produce affordable, high-qual- ity and chromosome-scale *de novo* genome assemblies.   To broaden the scope of such a hybrid-scaffolding approach to highly fragmented NGS data, we described a novel workflow that utilizes more than one nicking endonucle- ase to increase the information density of genome maps. We generated two independent sets of genome maps, each with a different nicking endonu- clease and developed a novel algorithm that uses the NGS sequences as a bridge to merge single-enzyme genome maps into combined genome maps that contain the sequence motif patterns from both nicking enzymes. Since the genome maps were generated independently they serve as orthogonal sources of evidences to detect and correct assembly errors in either sequence assemblies or genome maps. The complementarity of different data type also greatly improves the contiguity of the merged genome map while doubling the information density. This in turn allows for anchoring of shorter NGS se- quences in the final scaffold.   We benchmarked our approach by scaffolding a sequence assembly from Illumina short-read sequencing with two sets of BioNano genome maps from the human NA12878 sample. The NGS datasets were obtained from the NCBI's short read archive. It contains 630,209,622 reads with 251 bp per read and were assembled using DISCOVAR de novo. We produced a final scaffold that increased the contiguity of the sequence assembly from an N50 of 0.179 Mb to 18 Mb and anchored more than 80% of the total NGS sequences in length while reducing the number of chimeric joins – two distal regions in the genome that were erroneously joined together by assembly algorithms -- from 52 to 9. Compared to the published single-en- zyme hybrid-scaffolding approach, the wo-enzyme approach improved the scaffold contiguity by 300% while anchoring ~30% more sequence contigs, leading to the incorporation of additional 200Mb of sequence in the final scaffolds.   This new approach can greatly expands the type of NGS data that can be integrated with BioNano genome maps to produce highly accurate and contiguous assemblies for complex genomes.

1879W

Bayesian latent variable models for single-cell trajectory learning. *C. Yau, K. Campbell.* Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxon, United Kingdom.

The transcriptomes of single cells undergoing diverse biological processes - such as differentiation or apoptosis - display remarkable heterogeneity that is averaged over in bulk sequencing. Single-cell sequencing itself offers only a snapshot of these processes by capturing cells of variable and unknown progression through them. Consequently, one outstanding problem in sin- gle-cell genomics is to find an ordering of cells (known as their pseudotime) that best reflects their progression, for which several computational methods have been proposed. To date, the vast majority of such methods emphasise transcriptome-wide 'data-driven' approaches that assume no prior knowledge of gene dynamics along the trajectory during inference. The suitability of the inferred trajectory is typically assessed by post-hoc examination of a set of marker genes to ensure the inferred behaviour aligns with prior assumptions. Furthermore, most current methods are algorithmic and rely on heuristics as opposed to probabilistic models, which in the context of bifurcations requires the pseudotimes to be first inferred prior to the identification of any bifurcation events. Here we introduce a general probabilistic framework for single-cell trajectory learning based on Bayesian non-linear factor analysis. First, we demonstrate how such a framework may be used to integrate prior knowledge of gene behaviour in trajectory inference. By assuming a parametric form of gene expression evolution across pseudotime we can place informative priors on parameters that govern gene behaviour and remove the need for subjective post-inference checks. We demonstrate that a small panel of marker genes is often sufficient to achieve comparable results to transcriptome wide alterna- tives and that such a method can be used to recover trajectories correspond- ing to known pathways in the presence of heavily confounding effects. Our second application is to modelling bifurcations in single-cell data. By consid- ering a Bayesian mixture of factor analysers we simultaneously infer both the pseudotimes and branching behaviour of the cells, which is unique compared to existing methods. We derive a Gibbs sampler that allows for fast inference across hundreds of cells while accounting for the zero inflation that is pertinent to single-cell RNA-seq data. Notably, by using a Bayesian framework we can integrate prior knowledge of branch-specific gene behaviour allowing for robust inference on challenging datasets.

1880T

SeqArray - A storage-efficient high-performance data format for WGS variant calls. *X. Zheng[1], S. Gogarten[1], M. Lawrence[2], A. Stilp[1], D. Levine[1], B. Weir[1], C. Laurie[1].* 1) Department of Biostatistics, University of Washington at Seattle, Seattle, WA; 2) Bioinformatics and Computational Biology, Genentech, Inc., South San Francisco, CA.

Whole-genome sequencing (WGS) data is being generated at an unprecedented rate. Analysis of WGS data requires a generic and flexible data format to store the different types of DNA variation. The variant call format (VCF) is a general text-based format developed to store genotypes and annotations, however VCF files are large and data retrieval is relatively slow. Here we introduce a new binary data format implemented in the Bioconductor package "SeqArray" for storing variant calls which provides the same capabilities as VCF. SeqArray stores variant data in an array-oriented manner with multiple high compression options including the Lempel-Ziv-Markov chain algorithm (LZMA). Benchmarks using real data sets show that the file sizes for 1000 Genomes Phase 3 genotypes are 14.0G (VCF), 12.3G (BCF, binary VCF) and 2.4G (SeqArray). For an unpublished WGS study with 9K subjects and 140M variants, the VCF file size is 66.4G, BCF 50.1G and SeqArray 14.4G. Calculating allele frequencies in SeqArray is at least 2x faster compared to other tools that use VCF or BCF files. Other Bioconductor packages built on top of SeqArray provide general functionality in WGS analysis. For example, parallel implementations of relatedness and principal component analysis with SeqArray format are enabled in the package SNPRelate. This work was supported, in part, by NIH Grant GM099568. .

1881F

Utilizing allele specific expression to identify cis-regulatory variants. *J. Zou[1], S. Shifman[3], J. Sul[4], J. Ernst[1,5], E. Eskin[1,2].* 1) Department of Computer Science, University of California, Los Angeles, CA, USA; 2) Department of Human Genetics, University of California, Los Angeles, CA, USA; 3) Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; 4) Deparment of Psychiatry and Biobehavioral Science, University of California, Los Angeles, CA, USA; 5) Deparment Biological Chemistry, University of California, Los Angeles, CA, USA.

Understanding cis-regulatory control of gene expression is crucial towards understanding complex diseases. Traditional methods for identifying sequence variants, such as the expression quantitative loci (eQTL) approach, have identified many genetic variants correlated with changes in gene expression. However, these methods use total RNA expression, which fails to account for differences between the two alleles of diploid organisms. Allele-specific analyses account for these differences and benefit from reduced environmental and trans-acting influences. We implemented a method to identify variants that best explain observed patterns of allele specific expression (ASE) across individuals. We applied this method using RNA-seq data and genotype data collected from the Genotype-Tissue Expression (GTEx) Project. Increased power from allelic discrimination in our method allows us to identify variants with high functional relevance based on gene annotation and epigenomics data.

1882W

A scalable and secure genome archiving and communication system for the clinical enterprise. *R. Swaminathan[1], Y. Huang[1], E. Yu[1], J. Fitch[2], K. Lintner[1], P. White[2,3], S. Lin[1,3].* 1) Research Information Solutions and Innovation, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 3) Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, OH.

The plummeting costs and increased diagnostic yield of whole exome/genome sequencing have resulted in an increased use of this technology for clinical purposes.The resulting genomics data influx requires a novel technological approach to store the data, communicate with other Clinical Decision Support systems, interface with Health Information Exchange,and enable secondary use in research studies.The Hadoop framework has been at the forefront in dealing with "Big Data",harnessing its distributed storage and processing capabilities.The high scalability and low cost of Hadoop,along with its suite of authentication,authorization,and auditing schemes to maintain HIPAA compliance,makes it an ideal candidate for archiving large clinical genomic datasets.Similar to the Picture Archiving and Communication System (PACS) for radiology,we designed an open source Genome Archiving and Communication System (GACS) for clinical genomics that utilizes the Hadoop Filesystem (HDFS) for genomic data storage (FASTQ/BAM/VCF/GVCF);H-Base for real time querying of variants;Google Genomics compatible API for query interface;and a NIST-derived, enterprise-supported risk management framework for data security.We evaluated the performance and scalability of the GACS implementation for data ingestion and retrieval measures using both public data from the 1000 Genomes Project as well as a larger simulated dataset of 250,000 exomes.The average ingestion time was calculated as a function of the number of exomes/genomes in the database and the storage footprint,whereas retrieval is the response time as a function of the number of exomes/genomes in the database, tested for both clinical and research use cases.We also compared the performance from Hadoop with existing genomic analytical tools such as tabix(file data store) and GEMINI(relational database) to store and query genomic information,both of which are limited in their scalability.With increasing exomes within HBase,the average ingestion time for a batch of 100 exomes using the Hadoop system was approximately 3 minutes, 4 times faster than tabix.For data retrieval, although HBase outperforms tabix,the average response time for both tabix and HBase to process 100 concurrent requests is under 1 second.In summary, the reference implementation suggests the feasibility of using a highly scalable and secure archiving system for clinical genomics, which is also conducive for integrating clinical phenotypes and public genomic information.

1883T

IGLCBP - A cloud based clinical bioinformatics pipeline. *X. Liu, P. Kothi, D.C. Thach, W.S.W. Wong, S. Ravulapalli, K.B. Tokarz, S. Madhappan, A. Black, G. Eley, B.D. Solomon, R.K. Iyer, J.F. Deeken, J.E. Niederhuber.* Inova Translational Medicine Institute, Fairfax, VA.

Inova Genomics Laboratory (IGL) is a CLIA-certified clinical genomics laboratory within Inova Translation Medicine Institute that performs high-complexity molecular testing including NextGen sequencing. To meet the fast-expanding service needs we developed a flexible, scalable bioinformatics pipeline framework IGLCBP(**IGL C**linical **B**ioinformatics **P**ipeline).   IGLCBP is based on the open-source NextGen sequence analysis workflow system bcbio-nextgen (https://github.com/chapmanb/bcbio-nextgen). IGLCBP takes advantage of several built-in features from bcbio-nextgen that match well with clinical analysis's needs including a) pre-installed /configured top sequence analysis tools from the community; b) ease of benchmarking and comparing different tools; c) support of parallelization on the Amazon Web Service (AWS) cloud; d) built-in validation module against NIST's Genome in a Bottle (GIAB) genome standard for accuracy evaluation; e) support for Docker container service (www.docker.com) that makes deployment, versioning, and reproducibility all much more manageable. We further customize IGLCBP to suit our specific clinical mandates in the following areas: a) various QC report capabilities on sequencing, alignment, coverage, and variant calls; b) functions that automate the generation of comparison matrix on accuracy, and target coverage from using different variant calling workflows (e.g. GATK vs. Samtools, single caller vs. ensemble, etc.); c) data transfer functions to meet hospital's IT security and HIPPA compliance requirements; d) modules to evaluate reproducibility and reliability from sample re-sequencing; e) utilities that connect to the database backend for fast annotation and clinical interpretation.   Here we present our process to generate a production analysis pipeline for a clinical test using IGLCBP. To demonstrate the flexibility and accuracy of our system, we tested IGLCBP on a subset of clinically relevant genes enriched by Agilent's OneSeq Comprehensive Target Panel using the CEPH NA12878 sample. Results of comparing different variant discovery workflows using IGLCBP and with Agilent's own SureCall program are presented. We plan to open source IGLCBP to the clinical genomics community to make precision medicine more easily accessible in the near future.

1884F

A standard allele registry for the human genome. *S. Sherry, B. Holmes, E. Yaschenko.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

The NCBI Allele Registry addresses the problem of allele ambiguity in human genetic data. Sequence Variation objects are currently modeled as: a single identifier attached to an interval of sequence and alternative alleles; a semantic string in a recognized nomenclature system like HGVS; a structured (but possibly un-registered) row of data in a set of called human variants in VCF format; or a variant implicit in the CIGAR string of aligned next-gen sequence fragments. The lack of cardinality and precise identification of individual sequence alleles across these data sources requires each user of the data and every global repository of allele data to individually track these variants, discern equivalencies, and track or link accordingly. Given the tremendous recent growth in the size of the data sets– hundreds of millions of total genomic variants, and hundreds of thousands of them with clinical assertions – the Allele Registry was developed to provide a standardized, comprehensive set of reference-grade allelic identities that can be assigned and indexed within NCBI archives including ClinVar, dbSNP, GTR, dbGaP, and the Genotype Server. The opportunity to correctly query for alleles across resources offers important opportunities to eliminate data redundancy, join properties of alleles across resources, and present them coherently and efficiently to users as sequence annotation, population frequency estimates, phenotype associations, and clinical assertions. The reference set of allele identifiers developed by NCBI are freely available to external communities for clinical lab management, GWAS research, array development, and medical record integration.

1885W

High-throughput clinical reporting of gene panels with the Neptune Pipeline. *E. Venner[1], M. Bainbridge[1], C. Kovar[1], T. Chiang[1], S. White[1], M. Leduc[2], M. Murugan[1], W. Salerno[1], D. Muzny[1], R. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX;; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.

High throughput clinical analysis of DNA capture panels demands automated processing in order to provide timely and cost-efficient reporting for patient benefit. We developed Neptune, an automated analytical pipeline to sign-out and deliver clinical reports any site ordering a test. Our process relies on a highly-scalable Data Warehouse of curated variant and gene data that feeds into an automated reporting engine. After initial processing with the Human Genome Sequencing Center's Mercury Pipeline, custom annotation software within Neptune identifies variants of putative clinical relevance for manual review and possible addition to a dynamic "VIP" database of clinically relevant variation (cSNP, SNV, indel, CNV), drawing on both public resources (ClinVar, literature review) and internal data sets. Neptune relays these variants to a manual review interface, which captures the reviewer's input and updates the VIP database accordingly. Once all variants have been categorized, Neptune extracts known pathogenic variants using the VIP set. Neptune outputs an automated clinical pre-report populated with prioritized variants (or a negative report if no relevant variants are found), descriptive text and coverage statistics produced by the HGSC's ExCiD software. The VIP database currently houses 20,872 SNPs and 3,946 indels.   Initial data intake occurs in a HIPAA compliant environment, and samples are de-identified before moving into the CLIA lab. Neptune is engineered to provide the shortest possible turn-around times while still maintaining consistency and CAP/CLIA compliance. During initial validation, 20% of samples contained at least one reportable variant while the remainder generated a negative report. Future work will move further towards fully automated approaches for use on future clinical sequencing projects.   Early application includes reporting for the National Institutes of Health eMERGE network where more than 12,500 samples and a panel of ~100 genes will be processed in less than three years.

1886T

Developing an integrative method to improve GWAS inference using enhancer annotations. *Y.H. Hsu[1,2,3], T.H. Pers[4,5], R.S. Fine[1,2,3], J.N. Hirschhorn[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 3) Broad Institute of the MIT and Harvard, Cambridge, MA; 4) Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 5) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

Genome-wide association studies (GWAS) have identified many genetic variants associated with human traits and diseases. However, because more than 90% of GWAS variants fall in noncoding regions of the genome, it is often difficult to predict the functional effects of these variants and to uncover trait-relevant biology from GWAS results. We have previously developed a GWAS inference method, DEPICT (Data-driven Expression Prioritized Integration for Complex Traits), that combines datasets of gene expression, protein-protein interactions, mouse knockout phenotypes, and gene sets to prioritize important genes, pathways, and tissues/cell types for complex traits. In this project, we incorporated enhancer annotations from the FANTOM5 project into DEPICT to prioritize transcriptional regulators. We mapped the enhancers to downstream target genes by correlating the expression of proximal enhancer-gene pairs across cell types. We also identified upstream regulators for the enhancers by scanning their DNA sequences for transcription factor binding motifs from the Cis-BP database. We then summarized these regulatory relationships into a gene by regulator data matrix, in which each score indicates how strongly a gene is controlled by an upstream regulator. We used the gene by regulator matrix in DEPICT to prioritize regulators that are important for various traits, including height, inflammatory bowel disease, low-density lipoprotein, and schizophrenia. In some cases, we were able to further validate the prioritized regulators by showing that, in trait-associated loci, SNPs are more likely to overlap enhancers containing motifs for the prioritized regulators compared to other enhancers ($p < 0.05$). While additional work is needed to improve the gene by regulator matrix, our preliminary results indicate that integrating enhancer regulatory information into DEPICT can expand its utility for interpreting GWAS.

1887F

Using data-driven approaches to address clinical heterogeneity in complex traits. *J.B. Leader[1], A.O. Basile[2], H.L. Kirchner[1], C.B. Manney[1], A. Verma[1], M.D. Ritchie[1,2].* 1) Department of Biomedical and Translational Informa, Geisinger Health System, Danville, PA; 2) Department of Biochemistry and Molecular Biology, Center for Systems Genomics, The Pennsylvania State University, University Park, PA.

Complex diseases are often heterogeneous in nature; instead of representing a single disease, these conditions may in fact be comprised of multiple disorders, each with varying symptoms, clinical presentations, and thereby differing etiology. Trait heterogeneity is a confounding variable that has largely been overlooked in traditional genetic approaches. It complicates clinical patient diagnoses, and has been shown to decrease both the statistical power to detect associations and the risk attributed to susceptibility variants. The copious amounts of biomedical data present in the Electronic Health Record (EHR) system can be leveraged using data-driven approaches to overcome clinical heterogeneity and detect homogeneous patient subgroups. We performed *in silico* dataset evaluations to examine the performance of multiple machine learning algorithms in homogeneous data classification. The evaluated algorithms included classical clustering approaches such as hierarchical clustering and k-means analysis, random forests, and dimensionality reduction strategies such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). Further, various proximity and similarity metrics were evaluated to account for mixed data types within the EHR, e.g. binary, continuous, ordinal, and missing variables. We demonstrate the utility of the given algorithms by using Chronic Obstructive Pulmonary Disease (COPD), a condition with known clinical heterogeneity, for our proof of concept. Approximately 49,000 patients from the Geisinger Health System MyCode® Community Health Initiative with a known COPD diagnosis were subset using spirometry data, clinical lab variables, presence/absence of co-morbidities, clinical diagnosis codes, metabolic panels, medication usage data, vitals readings, and exercise testing measures. Using COPD as our proof of concept disease model, we demonstrate that data-driven approaches can be used as a means to overcome clinical heterogeneity of disease by embracing the complexity and wealth of data within the EHR. Moreover, these methods have the potential to build more accurate predictive disease models that can aid in uncovering the genetic basis of complex traits.

1888W

A framework for the construction of information maps. *H.S. Ooi[1,2,3], J. Grove[1,2,3,4], M. Schierup[4], A. Borglum[1,2,3], M. Mattheisen[1,2,3].* 1) Department of Biomedicine, Aarhus Univeristy, Aarhus, Select a Country; 2) iPSYCH, The Lundbeck Initiative for Integrative Psychiatric Research; 3) iSEQ, Centre for Integrative Sequencing; 4) BiRC, Bioinformatics Research Centre, Aarhus University, Denmark.

Molecular networks are undoubtedly an important resource for large-scale genetics studies. However, most of the existing methods only present the users with the network, without any auxiliary information to derive context-specific sub-networks for data analysis. Here, we present a flexible framework for the construction of an information map for large-scale genetics analysis. The information map is derived from more than 60 network-related resources and cover a wide-range of data types, including protein-protein interactions, gene co-expression, tissue expression, subcellular localization, drug targets, co-participation in pathways and co-association in diseases. Our framework first integrates and evaluates a molecular interaction network using a Bayesian-based approach, which transforms the heterogeneous information into an easily interpretable model. Unlike many of the existing approaches which ignore importat aspects of the raw information, we transform the data during the integration process into pair-wise relationships that indicate co-associations between genes or proteins under a particular conditions. This auxiliary information is then combined with the scored network to form the final information map. The base network in our information map is derived from multiple experiments spanning across various conditions and tissues. However, the biological processes are likely context dependent, where biological molecules interact with each other in a time- and space-specific manner. Here we illustrate the ability of our approach to derive a brain-specific network directly from the information map, and show the benefits of using such a brain-specific network in analyzing psychiatric disorders. The disease co-association information available in the information map allowed us to perform cross-disorder analyses of five major psychiatric disorders. This comparative analysis identified biological processes that are involved in multiple disorders or in contrast play a central role in one specific disorder. We could further demonstrate the flexibility of our information map by including disease-associated loci identified by our Danish iPSYCH study by means of genome-wide association studies or whole-exome sequencing experiments.

1889T

Predicting gene regulation in diverse global populations. *A. Badalamenti, N. Jachym, A. Almatrafi, V. Saulnier, J. Ng, S. Shah, H. Wheeler.* Department of Biology and Computer Science, Loyola University Chicago, Chicago, IL.

How the genetic architectures of complex traits vary between populations is not well understood, in part due to the historical paucity of GWAS in populations of non-European ancestry. For many complex traits, gene regulation is likely to play a crucial mechanistic role given the consistent enrichment of expression quantitative trait loci among trait-associated variants. We have previously developed PrediXcan, a gene-based association method that tests the mediating effects of gene expression levels by quantifying association between genetically regulated expression levels and the phenotypic trait of interest. The goal of our research is to build and test genetic predictors of gene expression within and between global populations. Utilizing genetic and transcriptomic data from diverse populations, we have built statistical models that accurately predict gene expression using variant genotypes. Our genotype data comes from the third phase of the International HapMap project, containing approximately 1.1-1.3 million SNPs per population. We used gene expression levels measured in lymphoblastoid cell lines from a subset of these individuals from seven non-European populations [Han Chinese (CHB), Gujarati Indians (GIH), Japanese (JPT), Luhya in Kenya (LWK), Mexican (MEX), Maasai in Kenya (MKK), and Yoruba in Nigeria (YRI)]. We performed elastic net modeling with 10-fold cross-validation to select genotypes and their weights that best predict expression of each gene. As expected, predictive performance (measured by R2) across genes correlated across populations. The best predicted gene in each population was the same, HLA-DRB5, with R2 > 0.94 in each population. However, reflective of differing allele frequencies and effect sizes, the highest predictive performance correlation was found between populations from the same continent. For example, the Pearson correlation between CHB and JPT is 0.59 and between LWK and MKK is 0.51, whereas the correlation between LWK and MEX is only 0.29. We will impute genotypes in each population to potentially improve predictive power. We aim to see how these genetic predictors of expression vary between populations, as well as determine if differences in predicted expression of the same genes associate with a particular complex trait between these populations by applying our PrediXcan method. Our method advances biological knowledge of the underlying regulatory mechanisms of disease risk not assessed in GWAS alone.

1890F

SCONE: Correcting and evaluating the influence of unwanted variation on single-cell RNA-seq data. *M. Cole[1,2], D. Risso[3], A. Wagner[2,4], J. Ngai[5], E. Purdom[6], S. Dudoit[3,6], N. Yosef[2,4].* 1) Department of Physics, University of California, Berkeley, Berkeley, CA; 2) Center for Computational Biology, University of California, Berkeley, Berkeley, CA; 3) Division of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, CA; 4) Department of Electrical Engineering & Computer Science, University of California, Berkeley, Berkeley, CA; 5) Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, CA; 6) Department of Statistics, University of California, Berkeley, Berkeley, CA.

  Single-cell RNA sequencing (scRNA-Seq) technologies are opening the way for transcriptome-wide profiling across diverse and complex mammalian tissues, facilitating unbiased identification of novel cell sub-populations and their functional roles. As in other high-throughput assays, a fraction of the heterogeneity observed in scRNA-Seq data results from batch effects and other technical artifacts. In particular, single-cell protocols' reliance on miniscule amounts of starting mRNA can lead to widespread "drop-out effects," in which expressed transcripts are missed. Due to the biases inherent to this assay, data normalization is an essential step prior to any downstream analyses. However, relatively few studies have assessed the performance of common normalization methods employed by the growing scRNA-Seq community. We introduce SCONE (Single-Cell Overview of Normalized Expression), a new R package for single-cell expression data quality control (QC) and normalization; our data-driven framework uses summaries of expression data to assess the efficacy of normalization workflows. Starting from read count data, our template normalization workflow begins with an optional drop-out imputation step, followed by a choice of scaling normalization (e.g. DESeq scaling factor) and ending with a regression-based adjustment for factors of unwanted variation (e.g. RUVg normalization). We compare the performance of hundreds of normalized data matrices by dimensionality reduction, scoring clustering metrics over cell classifications (prior or *de novo*) and scoring correlation metrics over control gene sets (e.g. cell-cycle genes or "housekeeping" genes). We apply SCONE to a publicly available data set of 130 high- and low-coverage single-cell libraries obtained from 65 human neural cell lysates (Pollen et al. 2014). Our results highlight negative trade-offs between preservation of biological variance and removal of artifact and demonstrate the extent of confounding between experimental batch, biological variation, library quality, and expression estimates. Top-ranked normalization methods for low-coverage sequencing data recover the covariance structures present in matched high-coverage data, showing that SCONE properly guides the study-specific choice of normalization work-flow based on empirical evidence. We will emphasize the issues surrounding the design of scRNA-Seq experiments and highlight the utility of positive controls and QC metrics in combating unwanted variation.

1891W

SeqSQC: An R package for sample quality check with NGS data. *Q. Liu[1,2], Q. Hu[2], Q. Zhu[2].* 1) Department of Biostatistics, University at Buffalo, SUNY, Buffalo, NY; 2) Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, NY.

  Objective: To identify problematic samples in next-generation sequencing (NGS) data, including samples with gender mismatch, abnormal inbreeding coefficient, cryptic relatedness, and population outlier. **Method:** SeqSQC is an R package that automates the sample quality control (QC) of large-scale NGS data. 100 independent samples and 10 family-related samples were selected from 1000 Genomes Project as benchmark to facilitate sample QC of study cohort. SeqSQC takes VCF files from whole-genome sequencing or exome sequencing data as input and stores the genotypes in GDS format to accommodate the extremely large-scale dataset with limited memory. The package generates interactive plots in each QC step as an intuitive interface for visualization. Statistical methods are implemented in the package for identifying problematic samples. **Result:** SeqSQC is capable of dealing with sequencing data from different platforms, different capture regions, and in different sample sizes. With simulation dataset and real data selected from 1000 Genomes Project, we have successfully identified all problematic samples. **Conclusion:** SeqSQC is highly efficient and flexible tool, with interactive visualization functions and statistical classification methods for sample quality investigation.

1892T

Cross-Validated BLUPs: A novel and powerful summary statistic opens new doors to multi-phenotype analyses. *J.A. Mefford[1], N. Zaitlen[2], J. Witte[1].* 1) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA; 2) Department of Medicine Lung Biology Center, UCSF San Francisco, CA.

With The Precision Medicine Initiative are emerging mega-scale genetic cohorts with rich phenotypic data collected over hundreds of thousands of individuals. Current gold standard analyses ignore the high-dimensionality of phenotypic data, and even multivariate LMMs are computationally incapable of leveraging the shared genetics between tens, let alone hundreds of phenotypes. Here we overcome this problem by introducing a novel summary statistic that is efficiently computed and securely shared without providing access to individual level genotypes. We propose to produce and release, in addition to SNP-level summary statistics, a genetic prediction of the phenotype of each individual in the study. To prevent over-fitting and provide securely shareable data, we adapt a machine learning technology to produce out-of-sample (i.e. cross-validated) predictions, which we call "cvBLUPs". Similar to SNP-level summary statistics, subject-level summary statistics open a range of new analysis possibilities. We highlight two of these here in the form of (1) the most efficient computation of genetic correlation produced to date and (2) efficient multivariate LMM association studies with up to hundreds of phenotypes analyzed jointly. We show analytically that the correlation of a pair of vectors of cvBLUPs is an efficient estimator of the genetic correlation of those phenotypes and accord well with genetic correlation estimates made using bivariate REML. Because calculation of genetic correlations by cvBLUP is essentially free after univariate analyses have been run we easily computed genetic correlations for over 300 phenotypes analysis of 4367 unrelated Finnish men. We discuss our findings including our discovery of a genetic correlation between BMI and smoking of 0.81. Unlike other summary statistics based methods cvBLUPs can be used to compute conditional or partial genetic correlations helping to unravel underlying causal relationships as opposed to just pleiotropy. We also use cvBLUPs to efficiently model polygenic effects across multiple traits with sets of cvBLUPs as fixed effects in a standard linear regression. This boosts the power of association analyses beyond what is possible with existing methods. For example, in an association analysis of BMI in the Finnish study, adjustment by the cvBLUPs for total cholesterol and smoking status each boost chi squared test statistics a further 2 percent.

1893F

Identifying multi-tissue gene expression outliers to elucidate the functional impact of rare variants. *E.K. Tsang[1,5], J.R. Davis[2,5], A. Battle[3,6], S.B. Montgomery[2,4,6].* 1) Biomedical Informatics Program, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 4) Department of Pathology, Stanford University, Stanford, CA; 5) co-first authors; 6) co-senior authors.

Each human genome harbors more than 40,000 rare variants and over 95% are non-coding. We expect that a subset of these variants disrupt physiological processes and significantly impact cellular phenotypes such as gene expression. Typical methods for identifying common regulatory variants, like eQTL studies, lack the power to detect rare variants. However, we can investigate rare variants with large effects on gene expression through outliers—cases where an individual strongly over-expresses or under-expresses a given gene compared with the rest of the population. Our group and others have successfully used this approach in single tissues, but the impact of rare variants across tissues remains uncharacterized. Studying multiple tissues lends insight into their regulatory similarities and combining information across tissues may yield a larger, more confident set of outliers. Here, we have developed a method for identifying multi-tissue gene expression outliers that captures situations where each individual has a different subset of sampled tissues. One of our method's key features is that it directly models similarities between tissues. In contrast to methods that ignore the expression correlation between tissues, we can discover expression outliers that break the population correlation structure in addition to outliers with consistent effects across tissues. We evaluated our method on simulated data and applied it to expression data from 23 primary human tissues from approximately 330 individuals in the GTEx data set. We find that expression outliers discovered in the GTEx data are more strongly enriched for nearby rare variants than for common ones. Rare variants near outlier genes are more conserved and more often in known regulatory regions, such as promoters and splice sites, compared with variants near the same genes in non-outlier individuals. Genes with outlier expression are enriched for allele-specific expression across tissues, providing evidence that these rare variants act in *cis*. We have applied our method to multi-tissue expression data and show how it can be extended to discover outliers in multi-omics data sets. Our findings demonstrate that our method identifies multi-tissue expression outliers that help pinpoint rare variants with regulatory effects. Improved characterization of the rare variants categories with large expression effects is essential to integrating whole genome data in our understanding of health and disease.

1894W

Facilitating collaborative analysis in large-scale rare disease genomics. *E. Valkanas[1,2], B. Weisburd[1,2], L. Xu[1,2], H. Arachchi[1,2], M. Lek[1,2], T. Mullen[1,2], A. Topf[3], M. Bertoli[3], K. Johnson[3], V. Straub[3], M.J. Daly[1,2], D. MacArthur[1,2].* 1) The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) Newcastle University John Walton Muscular Dystrophy Research Centre and MRC Centre for Neuromuscular Diseases, Institute of Genetic Medicine, Newcastle upon Tyne, UK.

One of the critical challenges in rare disease diagnosis is ensuring that data can be effectively shared between informatics groups and clinical experts. We have developed an open source web-based tool, *seqr* (https://seqr.broadinstitute.org/), that provides a centralized platform for analyzing exome and whole genome variant call sets, capturing phenotype information, and communicating progress and results. *seqr* allows users to leverage a variety of different data sources when searching for causative variants in patient and family data. In addition to allowing easy searching of patient genotype information, *seqr* also give access to reference populations, known disease associations, CNV data, and allows users to interactively view NGS read support and sequence context for variants. The platform allows the collection of structured phenotype data (Human Phenotype Ontology) through integration with PhenoTips. For unsolved cases, users have the option to search for individuals with similar phenotype and genotype profiles through the Matchmaker Exchange. Finally, the system facilitates collaboration between groups by allowing joint analyses of the same data sets, and variant-level tagging and discussion of pathogenicity by multiple users. The *seqr* platform has been applied to rare disease projects spanning 9363 exomes and 1088 whole genomes. We demonstrate the utility and usability of this platform using data from the MYOSEQ project, a collaboration between the Newcastle University John Walton Muscular Dystrophy Research Centre and the Broad Institute that has sequenced 1,000 exomes from patients with undiagnosed limb-girdle muscle weakness. Using the *seqr* platform, our analysis team, distributed across two continents, has successfully diagnosed 281 of the 725 families analyzed to date, including 14 index cases where diagnosis has altered therapeutic outcomes. A critical component of such successes is connecting with collaborators to develop new features within the platform and increase usability. As our user community grows, we plan to continue this iterative process to facilitate rare disease research.

1895T

PheWeb: A tool for interacting with and visualizing PheWAS results. *P. VandeHaar[1], C. Clark[1], E. Schmidt[1], M. Flickinger[1], L. Fritsche[2], C. Brummett[3], S. Kheterpal[3], G. Abecasis[1].* 1) Biostatistics, University of Michigan, Ann Arbor, MI.Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI.

Phenome-Wide Association Studies (PheWAS) report the association between single genetic variants and many traits. Each trait is typically a disease state defined by a collection of ICD-9 (International Classification of Disease, 9th Edition) or ICD-10 codes. When PheWAS are combined with modern genotyping and sequencing technologies, they can result in billions of association statistics and hundreds of association signals. Translating these large result sets into new understanding of health and disease requires making the results accessible to many scientists. Here, we describe a tool that can make these very large result sets easily accessible and intuitively browseable. PheWeb enables interactive exploration of the associations between hundreds or thousands of traits and genetic variants. It can easily handle matrices summarizing association between millions of SNPs and thousands of traits, such as would be produced by running a GWAS on a modern electronic health record dataset.For each trait, the browser currently includes a Manhattan plot, a QQ plot, and a LocusZoom-like region view. For each SNP, the browser includes a PheWAS plot and a table of the most-associated traits. These plots are based on LocusZoom.js, allowing them to link to each other, zoom, and show tooltips, providing an exploratory interface.An instance of PheWeb is available at http://pheweb.sph.umich.edu to browse EHR-based association results from the Michigan Genomics Initiative. If you have association results for many phenotypes, we encourage you to use PheWeb to visualize your results to improve understanding and interpretation of signals.

1896F

Generic, protein-, and mechanism-specific computational tools for variant interpretation. *M. Vihinen, A. Niroula, S. Urolagin, J. Väliaho, H. Ali.* Lund University, Sweden.

   Next Generation Sequencing (NGS) technologies produce huge amounts of sequence information. Large genomic projects aiming at personalized medicine are underway; however, variation interpretation has been and still is a bottleneck. Due to the huge amounts of produced data, reliable and fast computational tools are needed to enhance variation interpretation and to cope with the datasets. We have developed several tools to predict the impact of variations which can be broadly grouped as generic tools to predict the tolerance of variants, locus-specific tools to predict the tolerance of variants in specific proteins, genes or domains, and mechanism-specific tools to predict the mechanism of variation effect. The generic and locus-specific tools rank and prioritize variants and the mechanism-specific tools help to interpret the mechanism of variation effect. PON-P2 is a generic tool for predicting variant pathogenicity with highest performance among related tools[1]. The method computes reliability for each variation and classifies the reliably predicted variants as pathogenic or neutral and the unreliably predicted variants as unclassified. PON-P2 is fast and it can be integrated into variation interpretation pipelines since there is a application programming interface API) available. Our locus-specific tools include PON-BTK for variants in the kinase domain of Bruton tyrosine kinase[2], PON-MMR2 for mismatch repair protein variants[3] and PON-mt-tRNA for mitochondrial tRNA variants[4]. Mechanism-specific tools include PPSC for protein stability affecting variants[5], PON-Diso for protein disorder affecting variants[6], and PON-Sol for protein solubility affecting variants[7]. All these tools are among the most reliable ones. They are freely available at http://structure.bmc.lu.se. References: 1 Niroula, A. *et al. PLoS One* **10**, e0117380, (2015). 2 Väliaho, J. *et al. Hum Mutat* **36**, 638-647, (2015). 3 Niroula, A. & Vihinen, M. *Hum Mutat* **36**, 1128-1134, (2015). 4 Niroula, A. & Vihinen, M. *Nucleic Acids Res*, (2016). 5 Yang, Y. *et al. Amino Acids* **44**, 847-855, (2013). 6 Ali, H. *et al. Hum Mutat* **35**, 794-804, (2014). 7 Yang, Y. *et al. Bioinformatics*, (2016).

1897W

Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *B. Wang[1], J. Zhu[2], E. Pierson[1], S. Batzoglou[1].* 1) Computer Science, Stanford University, Stanford; 2) Department of Electrical Engineering, Stanford University.

   Single-cell RNA-seq (scRNA-seq) technologies enable gene expression measurement of individual cells and allow the discovery of cell population heterogeneity. Recent advances in scRNA-seq have allowed simultaneous profiling of thousands of cells and increased the sensitivity in quantifying single-cell transcriptome landscapes of complex biological systems. However, scRNA-seq data sets are noisy with high degree of dropouts, and exhibit higher levels of diversity such as cell input heterogeneity and variation in cell cycle stages. These challenges make it difficult to define cell-to-cell similarity measures based on strict statistical assumptions that have been developed for bulk RNA-seq. Here, we propose a novel similarity-learning framework, SIMLR (single-cell interpretation via multi-kernel learning), which learns an appropriate distance metric from the data for dimension reduction, clustering and visualization. We profiled >50,000 peripheral blood mononuclear cells (PBMCs) with the ChromiumTM system from 10x Genomics, and used SIMLR to provide an unbiased classification of all major subpopulations at expected proportions. In order to evaluate the sensitivity and accuracy of SIMLR, we further analyzed individual purified populations from PBMCs and pooled the data in silico at varying proportions. In addition, we extensively evaluated SIMLR using published scRNA-seq datasets generated by several other micromanipulation and microfluidics platforms with varying sequencing depths. For all the datasets above, we first show that simple correlation-based or distance-based similarity measures are sensitive to noise, dropouts and outlier effects among the high dimensional data. In contrast, SIMLR can uncover clear similarity block structures by automatically learning appropriate cell-to-cell similarity specific to each dataset. Furthermore, we performed dimension reduction on these high dimensional datasets using the similarity learned by SIMLR, and compared it with 8 other popular alternative methods. We evaluated the effectiveness of SIMLR's dimension reduction both quantitatively and qualitatively by considering clustering accuracy and visualization. Dimension reduction performed via SIMLR yields a substantially higher clustering accuracy. When applied to visualization, we illustrate SIMLR's advantages over other methods in projecting the high dimensional data to 2-D where the different cell types are automatically projected in spatially distinct clusters. .

1898T

Scaling rare disease genomics to tens of thousands of samples. *B. Weisburd, E. Valkanas, H. Arachchi, M. Daly, M. Lek, D.G. MacArthur.* Broad Institute, Cambridge, MA.

Rapid advances in exome and whole-genome sequencing make it possible to identify the full spectrum of genetic variation in rare disease patients and family members. However, to make sense of this data, and discover the causal variants underlying rare diseases, powerful filtering and decision support tools are needed that can handle the very large data sets now being produced. We are addressing this challenge by developing seqr - an open source, interactive web-based tool that allows researchers to work together to search and annotate variant call-sets from exome and whole genome studies involving families of rare disease patients. This presentation will describe the technology stack underlying seqr, including the database, server-side and client side technologies involved, with a focus on the design choices made to facilitate the scaling of the system to tens of thousands of exome and genome samples. The need to allow interactive searching of terabytes of genomic sequencing data requires custom solutions and thoughtful design trade-offs. We will present the technical details of how seqr dealt with these challenges. Additionally this talk will describe how interactive visualization of read-level data, copy number variation, detailed phenotype data, and other relevant data types were brought together in a single user-friendly interface.

1899F

Pedigree reconstruction in the era of many thousands of samples. *A.L. Williams[1], R. O'Hern[1], J. Blangero[2].* 1) Biological Statistics & Computational Biology, Cornell University, Ithaca, NY; 2) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.

Members of sexually reproducing species are related to each other through a pedigree in which each individual has two parents, some number of offspring, and is related to all other individuals through these first degree relationships. While numerous genetic analyses require or benefit from knowledge of pedigree relationships, this information is not always available. Moreover, when a pedigree is reported, errors in it can confound downstream analyses so the structure must be confirmed. We present Attila, a novel method for pedigree reconstruction that is accurate, efficient, and scalable to thousands of samples. The algorithm works by first estimating the probability that each pair of samples is related via three possible relationship classes: parent-offspring, full sibling, or other (2nd degree or more distant). Given these probabilities, the program constructs a pedigree as a graph with nodes representing samples and edges labeled by the relationship between each sample pair. Attila inserts edges in order of the most likely relationship, and during this process, propagates information implied by the edge using biologically-derived constraints on the relationships between any three samples. If an inserted edge leads to a constraint violation, the program reverts to the previous graph and inserts the next most likely edge until all relationships are defined. Thus the algorithm uses backtracking search to efficiently reconstruct a pedigree with high likelihood. Many pedigree reconstruction methods scale exponentially in the number of samples, whereas Attila has polynomial runtime and memory scaling, with a theoretical runtime scaling limit of $O(N^3)$, where N is the number of samples. In practice, a set of samples can be divided into subsets based on simple measures of relatedness and each of these subsets individually analyzed for pedigree structure, thus reducing runtime significantly. We ran Attila and PRIMUS on samples from a Mexican American pedigree dataset containing 257 individuals and consisting of six separate pedigrees. PRIMUS only reconstructed the smallest (21 sample) pedigree among those examined, but did not detect others and reported no errors. By contrast, Attila reconstructed all six pedigrees in only 6.6 hours. Of the >32,000 pairwise relationships in the reported data, all but four (likely misreported) were classified correctly. We are further improving Attila's efficiency and will present results from analyzing >115,000 UK Biobank samples.

1900W

Population-scale SV detection and characterization using SVTools. *H.J. Abel[1,2], D.E. Larson[1,2], C. Chiang[1,3], A. Badve[3], L. Ganel[1,3], R.M. Layer[4], A.R. Quinlan[4], I.M. Hall[1,3].* 1) McDonnell Genome Institute, Washington Univ. School of Medicine, St. Louis, MO; 2) Dept. of Genetics, Washington Univ. School of Medicine, St. Louis, MO; 3) Dept. of Medicine, Washington Univ. School of Medicine, St. Louis, MO; 4) Dept. of Human Genetics, Univ of Utah, Salt Lake City, UT.

   Structural variation (SV), including copy number variation and balanced rearrangements, comprises a large part of human genetic diversity. The contribution of SV to human disease remains unknown, however, as the full range of SV cannot be detected from the SNP microarray or exome sequencing data collected in most large-scale genetic studies to date. The current wave of population-scale whole-genome sequencing (WGS) studies (eg, CCDG, Topmed) provides unique opportunities, both for the characterization of the SV landscape in the healthy population and for investigation of the role of SV in common disease. To succeed, such endeavors will require new approaches to SV detection, as current algorithms scale to at most several hundred genomes.   We have developed svtools (https://github.com/hall-lab/svtools), a suite of python tools for SV detection and characterization, scalable to tens of thousands of genomes. The svtools pipeline is loosely modeled on the GATK 'N+1' framework: single-sample SV callsets are generated in parallel using LUMPY, whose output describes the uncertainty in breakpoint position via probability distributions. SV calls are then merged across samples, with shared evidence in breakpoint distributions combined to provide improved precision. Each sample is then genotyped and annotated with read-depth information, in parallel, at all variants in this discovery set; the resulting multi-sample VCF describes cohort-level SV with greater sensitivity and precision than the original single-sample calls.   The SVTools suite enables joint SV discovery and genotyping on population cohorts, thus promising a wealth of new genetic variation to aid in the understanding of human disease. As proof of principle, we present the results of our initial SV callset comprising more than 7000 deep WGS datasets from individuals of diverse ancestry. We observe an abundance of rare, high impact SVs that are predicted to alter gene function, including a remarkable number of balanced and/or complex rearrangements that would not be detected by traditional CNV-centric methods such as microarrays or read-depth analysis. Finally, we discuss our plans for data-sharing and the creation of an SV catalog resource to improve the understanding of the landscape of genetic variation in the healthy population. Such a resource should prove invaluable, in particular for the evaluation of 'n-of-1' type variants often encountered in clinical settings and in the study of rare disease.

1901T

A population-level structural variant catalogue to improve the recall and consistency of common structural variant calls. *M.A. Bekritsky[1], P. Krusche[1], N. Johnson[2], A. Gross[2], S. Ivakhno[1], S.J. Humphray[1], D.R. Bentley[1], M.A. Eberle[2].* 1) Illumina Cambridge Ltd., Saffron Walden, Essex, Essex, United Kingdom; 2) Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA.

   Accurate and consistent copy number variant (CNV) and structural variant (SV) detection is an essential component of genomic research and necessary for maximizing the diagnostic yield of clinical sequencing. Yet, the quality of SV calling cannot be assessed accurately in part because there are no comprehensive benchmarking datasets for large variants. While there are high quality benchmarking databases for SNVs and indels, SV benchmarking resources are limited to either a few samples sequenced to high depth or many samples sequenced to low depth. Additionally, these resources are heavily biased towards deletions over duplications and other SVs. To improve the current benchmarking resources, we are developing a database of common SVs identified from whole-genome sequence data from over 4,000 samples sequenced to an average depth of 35x. We have developed algorithms that scan through aggregated SNV datasets to identify population signals consistent with common underlying deletions and duplications. Using this approach we identified ~13,000 strong signals based on deviations from Hardy-Weinberg equilibrium and depth correlations that were consistent with deletions and duplications. A primary differentiator of this method is that it uses population signals to identify CNVs, in contrast to traditional CNV callers, which use read depth and/ or split reads. Using the population signals as hypotheses, we have identified a high-quality subset of CNVs whose depths are strongly correlated to the genotypes of a nearby SNV. We have also de novo assembled an initial subset of over 6,000 deletions (mean size ~500bp) to breakpoint resolution and are developing variant callers that leverage population information by using graph-based methods. This approach will allow these deletions to be reliably and consistently called in future studies. For variants that are not amenable to graph-based detection methods we are using population-level depth signals so that these variants can also be called consistently and reliably. To provide a public resource underpinning this variant catalogue for further benchmarking and algorithm improvement, we are sequencing 150 Coriell samples from disparate ethnicities to high depth using 2x150bp reads generated on a HiSeqX and releasing the sequence data and variant calls on these samples. We will present the results of our analysis and ongoing additional improvements to increase the completeness of this SV resource.

1902F

SVScore: An impact prediction tool for structural variation. *L. Ganel[1,2], I.M. Hall[1,2,3].* 1) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA; 2) Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 3) Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA.

Structural variation is an important source of human genome variation that includes deletions, duplications, inversions, mobile element insertions, translocations, and complex rearrangements. Over the past several years, much progress has been made in the area of structural variant (SV) detection, and we are now able to routinely detect 5,000-10,000 SVs in a typical deeply sequenced human genome. However, predicting the functional impact of SVs discovered in whole genome sequencing (WGS) studies remains extremely challenging. Accurate SV impact prediction is especially important for WGS-based rare variant association studies, such as those undertaken by the NHGRI Centers for Common Disease Genomics, and for WGS-based studies of rare disease. Here we present SVScore, a novel computational tool for *in silico* structural variant impact prediction. SVScore leverages existing single nucleotide variant (SNV) pathogenicity scores such as CADD (Kircher et al., 2014). SVScore aggregates per-base pathogenicity scores across relevant genomic intervals for each SV in a manner that considers variant type, gene features, and uncertainty in breakpoint location. To evaluate its effectiveness in a large dataset, we used allele frequency (AF) as a proxy for pathogenicity by assuming that, due to selection, pathogenic SVs are enriched among rare variants and benign SVs are enriched among common ones. We show that the AF spectrum of SVs that received high impact scores was skewed toward lower frequencies and vice versa, suggesting that high-scoring variants are indeed under negative selection. We quantify this effect and show that the allele frequency skew of high-scoring SVs in SVScore is stronger than the skew produced by various naïve approaches that merely consider variant size or genome annotations, suggesting that SVScore is more effective in identifying deleterious SVs than existing alternative methods. We further show that SVScore identifies pathogenic SVs in both coding and noncoding regions and compare the strength of the AF skew imposed on SVs by our scoring system to that of comparable SNV methods. Finally, we demonstrate that high-scoring duplications are under surprisingly strong selection relative to high-scoring deletions, suggesting that duplications can be more deleterious than previously thought. In conclusion, SVScore provides pathogenicity prediction for SVs that is both informative and meaningful for understanding their functional role in disease.

1903W

High resolution measurement of DUF1220 domain copy number from whole genome sequence read depth. *I.E. Heft, D.P Astling, K.L. Jones, J.M. Sikela.* UC-Denver Anschutz Medical Campus, Aurora, CO.

DUF1220 protein domains show the most extreme human lineage-specific increase in copy number of any coding region in the genome. There are 302 haploid copies of DUF1220 in hg38 (~160 of which are human-specific) and the majority of these can be divided into 6 different subtypes (clades). Copy number changes of specific DUF1220 clades have been associated in a dose-dependent manner with brain size variation (both evolutionarily and within the human population), cognitive aptitude, and autism and schizophrenia severity. However, DUF1220 sequences have been difficult to measure and, as a result, have typically not been directly examined in searches for human disease genes. To address these challenges and maximize the utility of emerging whole genome sequence (WGS) datasets, we have developed a high resolution method for the measurement of DUF1220 copy number from WGS data. We have characterized the effect that various sequencing and alignment parameters have on the accuracy and precision of the method and defined the parameters that lead to optimal DUF1220 copy number measurement and resolution. We show that copy number estimates obtained using our read depth approach are highly correlated with those generated by ddPCR for three representative DUF1220 clades. By simulation, we demonstrate that our method provides sufficient resolution to analyze DUF1220 copy number variation at three levels: (1) DUF1220 clade copy number within individual genes & groups of genes (gene-specific clade groups) (2) genome wide DUF1220 clade copies and (3) gene copy number for DUF1220-encoding genes. This is the first approach that measures DUF1220 copy number with clade-specific resolution and, to our knowledge, represents the most detailed method for the copy number measurement of DUF1220-related sequences. This approach greatly enhances our ability to analyze the role that these sequences play in human variation and disease.

1904T

CoNVaDING: Single exon variation detection in targeted NGS data. *L.F. Johansson[1,2], F. van Dijk[1,2], E.N. de Boer[1], K.K. van Dijk-Bos[1], J.D.H. Jongbloed[1], A.H. van der Hout[1], R.C. Niessen[1], H. Westers[1], R.J. Sinke[1], M.A. Swertz[1,2], R.H. Sijmons[1], B. Sikkema-Raddatz[1].* 1) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 2) University of Groningen, University Medical Center Groningen, Genomics Coordination Center, Groningen, The Netherlands.

   1. Introduction We have developed a tool for detecting single exon copy number variations (CNVs) in targeted next-generation sequencing data: CoNVaDING (Copy Number Variation Detection In Next-generation sequencing Gene panels). CoNVaDING includes a stringent quality control metric, that excludes or flags low quality exons, making the tool useful in both research and clinical diagnostics.   2. Materials & Methods The method is based upon a read depth comparison of selected target regions between a sample of interest and a set of control samples. Existing tools, such as XHMM, CoNIFER and CODEX consider all control samples equally informative even though there are sample to sample variations caused by differences in PCR and capturing efficiency. CoNVaDING, however, selects the control samples showing a coverage pattern most similar to that of the sample analysed. Data is then normalized in two different ways, using within the sample all autosomal targets or all targets within the same gene. Based on the normalized data, for each target the ratio of the normalized average read depth of the sample to that of the controls and a distribution analysis using a Z-score are calculated and a prediction is made for each target to determine whether a CNV is present or not. CoNVaDING provides for each sample quality metrics to distinguish high quality samples and targets from low quality ones.   3. Results We compared the performance of CoNVaDING with XHMM, CoNIFER and CODEX in 320 samples captured with two different targeted gene-panels (Agilent Sure Select custom design 0421101 and 0679001) containing in total 376,440 exons. For all CNV calls made by one of the first three methods MLPA was performed. CoNVaDING detected all known CNVs in high-quality targets, giving 100% sensitivity, at a 99.998% specificity for 308,574 high-quality exons.   4. Discussion CoNVaDING outperforms XHMM, CoNIFER and CODEX by exhibiting a higher sensitivity and specificity and by precisely identifying low-quality samples and regions, which have a high risk of false positive or false negative results. Since the quality control metrics show exactly which exons can be reliably analyzed and which exons are in need of an alternative analysis method, CoNVaDING can not only be applied for CNV detection in a research setting, but also in clinical diagnostics. CoNVaDING software is available under the GNU GPL open source license and can be freely downloaded from https://github.com/molgenis/CoNVaDING.

1905F

ClinVar: For researchers and medical practitioners alike. *M.J. Landrum, J.M. Lee, M. Benson, G. Brown, C. Chen, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Katz, Z. Maddipatla, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin, W.S. Rubinstein, D.R. Maglott.* National Center for Biotechnology Information, Bethesda, MD.

   ClinVar is a public archive of interpretations of clinical significance and functional effects of variants in any gene and for any disease. Interpretations are submitted by many sources, including research laboratories, clinical testing laboratories, locus-specific databases, expert panels, practice guidelines, as well as OMIM® and GeneReviews™. Currently ClinVar holds more than 9700 variant interpretations submitted as research results from 196 organizations. A submission may be held private until the corresponding paper is published; submitters are encouraged to start their submission to ClinVar when they start the publication process, to allow time for NCBI staff to process the data. ClinVar accessions for submissions can be provided early so that the manuscript can reference the ClinVar record. The PubMed ID for the submitter's paper may be included with the original submission or added later in an update. Submissions of a single variant interpretation are welcome, and an online submission wizard guides you through the process. A research submission may report the clinical significance of the variant, if patients were evaluated as part of clinical research. Alternatively, the functional significance of the variant may be provided if the submitter describes experimental results, such as a functional assay or an animal model. Submitters are encouraged to send updates to ClinVar when further research indicates that their classification should be changed or provides additional evidence. ClinVar can be used to inform research as well. The full dataset is available as XML for download; summary data is available as a VCF file and tab-delimited summary files. ClinVar data may be accessed programmatically with E-utils, NCBI's API for data retrieval. The data may also be viewed on the ClinVar website. The database may be searched by variant, gene symbol, disease or phenotype, PubMed ID, submitter, and more. Search results may be filtered in several ways, including allele origin, type of submission, variant type, functional consequence, and clinical significance. Each search result links to a variation page for the variant or set of variants that were interpreted. The variation page includes the aggregate clinical significance and review status based on all submissions. Details from each submission are also displayed, including each submitter, each submitted interpretation, criteria used to evaluate the variant, citations, and evidence for the interpretation.

1906W

A robust statistical approach to refine frequency thresholds for clinical variant interpretation. *N. Whiffin[1,2*], E. Minikel[3,4*], R. Walsh[1,5], A. O'Donnell-Luria[3,4], K. Karczewski[3,4], A.Y. Ing[6], B. Funke[6,7], S.A. Cook[1,2,8#], D. MacArthur[3,4#], J.S. Ware[1,2,4#].* 1) National Heart and Lung Institute, Imperial College London, London, UK; 2) MRC Clinical Sciences Centre, Imperial College London, London, UK; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 5) NIHR Royal Brompton Cardiovascular BRU, London, UK; 6) Partners HealthCare Personalized Medicine, Laboratory for Molecular Medicine, Cambridge, MA; 7) Department of Pathology, Harvard Medical School/Massachusetts General Hospital, Boston, MA; 8) National Heart Centre Singapore, Singapore.

Distinguishing pathogenic variants from benign bystanders is a principal challenge in clinical genetics. Allele frequency (AF) is a key potential discriminator, as variants found in the general population can often be excluded as causes of penetrant Mendelian disease. The Exome Aggregation Consortium (ExAC) dataset provides unprecedented power to characterise rare variants, but it is often unclear whether a variant present at very low frequency should be classified as benign, given that Mendelian diseases may also be present in the population. Consequently, there is large variation in practice, and a conservatively high threshold is often used, leading to inconclusive variant interpretation. We have developed a rigorous framework to evaluate whether an AF observed in a population is compatible with pathogenicity. This permits the safe and appropriate use of much more stringent AF thresholds. We consider disease prevalence, genetic/allelic heterogeneity and penetrance to calculate a maximum population frequency for a variant causing a given disease. We then assess the frequency of a variant in a fixed population sample, to determine whether it is compatible with disease causation. We apply this method to a range of clinically important inherited cardiac conditions, with ExAC as a population reference dataset. To evaluate our approach, we aggregated high quality reference variants from 7,855 cardiomyopathy cases classified by two clinical laboratories. Compared to a conventional 0.1% AF filter, our more stringent filter reduced the number of 'Variants of Uncertain Significance' by 10% (69 variants), while preserving 375/376 'Pathogenic' variants. The filtered 'pathogenic' variant causes a low penetrance phenocopy. To further validate our approach, we curated the 28 ClinVar 'pathogenic' variants that failed our stringent filter: none had robust evidence of disease-causation. In a clinical setting, stringent AF thresholds allow for much more confident variant interpretation: in patients with HCM the proportion of *MYBPC3* variants that are pathogenic increases from 78% at AF<0.1% to 98% at AF<0.001%. In a whole exome approach, the mean number of potential disease-causing variants for a dominant disease is reduced ~3-fold from 179 (AF<0.1%) to 67 (AF<0.001%) per exome. To facilitate variant interpretation, we have pre-computed our frequency annotations for all variants in the ExAC dataset, and are releasing these data through the ExAC website. *Contributed equally.

1907T

Cell type-specific annotation and prioritization of regulatory variants. *M. Li[1,2], Z. Pan[2,3], J. Wu[1], P. Sham[2,3,4], J. Kocher[5], M. Li[2,3], J. Liu[1], J. Wang[5].* 1) Department of Statistics, Harvard University, Cambridge, Boston, MA 02138-2901, USA; 2) Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 3) Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 4) State Key Laboratory in Cognitive and Brain Sciences, The University of Hong Kong, Hong Kong SAR, China; 5) Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, MN 55905, USA.

Interpreting human regulatory variants in the noncoding genomic region is critical to understand the regulatory mechanisms of disease pathogenesis and promote personalized medicine. With the accumulation of functional annotation data, several computational methods have been developed to predict and prioritize noncoding regulatory variants. However, we observed significant inconsistency in the predicted results among these tools, and the strengths and weakness of each tool are not known, which will hinder the selection of functional candidates in the follow-up experimental validation. Furthermore, the predictions made by these tools are rarely based on specific cell type conditions. But the function of regulatory variants, associating with concrete chromatin signatures and other cellular environments, is usually cell type-specific. Prediction based on cell type-specific function genomic/epigenomic data can improve the chance and accuracy of causal variants discovery. Prediction algorithms that use these information to prioritize and fine-map the putative regulatory variants are in urgent demand. In this study, we first compiled an integrative database for predictions from eight latest cutting-edge algorithms on function annotation/prediction of noncoding variants. We developed a composite strategy to integrate multiple predictions and compute the composite likelihood of a given variant being causal. We found that the composite model significantly improves the prediction performance, achieving significant AUC improvement. We further summarized a set of consolidated chromatin features using cell type-specific chromatin marks for eleven uniformly processed expression quantitative trait loci (eQTLs) datasets. We developed a logit model to measure the regulatory potential for a given variant in selected conditions. Using GWAS fine-mapped SNPs and 127 recently profiled human epigenomes, we have shown that our condition-dependent prioritization can identify the ensemble effect of chromatin signals that fine-maps the causal regulatory variant in phenotypically relevant cells, rather than the previous approach of using single chromatin marks such as H3K4me3 or H3K27ac. Finally, we combined the composite model and condition-dependent model into a unified probability to measure variant regulatory potential. We demonstrated that our unified model has better performance than state-of-the-art and cell type-specific algorithms in prioritizing regulatory variants. .

1908F

A novel 2D genome segmentation method for studying epigenetic variation in many cell types. *Y. Zhang[1], L. An[2], F. Yue[3], R.C. Hardison[4].* 1) Department of Statistics, Pennsylvania State Univ, University Park, PA; 2) Bioinformatics and Genomics Program, Huck Institutes of the Life Sciences, Penn State University, University Park, PA; 3) Department. of Biochemistry and Molecular Biology, Penn State School of Medicine, Hershey, PA; 4) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA.

A critical step for understanding the impact of genetic variation on disease is through identification of regulatory elements and understanding how they act on target genes. With thousands of whole-genome epigenetic data sets generated in hundreds of human and mouse cell types, a major challenge is to build advanced quantitative models to exploit this rich source of information for studying the roles of functional elements in disease in cell type specific context. Most existing methods are developed for studying one cell type at a time. Frequently these methods are adopted to study multiple cell types via overly simplistic strategies that are inadequate to capture the complex interrelationship between cell types. The very few methods for joint modeling of multi-cellular epigenomics cannot scale up to studying a large number of cell types. We have developed a novel and efficient method called IDEAS that can jointly analyze multiple epigenetic marks in hundreds of cell types simultaneously, and detect differential regulatory modules across both the genome and time (e.g., cell types, conditions, individuals, differentiation stages). The novelty of IDEAS lies in its integrative and dynamic combination of information from multiple cell types within a coherent probabilistic model. The inferred functional elements are therefore not only more accurate, but also coherent and comparable across cell types, which enables robust comparative analysis of differential gene regulation in a multi-cellular context. Application of IDEAS to ENCODE has demonstrated its superior power for predicting functional elements and their impacts on expression over existing algorithms. The cell type specific regions predicted by IDEAS are more significantly enriched in disease variants than those inferred from existing genome annotations. IDEAS has linear inference complexity with respect to the number of cell types, and we have used IDEAS to annotate the 127 epigenomes in Roadmap Epigenomics. A useful application of IDEAS is to leverage existing data sets in the public domain to study specific cell types of interest by individual labs, without requiring complete data generation. This will enable studying differential gene regulation in a wide range of cell types at low cost. Our improved annotation of functional elements in the genome and their cell type specificity will further help to identify and interpret non-coding variants in disease.

1909W

Incorporating variant analysis in microRNA coding genes and microRNA target sites into the standard exome analysis pipeline. *T.S. Frisby, C.C. Lau, C.J. Adams, M.J. Warburton, N.I. Balanda, B.N. Pusey, W.A. Gahl, D.R. Adams.* NHGRI-UDP, NIH, Bethesda, MD.

Exome sequencing has become a routine assessment for patients in the NIH Undiagnosed Diseases Program. To date, several hundred patients and their nuclear families have been sequenced. Our current exome analysis methods, which attempt to expand on standard-of-care clinical exome analysis, nonetheless focus largely on variants in and around the coding regions of the genome. In order to increase our success rate, as well as to elucidate novel causes of rare diseases, we are extending our analytic pipeline to incorporate an expanding set of non-coding DNA variation. One area of interest includes microRNA (miRNA) coding genes and miRNA target sites within the 3'-UTR of protein coding genes. miRNA are small, non-coding RNA that have been shown over the past decade to regulate gene expression by binding to mRNA, causing degradation and translational repression. Variants found in miRNA coding genes and target sites have also been shown to cause diseases in humans. As part of a pilot analysis, we have investigated variants within these elements by extending our exome analysis pipeline. In the 5 nuclear families that we studied, we found a range of 70 – 323 variants (mean 220, median 278) that are rare in the general population and segregated with the diseases within the families that were previously excluded from our exome analysis. By annotating the miRNA target sites using target binding prediction tools (miRanda-mirSVR, TargetScan), we were able to prioritize variants that fell into predicted or conserved miRNA target sites. We then further prioritized variants that fell within the 95th percentile of miRNA target predictions (defined as miRNA-target mRNA predicted interactions that have 50% probability of decreasing target mRNA level by 1 standard deviation, or 70% probability of decreasing target mRNA level by 0.5 standard deviation). Finally, we computed the effect of the variants to the binding prediction, as a way to prioritize variants that may be biologically and clinically significant. Further work needs to be done to validate these findings, but this represents the first step in the analysis and interpretation of variants that had been previously overlooked. This research is supported by the Intramural Research Program of the National Human Genome Research Institute and the Common Fund of the National Institutes of Health.

1910T

dbLoF: An online database of human loss-of-function variants. *P. Singh[1], D. Mohs[1,2], E. Minikel[1,2], I. Armean[1,2], K. Karczewski[1,2], S. Farhan[1,2], D.G. Macarthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Genetic variants predicted to cause loss-of-function (LoF) can be powerful models for the inactivation of human genes, providing insight into disease mechanisms, gene function, and the potential therapeutic relevance of possible drug targets. Here we describe dbLoF, a database designed to provide systematic information on the distribution and (inferred or observed) phenotypic impact of LoF variants in each human gene. Drawing from a collection of population sequence data, most heavily the Exome Aggregation Consortium (ExAC) data set, dbLoF provides the positional distribution of LoF variants for each gene, their aggregate frequency distribution across human populations, information about each variant, the inferred LoF constraint of each gene, and links to further variant- and gene-level resources. In addition, we are currently populating the database with curated gene-level summaries drawn from the primary disease literature as well as population databases of genotype/phenotype associations. We will describe the back and front-end architecture of dbLoF, the underlying resources used to create the database, and typical workflows for users of the resource.

1911F

An automated method to identify familial relationships in electronic health records. *F. Polubriaginof[1], K. Quinnies[1,2], R. Vanguri[1], A. Yahi[1], M. Simmerling[3,4], I. Ionita-Laza[5], H. Salmasian[3,1], S. Bakken[1,6], K. Kiryluk[7], D. Goldstein[2], D.K. Vawdrey[3,1,9], N.P. Tatonetti[1,2,7,8,9].* 1) Biomedical Informatics, Columbia University, New York, NY; 2) Institute for Genomic Medicine, Columbia University, New York, NY; 3) Value Institute, NewYork-Presbyterian Hospital, New York, NY; 4) Division of Medical Ethics, Weill-Cornell Medical Center, New York, NY; 5) Mailman School of Public Health, Columbia University, New York, NY; 6) School of Nursing, Columbia University, New York, NY; 7) Department of Medicine, Columbia University, New York, NY; 8) Department of Systems Biology, Columbia University, New York, NY; 9) Co-senior author.

Family history is a prognostic indicator of inherited disease and is critical to precision medicine. In general, heritability is defined as the variance of the trait attributable to genetic factors. Knowledge of patient's family history in conjunction with knowledge of disease heritability enables individualized disease diagnosis, treatment, and prevention. Electronic health records (EHRs) have become ubiquitous and contain vast amounts of phenotypic data. Combined with knowledge of family histories, these data could greatly impact genetic study design, recruitment, and analysis. Unfortunately, family history and familial relationship data are poorly collected in EHRs. We developed a novel method to extract familial relationships from EHRs using emergency contact information collected during hospital admissions. The method matches an emergency contact person to his or her existing medical record in our healthcare system and documents any familial relationship between the emergency contact person and the patient being admitted (e.g., child, sibling, parent). A match was established when only one medical record matched the emergency contact information. Once the family relationships were extracted, we inferred additional relationships based on the family structure. We then identified families by graphical analysis of these relationships. We repeated this analysis independently at two large academic medical centers (AMCs). Among 1.1 million patients, we identified 379,190 families spanning up to four generations (4.7 million familial relationships in total). Pedigrees included from 2 to 134 members, and 199 families with up to fourth-degree relatives (e.g., a great-great-grandchild). We evaluated inferred relationships against two reference standards, one clinical and one based on genetics. Against the clinical reference, we obtained 92.9% sensitivity with 95.7% positive predictive value (PPV) at the first AMC and 96.8% sensitivity with 98.3% PPV at the second AMC for mother/child relationships. Using the genetic relatedness of 117 individuals, we found that parent/child relationships and sibling relationships had a PPV of 98.1% and 71.4%, respectively. This novel method for identifying familial relationships using EHR data unlocks new opportunities such as disease risk prediction and heritability estimates for phenotypes that have not been previously studied.

1912W

Bioinformatics workflow for whole genome sequence linkage analysis of multiple families afflicted with rare disease of unknown heredity and penetrance. *J.L. Rodriguez-Flores[1], M.D. Ramstetter[1], M.R. Staudt[1], A. Robay[2], K.A. Fakhro[3], J.G. Mezey[1], J. Salit[1], J. Malek[2], C. Abi Khalil[2], R.G. Crystal[1]. 1) Genetic Medicine, Weill Cornell Medical College, New York, NY USA; 2) Genetic Medicine, Weill Cornell Medical College - Qatar, Doha, Qatar; 3) Translational Medicine, Sidra Medical and Research Center, Doha, Qatar.*

Linkage analysis is experiencing a renaissance thanks to the wide availability of whole genome sequencing, which enables identification of rare variants that segregate with a rare disease phenotype in one or more families. In many cases of rare disease, the inheritance mode (dominant, recessive), the region of the genome affected (autosomal, sex-linked), the number of linked variants, and the degree of penetrance is unknown. In addition, it is not known a priori if the same variant, gene, or pathway will segregate with the phenotype in every family. A bioinformatics analysis workflow that efficiently explores all potential scenarios and accounts for ancestry and relatedness is needed, and was developed and tested in this study. Ancestry-informative markers were used to infer individual ancestry, followed by mapping reads to a population-specific reference genome that incorporates major allele SNP and indel variants. Analysis of disease families using a population-specific reference genome produced a >3x increase in mapped read depth while cutting the number of linked variants for functional analysis by 44%. Variant calling using GATK best practices followed by functional annotation identified an average of 44 thousand potentially deleterious minor allele coding variants per genome. Using an analysis package written in Python, both recessive and dominant models were tested, as well as both complete and variable penetrance. This workflow identified rare potentially deleterious candidate variants that segregate with the disease phenotype. Candidate variants were filtered based on prevalence in global populations, and a set of candidate genes were identified. Gene set enrichment analysis of the linked genes identified enriched pathways relevant to the phenotype. Candidate causative potentially deleterious variants were identified, including both cases where the same variant segregates with the phenotype in all families, as well as cases where a different variant in the same gene or pathway segregates with the phenotype in each family. By calculating the degree of relatedness between families, the likelihood of all affected individuals inheriting causative variants from a common ancestor was estimated, suggesting the possibility of inheritance from a distant common ancestor. This approach is applicable to a wide spectrum of rare diseases, and identifies candidate genes missed by standard approaches.

1913T

GeneHancer: A comprehensive database of annotated human enhancers, key to interpreting intergenic regions in whole genome sequences. *D. Lancet[1], S. Fishilevich[1], Y. Golan[2], D. Dahary[2,3], M. Safran[1], R. Nudel[1], D. Cohen[1]. 1) Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel; 2) LifeMap Sciences Inc., Marshfield, MA 02050, USA; 3) Toldot Genetics Ltd., Hod Hasharon, Israel.*

This year, the $1,000 whole genome target became a reality, leading to an avalanche of whole genome sequencing (WGS) projects worldwide. Yet, deciphering the ensuing data poses a major challenge, necessitating significant improvements in the annotation and interpretation of intergenic variants. Enhancers are remote regulators of gene expression, whose aberrations underlie certain diseases. They are likely the most abundant functional entities in non-coding DNA, hence of great significance to WGS analyses. We present GeneHancer, a novel database of human enhancers and their inferred target genes, in the framework of GeneCards (www.genecards.org). GeneHancer computes combinatorial likelihood scores for enhancer-gene pairs, and displays charts of probable enhancers for every human gene. It thus makes it possible to map variants to enhancers, and forms a basis for their gene-phenotype interpretation. These capacities will soon be incorporated into the GeneCards Suite next generation sequencing tools, TGex and VarElect (BMC genomics, in press). GeneHancer integrates information on ~258,000 enhancer entries from three genome-wide enhancer databases. The integration includes: a) 213,000 elements from the Ensembl regulatory build, predicted based on the ENCODE and Roadmap Epigenomics projects; b) 43,000 elements from FANTOM, identified via enhancer RNA (eRNA) transcriptomics; c) 1,700 elements from the VISTA Enhancer Browser validated by transgenic mouse assays. GeneHancer thus portrays 237,000 integrated enhancers, with ~17,500 derived from more than one source. Additional scheduled integration includes DENdb predicted enhancers in cell-lines. GeneHancer subsequently links enhancers to genes, using the following methods: 1) Employing FANTOM expression correlation between eRNAs and candidate target genes; 2) Making use of GTEx expression quantitative trait loci (eQTLs), exploring genetic association between variants within enhancers and the expression of candidate target genes; 3) Seeking across-tissue expression correlation (PMID 27048349) between a transcription factor interacting with enhancers, and candidate target genes; 4) Using Hi-C chromosome conformation capture data as indications of genomic looping between a gene promoter and enhancer. The individual scores based on these four methods, along with gene-enhancer genomic distances, form the basis for GeneHancer's combinatorial likelihood scores for enhancer-gene pairs.

1914F

Expanding and improving the 1000 Genomes Project data resources in the International Genome Sample Resource (IGSR). *S. Fairley, R. Amode, P. Harrison, E. Lowy, E. Perry, I. Streeter, S. Trevanion, X. Zheng-Bradley, L. Clarke, P. Flicek.* EMBL-EBI, Hinxton, United Kingdom. CB10 1SD.

The International Genome Sample Resource (IGSR) builds on the data resources created by the 1000 Genomes Project, moving the data to GRCh38, expanding the collection to new samples and new data types, and improving the usability of the resources. Data from phase three of the 1000 Genomes Project has been mapped to GRCh38, using the alt-aware BWA alignment algorithm. These alignments provide a foundation for generating variant calls directly on GRCh38. Preliminary single-tool call sets are being generated using SAMtools, FreeBayes and GATK. Samples beyond those included in the 1000 Genomes Project have been added to IGSR. These share similarly open consent to samples from the 1000 Genomes Project. Such samples include those in the HGDP-CEPH collection and new samples from Gambian populations. It is planned to add further populations when suitable data sets are available and IGSR welcomes potential collaborators, with IGSR being able to provide support for ethical review, data coordination, data distribution, alignment and variant calling. IGSR provides links to raw data generated from the samples it contains. This includes various new data types including 10X Genomics, BioNano optical maps, PacBio data and RNA-seq. Where possible, alignments and variant calls are also made available. To facilitate finding data, a new portal has been developed. The portal gives an overview of the available data, both by project and data type, and can be used to browse collections of data, search for data on specific samples and to download files. Further improving the usability of the data, tools from the 1000 Genomes Project genome browser are being introduced in Ensembl, allowing users to work more easily with the data alongside the rich collection of data available in Ensembl. IGSR plans to continue to develop and improve the resources it hosts, with the aim of maintaining and enhancing access to openly consented genomic data, which provide a valuable resource both as a reference panel of human genetic variation and as a data set for use by the methods development community.

1915W

NGS-SwiftCluster: A parallel computational framework for large scale genetic variation analysis in cloud. *C. Xiao, S. Sherry.* NIH, Bethesda, MD.

Large scale genetic variation analysis plays an important role in elucidating the causes of various human diseases. However, the tera- and peta-byte scale footprint for sequence data imposes significant technical challenges and financial burdens for data management and analysis, including the tasks of collection, storage, transfer, sharing, and privacy protection. Cloud Computing has recently emerged as a compelling paradigm for large scale data management and big data analysis due to the availability, scalability, and cost effectiveness of computing resources across the global. It is still a huge challenge, and often big headache, for researchers to configure and set up appropriate instance clusters in a cloud for parallel computing with all the necessary software environments before a meaningful analysis tasks can be performed. Here, we provides a parallel computational framework, NGS-SwiftCluster, for the large scale genetic variation analysis using next-generation sequencing data in cloud. This framework not only includes a customized machine image with preconfigured tools and pre-selected analysis pipelines for read alignment and variant calling (https://github.com/ngs-swift/Introduction), but also enables users to launch and manage grid-engine cluster with multiple instance nodes so that the time consuming analysis can be expedited by distributing vast computing jobs into multiple nodes. Performance of this framework has been evaluated in cloud using dbGaP study phs000710.v1.p1 (http://www.ncbi. nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000710.v1.p1), and security handling in cloud environment when dealing with control-accessed sequence data has also been addressed.

1916T

Opening the door to large scale use of clinical lab measures for association testing and phenomic profiling: Sparse multivariate analyses in electronic health records reveals novel genetic variants affecting disease risk. *C.R. Bauer[1], J.M. Mahoney[2], S. Setia[1], D.R. Lavage[1,3], J. Snyder[1,3], J. Leader[1,3], H.L. Kirchner[1], S.A. Pendergrass[1].* 1) Biomedical & Translational Informatics, Geisinger Health System, Danville, PA; 2) Department of Neurological Sciences, University of Vermont College of Medicine, VT; 3) Phenomic Analytics and Clinical Data Core, Geisinger Health System, Danville, PA.

Personalized genomic medicine promises to improve medical screening, diagnosis, and treatment but this first requires construction of an accurate genotype-phenotype map. Association studies have linked many genetic variants to disease phenotypes, but most have relied upon rigorous assessment of case-control status. Electronic health records contain a wealth of quantitative clinical lab measures that can also be used for association testing. While patients generally have few specific diagnoses, many have results from dozens of different clinical assays providing an opportunity to construct high dimensional phenotypic profiles. Factor analysis and sparse discriminant analysis allow us to project this high dimensional data into a space that aligns with the etiology of disease states and the primary effects of genetic variants. We analyzed a dataset of 143 outpatient clinical lab measures across 859,010 individuals from the Geisinger Health System. By clustering measure co-occurrence and imputing missing values, we identified latent phenotypic factors in modules of clinical variables that can serve as useful physiological traits for association testing. We then used sparse linear discriminant analysis to identify the precise profiles of physiological states that are altered in individuals with specific gene variants by integrating these quantitative traits with genotypic data for 38,813 unrelated European Americans from the DiscovEHR cohort. Even for a module of four heavily studied measurements of serum lipids, we identified novel genetic associations with linear combinations of these variables that are not significant with any of the individual lipid traits. For example, SNP rs11080055 shows a relationship with increased triglycerides (b=0.022, p=4.5x10^{-3}) and HDL cholesterol (b=0.017, p=0.035) but reduced LDL cholesterol (b=-0.015, p=0.059). While none of these effects approach genome-wide significance independently for this SNP, discriminant analysis identifies that the vector of main effect is related to reduced LDL cholesterol and increased levels of all other lipid measurements (b=0.05, p=3.5x10^{-9}). Our approach demonstrates the feasibility of high-throughput analyses of quantitative clinical lab data and highlights important considerations for their use. These high-dimensional data allow us to identify novel genetic variants that impact disease risk and provides deeper insights into the structure of the genotype-phenotype map.

1917F

Exome sequencing is the most cost-effective approach for rare variant-based novel gene discovery. *L.C. Francioli[1,2], A. Manning[2], M. Lek[1,2], M.J. Daly[1,2], D.G. MacArthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

Although genome-wide association studies (GWAS) in complex diseases and sequencing studies in rare Mendelian disorders have been extremely successful in identifying disease loci, much of the genetic disease burden remains unexplained. Analysis of common variation generally uncovered only alleles of weak effect, reflecting negative selection on deleterious alleles. Some Mendelian studies have discovered *de novo* or very rare alleles of large effect, but these studies are too small to find modest effect contributors. Power to detect novel disease genes is a function of sample size, risk allele frequency and effect size. It is estimated that ~25k cases are needed to find associations based on rare moderate effect variant burden tests (Zuk *et al.* 2014). Here, we compare whole-genome (WGS) and whole-exome sequencing (WES) study designs for novel gene discovery. Dense SNP microarrays and imputation have provided a relatively unbiased sampling of common human genetic variation; thus GWAS results should reflect the properties of the variation underlying common diseases. Using highly confident and replicated disease-associated loci from the GWAS catalog, we show that while 87% of associated loci are non-coding, those with higher odds ratio (OR) are enriched for coding variants. Indeed, 13 of the 26 loci with OR > 2.0 are either coding or in high linkage disequilibrium (r^2 > 0.8) with a coding SNP. This enrichment is unlikely due to differences between disease as all 7 associations with OR > 2.0 are coding when restricting our analysis to diseases with >10 associated loci. Similarly, we outline a number of independent lines of evidence from rare disease studies (including low proportion of unsolved families following successful linkage analysis, and exome solve rates in multiplex recessive pedigrees) pointing to >80% of causal variants in Mendelian disease being coding. While some studies reported high yield from WGS compared to WES, we find that this difference is largely due to comparison against outdated WES technologies. These results indicate that the majority of the variation we can expect to have power to detect with large-scale sequencing will be detectable with WES. Through simulation of different cost structures and allelic architecture, we show that WES is likely to remain the optimal approach for rare variant-based gene discovery in both common and rare disease for several more years, and provide guidelines for the design of large-scale sequencing studies.

1918W

GenAMap on the web: Intuitive and scalable machine learning for structured association mapping. *B.J. Lengerich[1], H. Wang[2], M.K. Lee[3], E. Xing[2,3,4,5].* 1) Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA; 2) Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA; 3) Center for Machine Learning and Health, Carnegie Mellon University, Pittsburgh, PA; 4) Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA; 5) Computer Science Department, Carnegie Mellon University, Pittsburgh, PA.

The proliferation of genomic data has increased the usefulness of complex machine learning algorithms for structured association mapping. Such methods can effectively relate genetic polymorphisms with phenotypes, but correct use requires algorithmic expertise to run code and domain expertise to analyze results. To overcome these challenges, the GenAMap software platform was developed and released in 2010. Since then, the sizes of available biological data have continued to increase exponentially. To address these challenges, GenAMap is redesigned for scalability and updated with state-of-the-art methods for efficient calculations on human genome-scale data. The user experience is overhauled as an intuitive web application with a focus on simplicity and ease of use. GenAMap is available through http://sailing. cs.cmu/main/genamap.

1919T

Bayesian multivariate analysis of large genetic studies identifies novel associations. *M.C. Turchin[1], M. Stephens[1,2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL.

Genome-wide association studies (GWAS) are now a common tool to identify genetic variants that affect traits of interest. To date, the NHGRI GWAS Catalog has over 24,000 SNP-phenotype associations. However, the vast majority of these GWAS are conducted in univariate frameworks, ie when genetic variants are only tested against a single phenotype one at a time. This is in contrast to multivariate frameworks where genetic variants are tested against different combinations of traits simultaneously. Under many biological scenarios, multivariate frameworks drastically increase power. Additionally, by testing combinations of traits, multivariate frameworks allow researchers to investigate a greater level of biological complexity. Despite these clear advantages, multivariate analyses are seldom conducted. Univariate GWAS already involve a large computational and statistical burden; performing an additional, exponentially greater number of tests is highly deterring. Furthermore, it is often unclear how to properly compare different multivariate models even when they can be efficiently conducted. Here, we present a framework and R package that alleviates these obstacles -- Bayesian multivariate analysis of association studies, or bmass. bmass runs using solely univariate GWAS summary statistics. bmass can quickly conduct all possible multivariate analyses for up to 8 phenotypes. And bmass provides Bayes factors for each multivariate analysis, thus allowing models to be directly compared. Running bmass on various publicly available GWAS datasets consistently show an increase in power up to 40% over univariate approaches while keeping FDRs as low as 15%. bmass identifies many new significant associations as well as the phenotypic combinations driving these associations, thus providing a novel level of biological insight. Overall, bmass is a powerful tool enabling multivariate analysis of GWAS.

1920F

Picopili: Development and evaluation of a pipeline for quality control, imputation, and genome-wide association in family-based studies with arbitrarily complex pedigrees. *R.K. Walters[1,2], S. Ripke[1,2,3], B. Neale[1,2,4], Substance Use Disorder Group of the Psychiatric Genomics Consortium.* 1) Analytic & Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, Boston, MA; 3) Department of Psychiatry and Psychotherapy, Charité, Campus Mitte, Berlin, Germany; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Boston, MA.

   Genome-wide association studies of case/control and trio cohort samples have greatly benefitted from the development of standardized QC and analysis pipelines implementing best practices and ensuring uniform, rigorous analysis of array data to facilitate genome-wide meta-analysis. In order to extend these benefits to include family-based studies, we introduce picopili: a flexible, freely-available pipeline for QC, imputation and genome-wide analysis of cohorts with study designs ranging from sib-pairs to complex, multi-generational pedigrees.   The development of picopili gives particular focus to evaluating three areas of GWAS analyses that lack a clear consensus for best practices in family-based data. First, picopili provides a unified interface for estimating genetic relatedness under models appropriate for homogenous or admixed samples, with the optional inclusion of reference samples to specify the ancestral populations of the admixed samples. Comparison of results from the available models demonstrates the risk of substantial upward bias in kinship estimates when population structure is misspecified.   Second, picopili automates imputation of family-based samples to standardize alignment of array data to the desired imputation reference, pre-phasing of haplotypes with a pedigree-aware hidden markov model, imputation, and calling of best-guess genotypes with identification of mendelian errors. Marked increases in imputation quality are observed from this approach, with a net gain of more than 10,000 SNPs with info > 0.9 for sib-pair data compared to a case/control design, with much larger gains anticipated for larger pedigrees.   Finally, picopili enables association testing under multiple modeling frameworks, including transmission-based tests, generalized estimating equations (GEEs), and logistic mixed models. Comparison of these testing methods in simulated and real data suggests that the optimal model is highly depending on the study design, with computational advantages observed for GEEs in simple pedigrees versus markedly better statistical properties for logistic mixed models in larger, more complex pedigrees.

1921W

Incorporating 'topologically associating domain' content into pathway analyses to implicate causal genes at GWAS loci. *G.P. Way, C.S. Greene, S.F.A. Grant.* University of Pennsylvania, Philadelphia, PA.

   Genome wide association studies (GWAS) have revolutionized the field of complex disease genetics. However, GWAS only reports signals associated with a given trait and not necessarily the precise location of culprit genes. The past ten years of GWAS did not strictly represent a decade of gene target discovery; rather, it was a decade of signal discovery. Most of these association signals are found in non-coding regions and assigning the variants to causal mechanisms has proven challenging. Topologically associating domains (TADs) are cell-type independent, genomic regions that define interactome boundaries and therefore offer an opportunity to set limits to where the causal gene(s) at a given GWAS locus must reside. We calculated that the average TAD in human embryonic stem cells spans 852kb and harbors 17 genes, 1,585 repetitive elements, and 2,200 common single nucleotide polymorphisms (SNPs). We observed that protein coding genes strongly cluster near TAD boundaries (χ^2 = 1005.5, P = 1.2x10^{-220}), as do ALUs (χ^2 = 1809.3), while common SNPs are more frequent on the 5' side of TADs (χ^2 = 154.9, P = 1.4x10^{-35}). We developed 'TAD pathways' as a computational method to identify causal genes independent of nearest gene biases. We performed a Gene Ontology (GO) process analysis over TADs and treated all genes equally in TADs that harbor a GWAS signal. This revealed disease-associated GO processes and suggested causal genes. We applied our pipeline, which can withstand content that is predominantly presumably non-causal, to bone mineral density (BMD) and type 2 diabetes (T2D) GWAS catalogs, identifying, respectively, 'Skeletal System Development' (Benjamini Hochberg adjusted P = 1.02x10^{-5}) and 'Insulin Secretion' (BH adjusted P = 1.8x10^{-2}) as the top ranked pathways. In many cases the nearest genes to the given sentinel SNP were not implicated. For example, *ERC1* is closest to a BMD signal on chromosome 12, but our approach implicates *WNT5B* as the most likely causal gene. Likewise, and as a further example, *KCNK16* is closest to a T2D signal on 6p21, but our approach implicates *GLP1R*. While these have clear biological support for the 'correct' gene, we also identify genes in situations where no obvious gene resides within the TAD; for instance *DEAF1* at the '*BET1L*' BMD locus. In conclusion, our computational method can leverage TADs in order to infer causal genes at GWAS signals.

1922T

Phased human genome assemblies with single-molecule real-time sequencing. *C. Chin[1], F. Sedlazeck[2], G. Concepcion[1], P. Peluso[1], D. Rank[1], M. Schatz[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) Department of Computer Science, Johns Hopkins University, Baltimore, MD.

   In recent years, human genomic research has focused on comparing short-read data sets to a single human reference genome. However, it is becoming increasingly clear that significant structural variations present in individual human genomes are missed or ignored by this approach. Additionally, remapping short-read data limits the phasing of variation among individual chromosomes. This reduces the newly sequenced genome to a table of single nucleotide polymorphisms (SNPs) with little to no information as to the co-linearity (phasing) of these variants, resulting in a "mosaic" reference representing neither of the parental chromosomes. The variation between the homologous chromosomes is lost in this representation, including allelic variations, structural variations, or even genes present in only one chromosome, leading to lost information regarding allelic-specific gene expression and function. To address these limitations, we have made significant progress integrating haplotype information directly into genome assembly process with long reads. The FALCON-Unzip algorithm leverages a string graph assembly approach to facilitate identification and separation of heterozygosity during the assembly process to produce a highly contiguous assembly with phased haplotypes representing the genome in its diploid state. The outputs of the assembler are pairs of sequences (haplotigs) containing the allelic differences, including SNPs and structural variations, present in the two sets of chromosomes.   The development and testing of our de-novo diploid assembler was facilitated and carefully validated using inbred reference model organisms and F1 progeny, which allowed us to ascertain the accuracy and concordance of haplotigs relative to the two inbred parental assemblies. Examination of the results confirmed that our haplotype-resolved assemblies are "Gold Level" reference genomes having a quality similar to that of Sanger-sequencing, BAC-based assembly approaches.   We further sequenced and assembled two well-characterized human samples into their respective phased diploid genomes with gap-free contig N50 sizes greater than 23 Mb and haplotig N50 sizes greater than 380 kb. Results of these assemblies and a comparison between the haplotype sets are presented.

1923F

Beacon Network: A system for global genomic data sharing. *M. Cupak[1], M. Fiume[1], V. Rockai[1], S. Keenan[2,3], D. Lloyd[4], P. Goodhand[3].* 1) DNAstack, Toronto, Ontario, Canada; 2) EMBL-European Bioinformatics Institute, Hinxton, United Kingdom; 3) The Global Alliance for Genomics and Health, Toronto, Ontario, Canada; 4) Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

   Large volumes of genomic data have and will be collected to realize new discoveries and applications in medicine. No single institution has enough resources to process these vast quantities of information, which is why it is necessary to develop standards for interoperability and sharing across a federated network of genomic data resources. The development of standard protocols for genomic data discovery and exchange will open the Internet of Genomics and increase the collective power of genomics datasets being collected worldwide to resolve the causes of genetic diseases and accelerate the implementation of precision medicine. The Global Alliance for Genomics and Health (GA4GH) is a coalition of hundreds of organizations from around the world who work together to develop regulatory, ethical, and technical standards for genomics-derived advancements in human health. The Beacon Project is a GA4GH initiative that is writing an open technical specification for genetic variation discovery and sharing. It also develops regulatory, ethics, and security guidance and tools to ensure proportionate safeguards for distribution of data through the protocol according to the GA4GH-developed Framework for Responsible Sharing of Genomic and Health-Related Data. We introduce the Beacon Network, a global directory and search engine for Beacons. The system serves as a powerful, convenient, and interactive genomic data distribution channel through which users can determine the existence of alleles of interest and be directed to host organizations who have observed them. It enables global discovery of genetic data from patients, population cohorts, curated databases, and the scientific literature, federated across a large and growing network of shared genetic datasets. With tens of organizations and hundreds of datasets having joined the system, the Beacon Network is one of the largest aggregators of human genomic variation in the world.

1924W

The spectrum of loss of function tolerance in the human genome. *K.J. Karczewski[1,2], K.E. Samocha[1,2], M. Lek[1,2], D. Birnbaum[1,2], I. Armean[1,2], P. Singh[1,2], P. Natarajan[1,2], V. Narasimhan[3], A. Palotie[1,2], D. van Heel[4], R. Trembath[4], R. Durbin[3], S. Kathiresan[1,2], M.J. Daly[1,2], D.G. MacArthur[1,2], Exome Aggregation Consortium.* 1) MGH/Broad Institute, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) Wellcome Trust Sanger Institute, Hinxton, UK; 4) Queen Mary University of London, London, UK.

Loss-of-function (LoF) variants, which are predicted to severely disrupt the function of protein-coding genes, are found in every individual; however, there is a broad range of effects of these variants. Most LoF variants are benign, while some may even confer beneficial phenotypes: for instance, LoF variants in *PCSK9* have been causally linked to low LDL cholesterol levels, leading to the development of PCSK9 as a therapeutic target for cardiovascular disease. On the other end of the spectrum, some genes are essential, and thus, LoF variants can never be observed in living humans in the homozygous (or even heterozygous) state. Understanding exactly where each human gene lies along the spectrum between these extremes is important for prioritizing variants both as candidate disease genes and for the development of inhibitory therapeutics.   In order to define the spectrum of LoF variants across humans, we have developed an enhanced version of our prior LoF constraint model [Samocha et al., 2014], to incorporate a range of filters to remove common annotation errors as well as accommodating transcript-specific annotations. We apply this model to variants identified from over 100,000 exomes from the Exome Aggregation Consortium (ExAC) to identify genes with severe LoF depletion. In addition, we characterize LoFs present in this dataset and in other large datasets from collaborative efforts (including deeply phenotyped samples from Finnish national biobanks as well as over 1,000 parentally-related individuals from the UK) in order to identify a set of genes at the other end of the spectrum where homozygous LoF is consistent with survival into adulthood.   We correlate this spectrum with similar frameworks derived from mouse models and cellular assays. Additionally, we investigate how the network and expression properties of genes vary depending on their location on this spectrum. With these data, we can begin to characterize genes on the spectrum of LoF mutational tolerance across all human genes.

1925T

Grouped variable selection methods for microbiome compositional data. *A. Plantinga[1], M.C. Wu[1,2].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

Community-level and species-level analysis of microbiome sequencing data have revealed relationships between shifts in microbiome composition and a wide range of diseases and conditions. However, community level analysis does not allow identification of specific taxa that are associated with the outcome, whereas conventional species-level analyses may have low power due to multiple testing. Variable selection methods for high-dimensional compositional covariates, such as microbiome data, avoid the multiple testing problem while still allowing estimation of the effects of specific taxa. We propose an L2 penalized log-contrast model for estimating taxon-level regression coefficients and selecting groups of taxa (e.g., genera or families) that are associated with an outcome of interest. Compositional and subcompositional coherence of the regression model are enforced using a set of linear constraints on the coefficients. Simulation results show that this method correctly selects groups of taxa that are associated with the outcome. We apply the proposed method to an upper-respiratory-tract microbiome dataset comparing microbiome composition between adult smokers and non-smokers.

1926F

Navigating GRCh38 at NCBI. *V.A. Schneider, V. Ananiev, W.R. Anderson, N. Bouk, H.C. Chen, C. Clausen, M. DiCuccio, P. Kitts, A. Kuznetsov, P. Meric, T. Murphy, K. Pruitt, T. Rezaie-Jami, E. Weitz, the Genome Reference Consortium.* IEB, NIH/NCBI, Bethesda, MD.

   NCBI is a member of the Genome Reference Consortium (GRC), the group responsible for updates and improvements to the human reference genome assembly. Since the release of GRCh38 in December 2013, the GRC has also released more than 70 patches to this assembly. These fix assembly errors and introduce new sequence representations for variant or complex genomic regions not sufficiently represented in the haploid chromosome assemblies or current suite of alternate loci scaffolds, all without changing assembly coordinates. We will present examples of recent patches, including those of clinical importance. We will also discuss how the GRC is using new assemblies and genome maps to identify sequences missing from the reference and genomic regions that are candidates for improvement or variant representation. NCBI provides several tools that enable users to navigate GRCh38 and its patch updates along a variety of axes. Among these, the Genome Data Viewer (GDV), a new addition to the suite of NCBI browsers, supplies a graphical view of the assembly accessible from a variety of NCBI resources, such as Gene, Assembly and GEO. The displays in this browser are augmented with data tracks relevant to the referring resource and provide expanded access to the wealth of assembly annotations available at NCBI. Track Sets and Collections in GDV and other viewer instances at NCBI further enable users to quickly reconfigure and share browser displays. In addition, users can upload their own data in a variety of formats to GDV for visualization. We will present the features of GDV, including a demonstration of how it provides access to the patches and alternate loci. We will also discuss recent improvements to the NCBI Remapping Service, a tool that transitions user data between different assemblies and also supports mapping between reference assembly chromosome sequences and alternate loci and patch scaffolds. We will illustrate how the assembly-assembly alignments that underlie the service, along with its intrinsic support for the reference assembly structure, provide a highly accurate remapping of data associated with complex assembly updates or genomic regions. These, and other NCBI resources, facilitate the navigation of GRCh38 needed for scientific discovery.

1927W

Modeling the effects of linkage disequilibrium and allele specific expression on genotype/phenotype relations. *J. Dannemiller.* Rice University, Houston, TX.

   Allele Specific Expression (ASE) is thought to result from the effects of one or more regulatory variants in *cis* with specific alleles in the protein-coding region of a gene. Protein-coding variants (e.g., SNPs) of a gene can directly alter a phenotype downstream from and dependent on that gene. What impact might ASE have on such phenotypes by the mechanism of altering the expression *levels* of the protein-coding variants? *Purpose:* To model the effects of ASE on the hypothetical phenotypes produced by a single bi-allelic protein-coding SNP one of whose two alleles produced a protein assumed to be 25% more biologically active than the protein produced by the other allele. *Methods:* Analytic expressions were derived for estimating the mean effects at equilibrium of ASE on the Percentage of (phenotypic) Variance Explained (PVE) by the three genotypes of the protein-coding SNP in a sample of 600 hypothetical subjects. The modeling was done using a simple, linear regulatory-SNP—protein-coding-SNP system with either perfect linkage disequilibrium (LD) or perfect linkage equilibrium between the two bi-allelic loci. The phenotype depended both on the number of protein molecules of each type as well as on their relative biological activities perturbed by additional pathway noise with controlled variance. ASE was modeled using an empirical distribution of expression ratios for the gene *TGFBR1* shown in Valle et al. (2008). Minor allele frequency was set at 0.5 or 0.2 both for the regulatory and the protein-coding SNPs. *Results:* Several clear trends were evident across the different conditions: 1) The degree of LD has a very large impact on PVE, 2) ASE reduces PVE negibly when LD is 0, but substantially when LD is perfect, and 3) The phases of the alleles at the two loci can exert a large effect when LD is perfect; if the mean transcription rate in the ASE distribution is near the ratio of the biological activities of the two proteins as it was here, larger phenotypes ascribed to the more biologically active protein can be completely offset by the ASE up-regulation of the less biologically active protein molecules. This leads to apparent PVE values near 0. *Conclusions:* Allele specific expression as modeled using a single bi-allelic regulatory SNP can have a substantial range of effects on the phenotypic variance explained by the genotypes of a single, hypothetical bi-allelic protein-coding SNP. The impact of ASE is more evident when linkage disequilibrium is high.

1928T

Association analysis as a service. *J. LeFaive, E.M. Schmidt, S. Lee, P. VandeHaar, H.M. Kang, G. Abecasis.* Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI.

   We have designed a web-based interface where genomic researchers can carry out large scale association analysis combining phenotypes and genetic data on tens of thousands of individuals and hundreds of millions of genetic variants. Our tool allows researchers to focus on results interpretation by automating the process of running complex software pipelines and managing compute cluster infrastructure. With our platform, a large set of genotypes is prepared for analysis and stored in a cloud environment. Users can then request analysis of specific phenotypes and covariates to be queued and processed in parallel across multiple machines. The results are then tracked and summarized and can be explored through interactive plots, tables and downloads. The current version provides built-in analyses using EPACTS (http://genome.sph.umich.edu/wiki/EPACTS), including Wald, Firth and SKAT-O tests. Future versions will allow further customization of analysis options. Our tool is designed to facilitate trackable, reproducible research and dynamically allocate compute servers on virtual hosting services in a provider-agnostic manner. As genomic data increases in size, the complexity of analyzing this data in a timely fashion follows suit. Our goal is to simplify this process in an effort to streamline research.

1929F

HapRFN: A deep learning method to identify short IBD segments. *S. Hochreiter[1], D.-A. Clevert[2], G. Povysil[1].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Bayer Pharma AG, Berlin, Germany.

   For whole genome sequencing data HapFABIA was shown to be superior in detecting short IBD (identical by descent) segments that are tagged by rare variants. Nevertheless, HapFABIA still has several problems: (1) To decide whether individuals possess an IBD segment is often difficult because of the soft bicluster membership supplied by HapFABIA. (2) HapFABIA can only extract 10-30 IBD segments at once and therefore needs to perform multiple iterations. However, the IBD segments identified in different iterations may not be decorrelated, thus they may be redundant and overlapping or even split into smaller segments. (3) Very large data sets are time intensive. We recently introduced Rectified Factor Networks (RFNs) as an unsupervised deep learning approach. Each code unit of the RFN represents a bicluster and therefore an IBD segments, where samples for which the code unit is active share the bicluster (IBD segment) and features (SNVs) that have activating weights to the code unit tag the IBD segment. HapRFN overcomes the problems of HapFABIA. (1) RFNs provide sparser codes via their rectified linear units that immediately supply bicluster memberships as factors being different from zero. (2) RFNs can learn thousands of factors and therefore many IBD segments simultaneously. Therefore, all IBD segments are mutually decorrelated, thus are not redundant and do not overlap. (3) RFNs allow for much faster processing of very large data sets using techniques from deep learning like efficient matrix multiplications and implementations of networks on graphical processing units (GPUs). As a result HapRFN makes it possible to process very large data sets and to determine the size and number of IBD segments more precisely. With HapRFN we are able to accurately detect familial relationships, populations of origin, or interbreeding with ancient genomes in data sets with thousands of individuals. Furthermore, finding disease associations via IBD mapping becomes more reliable which might be the key to uncover unknown hereditary causes of multifactorial diseases.

1930W

Improving power of association tests using genotypes imputed from different reference panels. *W. Zhou[1], M.B. Elvestad[2], K. Hveem[3,4], J. Chen[5], C.J. Willer[1,5,6], G.R. Abecasis[7], H.M. Kang[7].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 3) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway; 4) Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, 7600 Levanger, Norway; 5) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 6) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 7) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, 48109, United States of America.

Genotype imputation allows variants that are not directly genotyped to be studied without other costs than computation. Improved genetic similarity between the reference and target samples increases the accuracy of imputation. The imputation accuracy also increases as the reference panels contain more individuals, especially for lower-frequency variants. When both a smaller ethnically-matched reference panel and a large cosmopolitan reference panel are available, it is unclear what is the optimal strategy to perform imputation-based association studies from these panels. Here, we systematically evaluated and compared genotype imputation based on three reference panels and we compared different approaches to incorporate multiple versions of imputed genotypes in the association tests. First, we genotyped 9,265 individuals from the population-based Nord Trøndelag Health Study (the HUNT study) for 350,270 polymorphic autosomal variants using an Exome array with GWAS tagging markers. Next, we whole-genome sequenced (WGS) 2,201 individuals from the same study matched for ancestry (mean depth 5x) for our reference panel. Then we imputed variants from our HUNT WGS reference panel as our ethnically matched panel, the Haplotype Reference Consortium (HRC) release 1 containing 32,488 individuals and 1000 Genomes Phase 3. We observed that imputation from the HRC resulted in most extremely rare variants (MAF < 0.05%) compared to the other two panels. Imputation from the HUNT WGS uncovered more variants with low frequency (0.05-1%) than HRC and 1000G. For variants with MAF >= 0.5%, the HRC and HUNT WGS panels showed comparable imputation accuracy, even though the sample size of the HUNT WGS is 16 times smaller than the HRC. The HUNT WGS panel added 4.5 million variants that could not be successfully imputed from the other two panels. Our findings have demonstrated the benefits of uncovering variants with low frequency (0.05-1%) by using ancestry-matched reference panels. In addition, simulations showed that using genotypes with the lowest association p value (best p-value method) results in higher power for association tests than keeping only one variant per panel with highest imputation quality metrics (best r2 method), even after adjusting for the additional variants tested, particularly for less frequent variants. For lower frequency variants, best-r2 method often retains variants with much less significant p-values due to imprecise estimate of imputation qualities.

1931T

Leveraging deep genealogical structure to estimate the phenotypic contribution of rare variants. *P. Palamara[1,2], J. Terhorst[3], Y. Song[3,4], A. Price[1,2].* 1) Harvard School of Public Health; 2) Broad Institute of MIT and Harvard; 3) University of California, Berkeley; 4) University of Pennsylvania.

The phenotypic impact of rare and low-frequency variants is poorly understood. Whole-genome sequencing studies offer one route to analyzing this class of variation, but sample sizes are currently too small to draw conclusions about the genetic architecture of a trait. Here, we develop a new method to study the phenotypic effect of rare variants using SNP array data, which is available in large sample size for many traits. Our approach is to first estimate the deep genealogical history for all pairs of individuals in a sample, and then use this information as a proxy for their sequence-level genetic relatedness to infer narrow sense heritability. To reconstruct deep genealogical relationships, we estimate the time to most recent common ancestor (TMRCA) of all pairs of individuals using a coalescent hidden Markov model (HMM) whose states represent TMRCAs along the genome. This type of model has been used in population genetic studies to infer historical effective population sizes from whole-genome sequence data (Li and Durbin, Nature 2011). To enable inferring TMRCAs using array data, we leverage a recently developed approach (Terhorst, Kamm, Song, under review) that enables incorporating site frequency spectrum information from a large panel of individuals into the HMM. We correct for biases due to SNP ascertainment on arrays, comparing the observed and expected allele frequency spectra for a given demographic model. To achieve scalability to large samples, we develop computational speed-ups that enable estimating pairwise TMRCA probabilities in time that is linear in the number of genotyped SNPs and the number of TMRCA states in the HMM. In coalescent simulations of a synthetic chromosome 1 using available models of European demographic history, we attained a squared correlation of 0.92 across pairs of individuals between the true average TMRCAs and the estimates obtained by our coalescent HMM method using array data. Using the reconstructed average TMRCAs for all pairs of individuals, we can infer the sharing of sequence-level genetic variants that arose at different time strata, including recent and rare alleles, which are generally not genotyped using array platforms. We use this information to estimate the phenotypic contribution of causal alleles of different ages, via variance components analysis. We test our approach using extensive simulations, and apply it to analyze the genetic architecture of 11 phenotypes in the UK Biobank data set (N=150k).

1932F

Systematic estimation of the environmental contribution to disease using the electronic health records. *R.S. Vanguri[1], F. Polubriaginof[1], K. Quinnies[1,6], A. Yahi[1], M. Simmerling[7,2], I. Ionita-Laza[3], H. Salmasian[7,1], S. Bakken[1,4], K. Kiryluk[5], D. Goldstein[6], D.K. Vawdrey[7,1], N.P. Tatonetti[1,5,6,8].* 1) Department of Biomedical Informatics, Columbia University, New York, NY; 2) Division of Medical Ethics, Weill-Cornell Medical Center, New York, NY; 3) Mailman School of Public Health, Columbia University, New York, NY; 4) School of Nursing, Columbia University, New York, NY; 5) Department of Medicine, Columbia University, New York, NY; 6) Institute for Genomic Medicine, Columbia University, New York, NY; 7) Value Institute, NewYork-Presbyterian Hospital, New York, NY; 8) Department of Systems Biology, Columbia University, New York, NY.

Accurate estimates of the effect of shared environment on disease is critical to understanding disease heritability. However, environmental factors are difficult to quantify and few datasets exist to derive their effects directly. Electronic health records passively collect a wide range of phenotypic data on millions of individuals who are actively seeking health care. Within these data are many related individuals and families. These relationships can be used to estimate the relative effects of shared environment on the incidence of disease. Through an analysis of the clinical records of 9 million patients at two large independent medical centers we identified 4.7 million familial relationships, from siblings to great-grandparents and great-grandchildren. We used these relationships to estimate the relative effect of shared environment over 900 diseases. This is done by leveraging two properties of heritability and shared environment. The first is that the heritability of a disease is a biological property and not affected by the familial relation. The second is that the shared environment between familial relationships is independent of the disease being studied. By minimizing a likelihood using these assumptions it is possible to isolate the component of observed heritability due to shared environment and calculate heritability more likely due to genetic factors. The diversity and size of the EHR, spanning thousands of clinical phenotypes and many degrees of relationships, allows for a novel method for estimating the effect of shared environment on diseases. Consequently, a measurement of shared environment allows for more accurate estimates of heritability across a broad range of phenotypes derived purely from EHR. .

1933W

VCFped: Identification of small pedigree structures in multisample variant call files. *M. Vigeland.* Dept Medical Genetics, Oslo University Hospital, Oslo, Norway.

We introduce a program, VCFped, which uses forced allele sharing to identify small pedigree structures (e.g. trios and quartets) in multisample VCF files. Correct relationships are crucial for successful analysis of family-based sequencing, e.g. detection of *de novo* variants in trios. Unfortunately, traditional methods for relatedness inference are not easily adaptable to sequencing data, which typically violate the assumptions of independent markers, accurate allele frequencies, and non-censored data. Our program sidesteps all of these problems by exploiting forced allele sharing between closely related individuals. In particular, the algorithm does not depend on allele frequencies or other variant annotations, allowing VCFped to be run as a quality control step before embarking on annotation and downstream analysis. Sequencing and genotyping errors are reduced using an automatically optimized filter based on the available quality metrics from the variant caller. VCFped is implemented as part of the pipeline for diagnostic trio analysis in the Department of Medical Genetics of Oslo University Hospital. In this study we tested VCFped in a range of family-based sequencing projects, including data from whole-genome sequencing, exome sequencing and smaller panels like TruSight One. We show that the VCFped algorithm is highly robust against low sequencing coverage, poor variant quality and inbreeding.

1934T

A versatile simulator for simulating and characterizing third-generation sequencing technologies. *B. Lau[1], M. Mohiyuddin[1], J.C. Mu[1], L.T. Fang[1], N. Bani Asadi[1], C. Dallett[2], H.Y.K. Lam[1].* 1) Roche Sequencing, Belmont, CA 94002; 2) Roche Sequencing, Pleasanton, CA 94588.

Third-generation sequencing (TGS) technology has demonstrated long read length and reduced sequencing bias. Such improvements enhance genome assembly, structural variant detection, haplotype phasing, and repeat resolution. The development, deployment and adoption of TGS can be accelerated by proof-of-concept bioinformatic analysis using simulated data; however, the sequencing characteristics and data formats of TGS have been rapidly changing, hampering the realism and versatility of simulators and bioinformatic tools when applied to different use cases. We present a simulator for profiling and simulating TGS single-molecule technologies. The software is highly adaptive, able to characterize varying sequencing characteristics and to perform I/O with different file formats. We use the software to profile CHM1 hydatidiform mole data sequenced using the P5 and P6 chemistries from Pacific Biosciences (PacBio). The software manages to distinguish the sequencing characteristics of the two chemistries. Such characteristics are automatically stored in numerical models, according to which simulation is performed. Moreover, the software can output the simulated reads in various formats, such as FASTQ, H5, and the latest PacBio BAM. This feature enables downstream analysis of a wide range of tools: PBHoney/Freebayes/Vardict, which require the widely used FASTQ, as well as PacBio's Quiver/CCS2, which require unconventional quality scores in H5/BAM from the primary analysis. The simulator is seamlessly integrated into the VarSim framework, allowing us to demonstrate unbiased comparisons among various variant callers applied to TGS data. The software enables researchers to characterize TGS sequencing technologies and to design experiments accordingly. Bioinformaticians can use simulation to tune against bias in sequencing data. Sequencing and primary-analysis developers can also use the software to predict the performance of downstream analyses.

1935F

RVS: An R package to integrate next generation sequencing (NGS) data across cohorts for association analysis. *J. Gong[1], Z. Baskurt[1], A. Derkach[2], A. Pesevski[1], L.J. Strug[1,3,4].* 1) Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, ON, Canada; 2) Biostatistics Branch, National Cancer Institute, Rockville, MD, USA; 3) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Ontario, Canada; 4) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada.

A growing number of cohorts are undergoing whole genome sequencing using next-generation sequence (NGS) technologies. Integrating NGS data across patient and/or control cohorts for association analysis is challenging. Different studies may use different sequencing platforms and parameters (eg. read depth), and alignment and variant calling algorithms, that can result in spurious findings unless statistical methodology implemented explicitly accounts for these differences. In Derkach *et al.* (2014) we developed a workflow to conduct case-control association analysis for common and rare variants with 'out-of-study' controls when aligned reads in BAM format are available. The proposed method, the robust variance score (RVS), eliminates read depth bias, controls type I error and has comparable power to studies when the true genotypes are known. Here we develop a user-friendly R package to make the RVS accessible, and extend its utility to the integration of NGS data for a variety of study designs. The '*RVS*' R package consists of three modules: (1) common and rare variant association analysis using RVS, (2) conventional association using variant calls, and (3) a simulation module. Module 1 takes a multi-sample variant call format file (VCF, for example, from Genome Analysis Tool Kit (GATK, DePristo *et al*, Nat Genet, 2011)) created from all samples to be analyzed as input, obtains the genotype probabilities for the reported variants, and uses them in the RVS statistic for association analysis. Module 2 uses a conventional score test to for association with variant calls rather than probabilities, for comparison; this module uses the plink (Purcell *et al*, AJHG, 2007) format file as input. Module 3 simulates sequence data for different read depths and error rates. Sample data with examples are provided for each module. The RVS package includes extensions beyond what was described in Derkach *et al* (2014), to analysis with multiple control samples, and for integrating NGS data across consortia for analysis of quantitative traits. The RVS package has been tested using simulated data, data downloaded from 1000 genomes project website (http://www.1000genomes.org) and NGS data from our studies of childhood-onset epilepsy. The package is available at www.tcag.ca. When used with algorithms such as GATK, RVS will provide a useful tool to help researchers identify loci associated with disease while protecting from confounding due to differences in NGS study designs.

1936W

A novel phenotype permutation method to optimize threshold selection in Random Forests. *E.R. Holzinger[1], S. Szymcak[2], J. Malley[3], A. Dasgupta[4], Q. Li[1], J.E. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, USA; 2) Institute of Medical Informatics and Statistics, University of Kiel, Kiel, Germany; 3) Center for Information Technology, National Institutes of Health, Bethesda, Maryland, USA; 4) Clinical Trials and Outcomes Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA.

Standard analysis methods for genome wide association studies (GWAS) are not robust to complex disease models, such as interactions between variables with small main effects. These effects very likely contribute to the genetic etiology of complex human traits. Machine learning methods that are capable of identifying interactions, such as Random Forests (RF), are an alternative analysis approach. One currently unresolved problem with RF or any other machine learning scheme is that there is no standardized method of selecting a threshold that identifies the correct variables while filtering out false positives. The optimal threshold is highly dependent on the underlying genetic model, which is typically unknown in biological data.   To address the long-standing threshold problem, we have developed an RF-based variable selection method called r2VIM. Among other novel features, this method incorporates a permutation scheme to optimize threshold selection. Briefly, we permute the phenotype and run r2VIM on the permuted data. This provides us with a null distribution of importance scores while maintaining the relationships between the predictor variables. A threshold is then selected based on the false positive rate in the permuted data. This threshold is then applied to the non-permuted r2VIM analysis. We refer to this permutation scheme as "phenoPerm." Our results show that phenoPerm provides an ideal balance between power and false positive selection when compared to other selection methods across different complex genetic risk models, including gene-gene interactions, both with and without main effects. We also assess the relationship between the false positive rate in the permuted data to the false positive rate and power in the non-permuted data, and show how it can be used as a guide to perform threshold selection according to individual analysis requirements.

1937T

Improved correction for tissue heterogeneity in DNA methylation data. *E. Rahmani[1], N. Zaitlen[2], Y. Baran[1], C. Eng[2], D. Hu[2], J. Galanter[2,3], S. Oh[2], E.G. Burchard[2,3], E. Eskin[4,5], J. Zou[6], E. Helperin[1,7,8].* 1) Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv, Israel; 2) Department of Medicine, University of California San Francisco, San Francisco, California; 3) Department of Bioengineering and Therapeutic Science, University of California San Francisco, San Francisco, California; 4) Department of Computer Science, University of California, Los Angeles, California; 5) Department of Human Genetics, University of California, Los Angeles, California; 6) Microsoft Research New England, Cambridge, Massachusetts; 7) International Computer Science Institute, Berkeley, California; 8) The Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Tel Aviv, Israel.

Mounting evidence for the role of DNA methylation in various cellular processes and its correlation with disease and environmental factors have led to an increasing interest in Epigenome-Wide Association Studies (EWAS). In a standard EWAS of primary tissue (e.g. whole blood), methylation data represent the epigenetic states of a heterogeneous mixture of cell types. Since different cell types are highly variable in their methylation profiles, correlation between the cell type composition and the outcome of interest leads to spurious associations. In principle, false discoveries due to tissue heterogeneity can be addressed by adding the cell proportions as covariates. However, cell-type compositions are typically hard to measure in practice, and therefore the problem of tissue heterogeneity is believed by many within the epigenomics research community to be amongst the foremost challenges faced by genome-wide epigenetic studies. Although computational methods for estimating cell compositions have been proposed, these rely on scarce reference data of whole-genome methylation levels from sorted cells. Such reference datasets do not exist for most tissues and may not represent well the epigenetic state of individuals with different methylation altering factors such as age and sex. Moreover, accounting only for a preselected set of estimated cell types may still result in false discoveries in EWAS if the outcome of interest is correlated with cell types that were not estimated. We developed a new unsupervised method for reconstructing sparse low rank signal in data, and demonstrated its use and benefits in accounting for the cell composition confounder in methylation data. Using real datasets including measured cell counts and simulated data we showed that our highly efficient method, ReFACTor, increases the correlation with the cell composition and as a result substantially reduces the number of false positives and improves power in EWAS compared with existing methods. ReFACTor does not require reference data; therefore it can be applied to any tissue. We also demonstrate that ReFACTor can be further used for detecting more low rank structures in methylation data such as population structure, thus leading to a more comprehensive understanding of the underlying model of genome-wide epigenetic states. Corresponding software is available from http://www.cs.tau.ac.il/~heran/cozygene/software/refactor.html.

1938F

Efficient user-friendly toolset for the analysis of high-throughput DNA-methylation data. *R. Yedidim[1], E. Rahmani[1], L. Shenhav[3], R. Schweiger[1], O. Weissbrod[2], N. Zaitlen[1], E. Halperin[1].* 1) Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv, Israel;; 2) Computer Science Department, Technion – Israel Institute of Technology, Haifa; 3) Department of Statistics, Tel Aviv University, Tel Aviv, Israel.

 Recently introduced high-throughput technologies for probing DNA methylation states have resulted in a rapidly increasing number of phenotype-methylation associations as well as mounting evidence for the importance of methylation in disease and complex cellular processes. In particular, Epigenome-Wide Association Studies (EWAS) have become a compelling and successful study design. The exponential growth in both data collection (50% of the methylation data currently in GEO were collected since 2014) and related research (66% of the papers including the terms "DNA methylation" and "epigenome-wide" were published since 2014) demonstrate the mounting interest in EWAS. As with other highthroughput biological data, appropriate EWAS analysis requires development of tailor-made methods and tools to address the unique properties of methylation data. Although there is a growing repertoire of available tools, most of these require programming proficiency in order to perform an end-to-end data analysis. User-friendly software, which does not require any programming proficiency and can be easily accessed by more researchers, will substantially benefit the epigenetics community. We developed a highly efficient command-line toolset for the analysis of DNA methylation data and conducting EWAS, analogous to the widely used toolset PLINK for GWAS. Some of the features of the toolset include a range of statistical tests for detecting differentially methylated sites, state-of-the-art methods to account for tissue heterogeneity and confounders including ReFACTor, principal component analysis, and linear mixed models, cell type composition estimation, a method for methylation imputation and ancestry estimation from methylation. In addition, the software provides an easy to use data management tools, and it automatically generates publication-quality figures. The software is highly efficient – we demonstrate how a small-scale EWAS can be performed in minutes using just a few simple commands, and large-scale EWAS with hundreds or thousands of individuals can be performed in less than an hour.

1939W

A novel, widely-applicable method for detecting long-term balancing selection. *K.M. Siewert[1], B.F. Voight[2,3].* 1) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA; 2) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA.

 Balancing selection preserves two or more alleles at a locus over long evolutionary time periods. This type of selection is known to play a role in human evolution, a classic example being the Hemoglobin Beta locus, which maintains alleles that protect against malaria, but cause sickle-cell anemia. Owing to large-scale genomic data sets, discovery and characterization of loci subject to balancing selection has become an active area of investigation. Our goal is to quantify the number of sites and determine which genes are subject to selection in a wide-range of species. To achieve this aim, we require a test that is powerful, robust to assumptions about demographic history that are difficult to model with certainty, not reliant on outgroup sequences, and is computationally efficient so as to quickly scan entire genomes. We are developing an algorithmic approach that utilizes the signature of allelic class build-up, in which neutral variants accumulate at the same frequency as the balanced allele they are linked to. Preliminary results from simulations demonstrate that our method performs comparably to more computationally intensive methods, despite our statistic not requiring knowledge of complex demographic parameters or outgroup sequences. This indicates our method is applicable in a wide range of species where this information is not known. We will next apply our method to population genomics data from humans and drosophila to detect loci under selection..

1940T

New features improve user experience of Analyst Portal-A distributed real-time data query system. *H. Qiu[1], F. Mentch[1], L. Hermannsson[1], H. Hakonarson[1,2].* 1) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

   Analyst Portal is the distributed and integrative data query system developed in-house where users, through web interface, perform complex queries on numerous modalities including but not limited to Electronic Medical Records (EMR), laboratory workflow information, genotyping and next generation sequencing data. Instead of involving a centralized data warehouse to store data from various sources, Analyst Portal uses distributed database query facility to query different source databases on the fly. Since production 6 years ago, Analyst portal has helped researchers search and access the data needed for hypothesis generation and making new discoveries.Based on past user activities, one of the most frequent search is selection of a cohort meeting certain criteria. In the past, Analyst portal used html form based approach in combination with intermediary patient list uploading. Although easier for implementation, it did not score well in terms of flexibility and it was not user friendly. We have developed a new AJAX based, more user friendly cohort selector that allows users to drag-and-drop search terms and build more complex ad-hoc queries that performs well in real-world scenarios.Encouraged by the early benefits seen in the past year from indexing raw SNP calls from different SNP arrays, we extend the same Word Aligned Hybrid (WAH) bitmap indexing to imputed genotypes and SNP calls from NGS sequencing data, making SNP-aware searches more unified and informational.Here we first review the architecture, implementation, lessons learned and user activity patterns of the Analyst Portal, and then present two new features: 1) drag-and-drop enabled query builder for improved cohort selection; 2) unified genotype call indexing, both of which are uniquely enabling for genotype-phenotype queries and association studies.

1941F

Interactive search over massive whole genome repository. *A. Telenti[1,2], V. Lavrenko[2], I. Bartha[2], A. Bernal[2], C. Chang[2], P. Garst[2], A. Harley[2], B. Hicks[2], E. Kostem[2], L. Pierce[2], J. Piper[2], S. Ramakrishnan[2], N. Shah[2], H. Tang[2], W.-Y. Yang[2], J.C. Venter[1,2], F. Och[2].* 1) J. Craig Venter Institute, La Jolla, CA; 2) Human Longevity Inc., San Diego, CA.

   Background. Interacting with thousands of genomes is a significant challenge: both computationally and in seamless integration with phenotype records, knowledge sources and quality metrics. Existing solutions for genomic data management do not scale to a large number of whole genomes. We developed a search engine that that enables access to genomic, phenotypic and annotation data with a high-level of interactivity and flexibility. **Materials and Methods.** We introduce innovative elements into every traditional component of a genomic search engine. The retrieval engine seamlessly integrates free-text keywords and specialized constructs, such as cohort definitions, variant effects, score/frequency filters, gene/chromosome queries and disease panels. The engine is designed to be highly-responsive and responds to almost all queries in under 500 milliseconds. The index makes use of a unified data structure to represent all the underlying information: individual genomes, phenotypes, natural-language and structured annotations. The index supports incremental updates: a whole-genome VCF can be added in 6 seconds and be fully-searchable. The engine supports multiple ranking mechanisms, including rule-based formulas, and a learning-to-rank algorithms. **Results.** We have indexed 11,981 genomes (30x coverage) comprising 217 millions of distinct variants (98% outside the exome). We indexed 1.16 million records from databases such as ClinVar, HGMD and CPDB, providing annotations for 263,305 variants, 58,143 genes and 3816 pathways. The index also contains 10,861 fully-searchable phenotype records. The engine allows individual or collective VCF calls and annotation with multiple annotation sources. Features include visualization of variants on 3D protein structures, a professional curation tool, support for crowd contributions, and interactive access to genomic read alignment for assessing variant call quality. Many analytic operations are supported: genetic association, trios, pathways, somatic oncology, and the definition of gene and variant panels for automated calls. **Conclusions.** We present a novel tool for extremely efficient and interactive access of genomic, knowledge and phenotypic data. It allows real-time free text or structured queries covering the main information needs in genome interpretation.

1942W

Population genotyping from low-coverage sequencing data using a reference panel. *L. Huang[1], P. Danecek[2], S. Batzoglou[1].* 1) Stanford University, Stanford, CA; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

In the past years, several large-scale whole-genome sequencing projects are completed, such as the 1000 Genomes Project, the UK10K Cohorts Project. These projects provide high-quality haplotypes and/or genotypes of a considerably large number of whole genomes. Besides its value in the original projects, this resource has the potential to benefit the quality of genotype calling from low-coverage sequencing data in future sequencing projects, because these existing genotype calls effectively captures the linkage disequilibrium structures of cohorts. Here we present reference-based Reveel (or Ref-Reveel in short), a novel method for population genotyping. Ref-Reveel is designed based on Reveel, our first generation of population genotyper, which has been demonstrated to be an effective tool. Ref-Reveel incorporates reference genotypes from completed projects to boost the genotyping quality of new data sets while maintaining high computational efficiency.   We demonstrate that using a reference panel substantially improves genotyping accuracy through extensive experiments on the whole genome. We also compare the performance of using three reference panels when we call genotypes of 99 CEU samples from the 1000 Genomes Project: (i) the genotypes of 3,781 UK10K samples provided by the UK10K Consortium, (ii) the Reveel-called genotypes of the UK10K samples, and (iii) the integrated call set of 2,535 samples from the 1000 Genomes Project Phase 3. We show reference panel (ii) outperforms the other panels. Critical characteristics of reference panels are then discussed based on our observations: reference panel size, ancestry of samples, genotype quality, and completeness of variants. In our experiment, the quality of two UK10K reference panels are evaluated by using the genotypes called from the high read-depth exome sequencing of the TwinsUK cohort were used as a truth set. Finally, we demonstrate the effectiveness of Ref-Reveel on an even more challenging population genotyping problem: to call genotypes of a small data set with merely ten samples from unknown populations. A public available reference panel, the 1000 Genomes Project Phase 3, is used in this experiment.

1943T

Automated, scalable quality control of heterogeneous exome sequence data. *R. Koesterer, M. von Grothuss, J. Flannick on behalf of the GoT2D, T2D-GENES, SIGMA, LuCAMP, and ESP consortia.* Medical and Population Genetics, Broad Institute, Cambridge, MA.

In the coming years, hundreds of thousands of samples will be sequenced to identify associations between genetic variation and complex phenotypes. To avoid false-positive findings but maintain power, stringent quality control (QC) procedures must be applied to these data. Current QC approaches require the direct attention of analysts, presenting scaling challenges to millions of samples.We developed a novel approach for sample QC, combining control for ancestry and other confounders with automated outlier detection. Our approach (a) projects genotypes onto principal components (PCs) of known reference populations; (b) computes a battery of metrics assessing sequence quality; (c) normalizes metrics according to ancestry and other confounders such as sequencing platform; (d) performs PCA on the normalized metrics to identify independent measures of sample quality; and (e) uses Gaussian Mixture Modeling to identify outliers according to both the individual metrics as well as their PCs. We evaluated our method relative to previously published manual approaches using 18 sample metrics computed across 12 ancestry groups from 26,171 exomes. Sample quality was evaluated based on the transition-transversion ratio (TITV) of singleton variants unique to samples that passed QC (QC+) relative to those that failed (QC-).We identified 336 QC-samples, within which singleton variants had a TITV ratio of 1.75 compared to a ratio of 2.66 for QC+ samples (p=1.63e-129), suggesting QC- samples as significantly enriched for sequencing artifacts. Among them, 210 problematic samples with a singleton TITV ratio of 1.81 (p=2.8e-76) were not identified by the current gold standard approach.Samples substantially skewed on any of the metrics, including low call rate, high number of alternate alleles, and fraction of alternate reads at heterozygous sites (allele balance), were enriched for sequencing artifacts. Metrics based on lower frequency variants were more effective than comparable metrics based on common variants. PCA of normalized metrics identified six additional outliers with a mean singleton TITV ratio of 2 (p=6.8e-3), suggesting samples skewed across multiple metrics are also enriched for sequencing artifacts.These results demonstrate that not only can automatic, scalable QC approaches be developed for heterogeneous exome sequencing data, but their performance can meet or exceed that of current gold standard approaches.

1944F

New RAREMETAL: An efficient and flexible tool for meta-analysis. *S. Chen[1,2], J. Yang[1], D. Liu[3], S. Feng[4], G. Abecasis[1]. 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Department of Bioinformatics, University of Michigan, Ann Arbor MI; 3) Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA; 4) Hitachi America Big Data Lab, Santa Clara, CA.*

Meta-analysis plays a critical role in large-scale genetic association studies. It provides researchers a convenient computational tool for achieving large sample sizes. Here, we report on several useful improvements to our RAREMETAL software for rare-variant meta-analysis. Specifically, we have: (1) updated score statistics for meta-analysis so that they specifically account for varying case-control ratios among the studies being analyzed, resulting in substantial power gains; (2) improved meta-analysis methods for dichotomous traits by allowing for logistic regression models in the analysis of component studies; (3) optimized workflows for large-scale datasets to enable analyses that include 10s of millions of variants across 100,000s individuals; (4) added SKAT-O and additional weighting schemes for gene-based rare variant association tests. We hope our new RAREMETAL software will continue to be a useful tool to facilitate the discovery of disease-associated genes in large studies.

1945W

Discovery and meta-analyses of trans-eQTLs from cross-tissue gene expression data using approaches that increase statistical power. *B. Jo[1], Y. He[2], I. McDowell[3], A. He[2], A. Battle[2], B. Engelhardt[4], The GTEx Consortium. 1) Department of Quantitative and Computational Biology, Princeton University, Princeton, NJ; 2) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 3) Program in Computational Biology and Bioinformatics, Duke University, Durham, NC; 4) Department of Computer Science and Center for Statistics and Machine Learning, Princeton University, Princeton, NJ.*

Expression quantitative trait locus (eQTL) studies seek to identify the genetic loci that regulate gene expression levels, helping to characterize the role that non-coding genetic variants play in the regulation of complex traits. As RNA sequencing (RNA-seq) experiments have advanced to the degree that we are able to obtain multi-tissue expression profiles per individual in a robust manner, so has our ability to understand context-specific roles that genetic variants play in expression profiles. In particular, the NIH Common Fund's Genotype-Tissue Expression (GTEx) project provides a rich collection of RNA-seq experiments that span 449 individuals over 44 different tissue types. This study expands upon the previous cis-expression QTL (eQTL) discoveries in GTEx by establishing strong evidence for trans-, or distal, eQTLs. Prior eQTL studies mainly focus on cis-eQTLs, as trans-eQTLs are difficult to detect because of the high multiple hypothesis testing burden, and do not often replicate because of smaller effect sizes and greater levels of tissue- and sex-specificity relative to cis-eQTLs. In this work, we describe 44 trans-eQTLs identified in the GTEx v6 data using standard association testing in each tissue. Then, we demonstrate improvements in these results by using various approaches that seek to increase power in trans-eQTL detection, including making use of known cis-eQTLs to reduce the number of SNPs tested, and leveraging known genetic networks and trait-associated genetic variants from GWAS studies. We also apply Meta-Tissue, a linear mixed model approach to meta-analysis, for identifying putative trans-eQTLs that are shared across tissues, improving power and providing estimates of tissue-specificity. We validate our trans-eQTLs using replication in other large-scale eQTL studies (DGN, MUTHER, and GEUVADIS), functional analysis in enrichment of regulatory regions, and comparisons to effect-size confidence intervals established via Kallisto, a fast expression quantification method with bootstrapping.

1946T

ASElux: An ultra-fast and accurate allelic read aligner. *Z. Miao[1,2], A. Ko[1,3], M. Alvarez[1], M. Laakso[4], P. Pajukanta[1,2,3].* 1) Department of Human Genetics, UCLA, Los Angeles, Los Angeles, CA; 2) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA; 3) Molecular Biology Institute at UCLA, Los Angeles, CA; 4) University of Eastern Finland, Kuopio, Finland.

A large proportion of genes display allele specific expression (ASE). Accurate assessment of ASE can help elucidate biological mechanisms related to disease via imprinted genes or cis-regulation of expression. Since reads from the alternative allele are more likely to be falsely mapped, mapping bias remains a major obstacle in analyzing allelic expression. Previous studies have used simulation to mark the potentially biased alignment regions, however, as ~30% of SNPs are identified as biased SNPs, a large number of SNP sites are excluded from the downstream analysis and the remaining SNPs still display some degree of mapping bias. Other methods such as SNP-o-matic and GSNAP integrate the SNP information with the reference genome to perform a SNP-tolerant alignment. However, both SNP-o-matic and GSNAP are relatively slow, making them less ideal for the analysis of large data sets. To perform a fast and accurate alignment for ASE analysis, we developed a new method, ASElux that focuses on a SNP-tolerant alignment of allele-specific reads. Since the allele-specific reads comprise only ~10% of the RNA-seq data, we save a significant amount of time by ignoring the non-allele specific reads during the alignment. With ASElux, we use personal SNP information to generate all possible allele-specific reads, followed by their direct counting during the alignment. This design makes ASElux 75 times faster than GSNAP (440 queries/s) and ~2 times faster than STAR (14,286 queries/s) and HISAT2 (16,666 queries/s). We also tested the accuracy of ASElux along with other tools both on simulated RNA-seq data and 20 adipose RNA-seq samples from the METSIM study. Using GSNAP as the golden standard, ASElux shows a significantly smaller ($p<2.2\times10^{-16}$) reference bias than STAR and HISAT2. In identifying ASE SNPs, ASElux also displays a better concordance with GSNAP than STAR and HISAT2. We further built a modified version of ASElux for aligning DNA-seq data and tested it with other tools in a simulated ChIP-seq data set. ASElux shows a top speed and accuracy when compared with GSNAP and Bowtie. A total of 99.7% of SNPs aligned by ASElux showed no bias, exceeding the performance of GSNAP (88.5%) and Bowtie (87.5%). Taken together, ASElux provides a great accuracy and ultra-high speed for allele-specific read alignment, making it a powerful new tool for large scale allele-specific analysis of not only gene expression but also transcription factor and histone modification data.

1947F

Genomewide allele-specific expression (ASE) analysis of RNA sequencing and exome array data in the COPDGene Study. *M.M. Parker[1], R.P. Chase[1], A. Lamb[1], B.D. Hobbs[1,2], M.H. Cho[1,2], E.K. Silverman[1,2], C.P. Hersh[1,2], P.J. Castaldi[1,3].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA.

Background: Allele-specific expression (ASE) analysis quantifies the variation in gene expression observed between the two haplotypes of an individual and can be used in analysis of cis regulatory variant effects, nonsense-mediated decay and imprinting. **Methods:** To assess genomewide ASE, we combined Illumina exome array and whole-blood RNA sequencing data from 127 individuals in the COPDGene study. For subjects passing standard quality control, allelic counts at heterozygous sites were quantified using GATK's ASEReadCounter. To minimize allelic mapping bias, we excluded non-uniquely mapping reads, reads with a mapping quality < 10 and variant sites shown in simulation to change the allelic ratio > 5%. To define significant ASE, we performed binomial tests across all heterozygous sites with a minimum of 10 reads with the null value equal to the mean reference ratio. Variant sites were annotated with gene name and functional consequence using GENCODE v19. **Results:** A total of 131,740 heterozygous sites were tested for ASE (median coverage = 24x, IQR coverage = 30). The mean reference ratio over all sites was 0.53, indicating residual allelic mapping bias, which was strongest for low depth sites (reference ratio at 10-20x coverage = 0.534 vs. reference ratio at > 100x coverage = 0.523, p-value difference in mean < 1 $\times10^{-6}$). At an FDR of 5%, 4,623 sites (in 1,358 unique genes) showed significant ASE, with 972 sites showing complete ASE. The majority (70.84%) of the 4,623 sites showed ASE in > 1 subject (range = 2-67 subjects), and there were 563 genes with >1 instance of significant ASE. Stopgain/stoploss variants were significantly more likely to show allelic imbalance as compared to nonsynonymous or synonymous (p-value < 6x10^{-10}, proportion of stop variants with ASE = 0.18, proportion of synonymous = 0.047, proportion of nonsynonymous = 0.037). To assess if ASE effects were consistent across individuals, we subset heterozygous sites to those covered at > 40x and observed in > 10 individuals (n=482). We identified examples of both: 1) sites with ubiquitous ASE (n=8, max number subjects with ASE =51), and 2) sites with ASE in only a subset of individuals (n=474, mean proportion with ASE = 0.13). **Conclusions:** Our findings from ASE analysis in 127 subjects from the COPDGene study: 1) document extensive evidence of ASE in whole blood; 2) confirm enrichment of ASE in stop variants; and 3) provide evidence for both ubiquitous and person-specific ASE.

1948W

Seqant 2.0: Whole-genome variant annotation in the Amazon cloud. *A. Kotlar[1], M. Zwick[1], D. Cutler[1], T. Wingo[2].* 1) Human Genetics, Emory University, Atlanta, GA; 2) Neurology and Human Genetics, Emory University, Atlanta GA.

As the cost of sequencing drops, whole genome and whole-exome experiments are increasingly common in labs with a diversity of computational experience. In these environments there is a need for cross-platform, easy-to-use computer programs that identify and analyze alleles relevant to researchers' phenotypes of interest. SeqAnt 2.0 is the first web-based program of this kind, called a variant annotator, that is backed by the Amazon AWS cloud. It runs on any modern browser without the need for plugins, and is capable of analyzing more than 1.5 million variants per minute for experiments in excess of 40 gigabytes in size. Its web interface is a JavaScript single-page application (SPA) that renders at 60 frames per second, achieving the simplicity and fluidity of offline applications. It features a minimalist interface that walks researchers through three simple submission steps. Once submitted, jobs are analyzed by a Perl multi-threaded annotator running in Amazon's data centers, which sends real-time progress updates with only 2 bytes overhead. Once completed, results are visualized in the browser using WebGL and JavaScript. Users can filter their results in the browser or download the full experiment for offline analysis. Furthermore, users are given accounts through which they can store their analytic steps and results in a HIPAA-compliant data center, which removes the need to worry about hard drive space limitations. SeqAnt currently serves *Homo sapiens*, *Mus musculus*, and *Drosophila melanogaster* annotations, and can be extended to support any additional eukaryotic model organisms. In conclusion, Seqant 2.0 is the first web-based variant annotator that leverages the Amazon cloud computing platform. It runs on any modern browser and on any recent computer or mobile device. Its performance and simplicity enables researchers with even limited computational experience to rapidly analyze alleles found in their whole genome and whole-exome experiments.

1949T

Decoupling pervasive transcription and real transcription reveals transposable element activity during the development of the human brain. *F. Navarro, M. Gerstein.* MB&B, Yale University, New Haven, CT.

Transposable elements (TE) constitute approximately half of the human genome. Despite the difficulties created by their highly repetitive nature, there is building evidence that transposable sequences play a significant role in the genome biology influencing gene regulation and creating variability across individuals and species. TEs show somatic activity in pathogenic and healthy cells. Among other tissues, the human brain is thought to harbor somatic transposable element somatic. Due to their high copy number and broad distribution across the genome, assessing transposable elements expression is specially affected by RNA fragments originating from pervasive transcription. We developed a new approach that deconvolutes pervasive transcription signal from real transcription and estimates TE subfamily expression level. We used RNA-seq experiments from ENCODE, BrainSpan, GTex and PsychENCODE to evaluate the activity pattern of LINE1s, SVAs and HERVs in 16 regions of the human brain across 12 developmental stages and different conditions. We observe that the majority of the reads mapped to transposable elements are due to pervasive transcription and removing this signal results in recent, and potentially active, TE subfamilies being transcribed in the human brain. Using ENCODE data, we find that pervasive transcription signal is generated in nuclear transcripts while the cytoplasm contains real transcription of L1Hs. Moreover, we found that specific brain regions harbor higher activity of L1HS and LTRs and no activity of SVAs. All together our results demonstrate that different but mostly recent TEs are active during the whole development of the human brain and might impact tissue biology.

1950F

FusorSV: A unified knowledge based framework for comprehensive Structural Variant discovery. *T. Becker[1,2], J. Leone[1], Q. Zhu[1], E. Cerveira[1], M. Romanovitch[1], C. Zhang[1], D. Shin[1,2], C. Lee[1], A. Malhotra[1].* 1) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Computer Science and Engineering, University of Connecticut, Storrs, CT.

Calling Structural Variants (SV) comprehensively and with high accuracy remains a challenge. Phase 3 of the 1000 Genomes SV project made 68,818 calls from nine different algorithms and integrated them into a single call set across 2,504 individuals from 26 human populations. Methods have since been developed that combine multiple algorithms under consensus or by counting the supporting reads for each algorithm to obtain a higher quality SV call set than any single algorithm on its own. However, the existing methods do not utilize discriminating SV features or prior SV information from projects such as the 1000 Genomes.We present a state of the art data fusion method driven by SV features and prior SV knowledge to produce a comprehensive and unified call set from an arbitrary number of SV callers. The method characterizes all possible input combinations using prior SV information as a true call set for each feature and then leverages that information in the discovery step to produce an accurate genotyped result. Each output SV region is annotated with a confidence measure from the model, information about the contributing input calls and when available, the aligned targeted assembly. We tested the approach on deletion, duplication and inversion SV call types by constructing a fusion model using ten of the most popular SV callers (including some ensemble callers) with the 1000 genomes Phase 3 data set that comprises 27 PCR-Free samples each with 250bp pair-end reads at 50X coverage. Using SV type and SV length as the discriminating features, fusorSV outperformed existing algorithms based on a 3-fold cross validation with 1000 permutations. We also performed (1) in silico validation by targeted breakpoint assembly of the calls and (2) experimental validations on a randomly selected subset of novel (not present in 1000 genomes Phase 3) deletions, duplications and inversion calls.

1951W

Analytical approaches for sequence variant interpretation in clinical exome analysis. *D.J. Corsmeier[1], G.E. Herman[2,4], R.E. Pyatt[1,3], J. Gastier-Foster[1,3,4], C. Astbury[1,3].* 1) Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital; 2) Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital; 3) Department of Pathology, The Ohio State University College of Medicine; 4) Department of Pediatrics, The Ohio State University College of Medicine.

With the rapid adoption of whole exome sequencing into clinical practice, the volume of genomic sequence variants that must be analyzed and interpreted accurately is growing exponentially. The 2015 ACMG Guidelines are an invaluable resource for a systematic approach to the interpretation of these variants. The 28 basic evidence classifiers give specific criteria by which to approach the classification and prioritization of genomic variation. While nearly all of the 12 classifiers supporting benign interpretation and the 16 that may support calling a pathogenic finding are based on objective information, approximately half of them require more than a cursory manual effort to evaluate. We demonstrate that by performing some preliminary bioinformatic analysis, in particular the mining of databases that include known variant pathogenicity information such as ClinVar and HGMD, we can eliminate much of this manual effort. For instance, in the case of the very strong evidence of pathogenicity predicted by a null variant for a disease in which loss of function is a known mechanism (PVS1), we have precompiled a list of all known genes associated with such diseases, making retrieval trivial. Evaluation of a given variant then becomes a matter of simply inspecting the variant type (e.g., nonsense, frameshift) to determine whether the associated predicted gene product would be null. With the more complex classifiers, we utilize more involved computational methods. For example, we have compiled lists of repeat regions to facilitate evaluation of the criteria assessing in-frame indels (BP3). We have also analyzed the variant databases to establish reasonable thresholds, omitted from the Guidelines, to assess rates of pathogenic and benign variation (PP2, PM1). Through this type of preprocessing, we effectively minimize the manual effort necessary to evaluate sequence variants on the basis of the evidence criteria. Through use of a more automated approach, variant interpretation can be performed more accurately, efficiently, and with less human error and subjectivity. This is a necessary requirement with the rapid growth of clinical exome sequencing. Our goal is to minimize manual effort in the evaluation of pathogenicity evidence. This will facilitate sequence variant interpretation and demonstrate how the variant interpretation process might be further standardized in the application of the ACMG Guidelines in clinical exome analysis and interpretation.

1952T

Comparison of variant analysis pipelines for genetic diagnosis of undiagnosed diseases. *A. Dries[1,2], J. Kohler[1,2], D. Waggott[1,2], P. Batra[1,2], C. Prybol[1], E. Fernandez[1,2], D. Zastrow[1,2], D. Fish[3], M. Grove[2,3], P. Zornio[1,2], U.D.N. Members[4], E. Ashley[1,2,8], P. Fisher[5], J. Bernstein[6], J. Merker[7], M. Wheeler[1,2].* 1) Stanford Center for Undiagnosed Diseases, Stanford, California; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, California; 3) Stanford Center for Genomics and Personalized Medicine, Stanford, California; 4) Undiagnosed Diseases Network, National Institutes of Health, Bethesda, Maryland; 5) Department of Neurology, Stanford University, Stanford, California; 6) Department of Pediatrics, Stanford University, Stanford, California; 7) Department of Pathology, Stanford University, Stanford, California; 8) Department of Genetics, Stanford University, Stanford, California.

For participants evaluated by the Undiagnosed Diseases Network (UDN), novel strategies analyzing genetic data are needed in the search for a genetic diagnosis. While clinical exome and genome laboratories are necessarily limited in the curation of novel genes and variants potentially causative of disease, the clinical sites of the UDN have methods to identify clinically relevant variants in rare undiagnosed diseases.   At the Center for Undiagnosed Diseases at Stanford (CUD), patients undergo whole exome or whole genome sequencing through the UDN sequencing cores. Samples are run with family member DNA in a trio or quad. In addition to a clinical report, cores return raw sequencing data to UDN clinical sites for genetic analysis and curation. The analysis team uses two parallel pipelines to identify variants of interest: Ingenuity Variant Analysis, a commercial phenotype driven tool that uses customizable filters to refine variants; and sequence to medical phenotypes (STMP), a bespoke and customizable curation pipeline that prioritizes variants using multiple levels of evidence integrating a broad array of metadata. STMP Tier 0 and Tier 1 variants are manually curated and represent rare ClinVar pathogenic or likely pathogenic variants and rare loss of function variants, respectively. Variants in tiers 2 through 4 are analyzed through a phenotype-driven approach. Variants from each pipeline are assessed by a standardized process of manual curation. In five exome cases using four trio analyses and one quad analysis, Ingenuity filters narrowed an average of 312,347 variants per proband to 115 variants moved to manual curation. Average outputs per filter showed: call quality – 181322; allele frequency – 28047; variant type – 1057; phenotype-driven gene list – 372; and inheritance pattern – 38. For the same cases, STMP analysis averaged 3 Tier 0 and 74 Tier 1 variants.   Clinical reports from the sequencing cores have revealed 4 pathogenic variants returned to 2 UDN patients. In 11 total cases, the CUD manual curation has resulted in 13 candidate variants being sent for further evaluation by the UDN Model Organism Screening Core (9), by functional or splicing assays (2) or by a novel mosaic pipeline (2). In its first 6 months, the CUD curation pipeline has resulted in 4 presumptive diagnoses. The CUD sequencing pipelines for genetic analysis are valuable for identifying potential genetic diagnoses in patients with rare and undiagnosed disease.   .

1953F

Scaling variant calling to hundreds of thousands of samples with GATK. *L.D. Gauthier[1], D. Benjamin[1], K. Bergin[1], J.M. Bloom[1], Y. Farjoun[1], M. Fleharty[1], S. Lee[1], M. Lek[1,2], H. Li[1], V. Ruano-Rubio[1], T. Sato[1], M. Shand[1], E. Banks[1], D. MacArthur[1,2].* 1) The Broad Institute of Harvard and MIT, Cambridge, MA; 2) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

The modern age of high-throughput sequencing has enabled sequencing studies of unprecedented numbers of samples. Whether designed to increase power in GWAS studies or establish allele frequencies in healthy populations, these projects require careful control of batch effects that can arise from non-uniform bioinformatic processing. GATK's joint calling workflow is one approach to achieving that necessary consistency of variant calling and filtration. Joint calling further benefits studies of any design by producing genotype and confidence information for every sample at every genomic position called in any sample, clearly differentiating between lack of coverage and lack of alternate reads. While the runtimes of other variant callers scale exponentially with the number of samples, the modern HaplotypeCaller algorithm in GATK enables a more efficient linear runtime increase. This implementation underlies the Exome Aggregation Consortium releases, with the most recent exceeding 100,000 samples. Using a 50/50 mixture of two deeply characterized haploid cell line samples as our control data set, we performed variant calling and filtration on cohorts of increasing size from a few dozen samples to over 100,000. We are able to show that variant calling accuracy improves as cohort size is increased up to a few hundred samples. As cohort sizes reach thousands, tens of thousands or even hundreds of thousands of samples, variant calling accuracy is preserved for both SNPs and INDELs over the samples in the intersection of those call-sets. Furthermore, we found that sensitivity to singleton variants in the well-characterized sample is preserved as additional samples are added to the cohort. These results can be used to inform approaches for variant calling in large sequencing studies and particularly highlight the need to call all samples together in order to minimize batch effects and harmonize variant representation. Based on our results, in order to achieve maximum variant calling accuracy for smaller cohorts while minimizing bioinformatic runtime, we recommend joint calling with a panel of reference samples such that the total cohort size exceeds 200.

1954W

SCOTCH: A novel method to detect insertions and deletions from next-generation DNA sequencing data. *R.L. Goldfeder[1], E.A. Ashley[2].* 1) Biomedical Informatics, Stanford University, Stanford, CA; 2) Department of Medicine, Stanford University, Stanford, CA.

Clinical-grade genome sequencing and interpretation requires accurate and complete genotype calls for all interrogated positions. While single nucleotide variant detection is highly accurate and consistent, these variants explain only a small fraction of disease risk. Other types of variation that disrupt the open reading frame, such as insertions and deletions (INDELs), have systematically been shown to have dramatic effects on phenotype. However, current methods have low sensitivity for larger INDELs (>= five bases), primarily due to challenges surrounding aligning sequence reads that span complex loci. We present SCOTCH, a novel INDEL detection method that leverages signatures of poor read alignment, through machine learning approaches, to accurately identify INDELs from next-generation DNA sequencing data. Using biologically realistic simulated genomes and sequence reads with technologically representative error profiles (generated by ART), we evaluate SCOTCH and several currently available INDEL callers. We show that SCOTCH outperforms current methods, particularly for larger INDELs and INDELs within repetitive or low complexity regions. Finally, we demonstrate the clinical utility of SCOTCH through the identification of a large number of high quality novel INDELs within patients enrolled in the Stanford Clinical Genome Service and Stanford Undiagnosed Disease Network. This method will enable researchers and clinicians to more accurately identify INDELs associated with previously unexplained genetic conditions.

1955T

A random forest model for identifying LCL-derived mutations enabling the use of LCL DNA in rare de novo variant studies. *D. Kashef-Haghighi[1,4], L. Perez-Cano[2,4], E. K. Ruzzo[2,4], J.Y. Jung[1], L.K. Wang[2], S. Sharma[1], M. Duda[1], G. M. McInnes[1], J. K. Lowe[2], D. H. Geschwind[2], D. P. Wall[1,3].* 1) Division of Systems Medicine, Department of Pediatrics, Stanford University, Stanford, California, USA; 2) Center for Neurobehavioral Genetics and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, California, USA; 3) Departments of Psychiatry and Biomedical Data Science, Stanford University, Stanford, California, USA; 4) These authors contributed equally to this work.

With recent technical advancements, next-generation sequencing has become a powerful and cost-effective tool for the discovery of rare disease-causing variants. But often times the difficulty of obtaining sufficient amounts of DNA for library preparation and validation limits the scope of such studies. Lymphoblastoid cell lines (LCL) have been introduced as renewable sources of DNA to be used in human molecular studies. These immortalized cell lines are developed by infecting the peripheral blood lymphocytes from a donor using Epstein Barr virus (EBV). Although the ease of their generation and cultivation combined with the resulting unlimited DNA supply from the donor renders LCLs suitable for next-generation sequencing, their genomic instability poses a new challenge for the analysis of the resulting data. During prolonged culture, LCL cells accumulate new genetic changes that can appear as *de novo* mutations when sequence reads from parents and the offspring are compared. These LCL-specific genetic mutations are not easily distinguishable from true *de novo* variants and will confound the downstream analysis of genotype-phenotype association. As a result, *de novo* variants are often excluded from analysis in LCL studies unless they can be experimentally validated in the matched whole-blood DNA, which is expensive and not always feasible. In this work, we present a novel machine learning approach for distinguishing true *de novo* variants from LCL-specific genetic aberrations or other sources of artifacts such as sequencing and mapping errors. As part of the Hartwell Autism Research and Technology Initiative (iHART), we generated high coverage (~30x) whole-genome sequence data for 75 pairs of LCL monozygotic (MZ) twins that we have used to train our model. For the training, *de novo* mutations identified in both MZ twins were considered as true events, whereas discrepant calls were assumed to be either LCL or sequencing artifacts. Additionally, we also sequenced the whole-genome of 19 pairs of matched whole blood and LCL samples to serve as an independent test set for the evaluation of our final model. We have used a diverse collection of features including various variant and genome properties, and also annotations from the "Genome in a Bottle" project to train a random forest model achieving an area under the ROC curve of 0.96. We anticipate that this work will open up a practical avenue for utilizing LCL-derived DNA to identify rare genetic variants.

1956F

Normalizing genetic variation data from more than 100,000 human samples. *G. Saunders, C.Y. Gonzalez, J. Kandasamy, J.M. Mut, P. Arce, T. Smith, D. Spalding, H. Parkinson.* Variation Archives, European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom.

A number of high-profile human genetic variation datasets are publically available e.g. 1000 Genomes, ExAC, deCODE and UK10K. However, these datasets are often prepared using independent methodologies or processing pipelines and can be accessed using only custom websites or FTP servers. We present the European Variation Archive (EVA) that acquires, normalizes and re-annotates variation data from the large datasets described above and datasets that are submitted directly to us. As the EVA variant collection includes all dbSNP variants and also consumes data from a number of other sources (e.g. ClinVar) our resource represents the largest collection of open-access genetic variation data available, worldwide. In total the EVA contains data from 57 studies, describing ca.154 million unique variants from more than 100,000 human samples. Access to these data is possible via the EVA website (www.ebi.ac.uk/eva) or API (http://www.ebi.ac.uk/eva/?API), where we offer the ability to search and/or download at the database, study or custom query level. The ability to download custom datasets is the single most powerful aspect of the EVA as this allows any user the ability to easily generate a single dataset, in a number of different file formats, which best addresses their research question (e.g. 'all missense variants within the BRCA2 gene with an allele frequency greater than 0.2'). From a technical perspective, our systems are based in modern and widely used technologies and all source code is completely open-access. Variation data are stored in a NoSQL distributed database and shared via a REST web services API, including a layer compatible with the Global Alliance for Genomics Health (GA4GH) API specification. Additionally, the website that consumes and displays these APIs has been developed using HTML5 and the ExtJS 6 framework for JavaScript. We shall also describe how our efforts are aligned with other initiatives such as Ensembl (http://www.ensembl.org/) and Open Targets (previously known as the Centre for Therapeutic Target Validation; https://www.opentargets.org/) and how the EVA serves as the data source for these resources thus providing examples of our long-term strategic plan where the EVA functions as a central genetic variation warehouse for the human genetics community.

1957W

Integrated transcriptome and methylome analyses identify novel genes involved in progression of non-alcoholic fatty liver disease (NAFLD) among Indians. *A. Chatterjee[1], K. Das[2], A. Basu[1], P. Singh[2], B. Dey[3], S. Roychowdhury[3], P. Pandit[3], B. Bhattacharya[3], N.P. Bhattacharyya[3], P.P. Majumder[1], A. Chowdhury[2,3], P. Basu[1,3].* 1) National Institue of Biomedical Genomics, Kalyani, West Bengal, India; 2) SDLD, IPGMER and SSKM Hospital, Kolkata, West Bengal, India; 3) Biomedical Genomics Centre, Kolkata, West Bengal, India.

Non-alcoholic Fatty Liver Disease (NAFLD), characterized by deposition of fat in liver, is a burgeoning health problem worldwide. Epidemiological studies suggest the prevalence of NAFLD in India is highly variable and ranges between 17% to 32%. The initial stage is Simple Steatosis (SS), characterized by the deposition of triglycerides (TG) in hepatocytes. SS is asymptomatic and selectively progresses to Nonalcoholic Steatohepatitis (NASH), characterized by the presence of hepatocyte injury and inflammation. NASH, in turn, is highly susceptible to end-stage liver disease. The molecular mechanisms underlying the heterogeneous outcomes of NAFLD remain vastly unclear. In this study, we aimed to identify the difference in transcriptome between the mild and advanced stages of NAFLD spectrum and its associated changes in DNA methylome. Whole transcriptome and DNA methylome data were generated on 25 mild and 20 advanced NAFLD liver biopsy tissue samples. 28 Differentially Expressed Genes (DEGs) were observed between the mild and advanced NAFLD. Among them, the most enriched pathway was found to be PI3-AKT-p70S6K pathway, which is involved in normal insulin functioning and its impairment cause insulin resistance. We also observed increased expression of fatty acid binding proteins (FABP4, FABP5, and FABP5L2) among advanced NAFLD, suggesting increased hepatic FFA uptake. Among the other DEGs, we observed upregulation of genes involved in protein synthesis regulation (ERAL1, RBM4B), cell cycle regulation (PRAP1, ZFP36), cellular structural integrity maintenance (KRT8, ITFG2, ATP6V1F) and response to oxidative stress and apoptosis (TMEM214, URM1), which provides evidence of deregulation in these pathways quite early in the disease spectrum. We also observed subtle differential methylation between the study groups. Significant negative correlation between promoter methylation and gene expression was observed for ATP6V1F and PRAP1.

1958T

Moving beyond GWAS: Characterizing the function of pubertal timing genes in early vertebrate development. *J.T Leinonen[1], Y.C Chen[2], H. Koivula[2], P. Panula[2], E. Widén[1].* 1) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland; 2) Neuroscience Center and Institute of Biomedicine, Anatomy, University of Helsinki, Finland.

Puberty represents a crucial developmental stage, the timing of which varies considerably between individuals. Genome-wide association studies have identified several genetic loci associating with pubertal timing. Some of these loci implicate evolutionarily conserved genes which are expressed already during the embryonic stages. Normal puberty requires appropriate fetal development, including the formation of the hypothalamus and gonadotrophin releasing neurons. Therefore, we hypothesized that some of the pubertal timing loci may interfere with these early developmental processes. Using zebrafish as a model, we aim to study the function of selected pubertal timing-associated genes including *lin28b,* establishing the spatio-temporal gene expression patterns in embryos, and assessing the impact of their downregulation and upregulation. We model gene downregulation using morpholino-antisense and CRISPR-Cas9 technologies, and upregulation by synthetic mRNA injections. The specificity of the morpholino-induced phenotypes is verified by mRNA-rescue. We assess the potential impacts on fish morphology and RNA expression both in embryos and adult fish. While the expression of *lin28b* in wild type fish is spatially unrestricted during the first 24 hours, and subsequently restricted to organ primordia including regions of the brain, our results suggest that transient alteration of *lin28b* expression in embryogenesis leads to profound alterations in fish development. Both mRNA-injections and morpholino-induced downregulation result in dose-dependent lethality and malformations visible already during 24hpf. However, our data based on non-lethal doses of morpholinos suggest that *lin28b* may also have a specific impact on genes regulating hypothalamic functions. Whereas the injection of non-lethal doses of *lin28b* morpholino had no effect on *gnrh3*–expression at the larval stages, *lin28b*-morphants showed reduced expression of *kiss2* in qPCR analysis (p<0.05). Moreover, our preliminary data show that fish injected with a dose of *lin28b* mRNA causing no discernible phenotypic effects in the embryos had significantly greater body length as adults compared to controls (p<0.005). These results suggest that even transient dysregulation of a puberty-associated gene during the embryogenesis may result in sustained alterations in growth during later development. We currently seek to assess how permanent *lin28b* knock-out based on our CRISPR-Cas9 model may impact on development and growth.

1959F

Epigenome-wide DNA methylation and non-alcoholic fatty liver disease in adolescence. *P.E. Melton[1], K.A. Lillycrop[2,3], K.M. Godfrey[3,4], D. Anderson[5], G.C. Burdge[6], M.S. Kobor[7], T.A. Mori[8], L.J. Beilin[8], J.M. Craig[9], J.K. Olynyk[10], J.D. Holbrook[11], C.E. Pennell[12], W.H. Oddy[13], E.K. Moses[1], L.A. Adams[8], R.C. Huang[8].* 1) The Curtin UWA Centre for Genetic Origins of Health and Disease, Faculty of Health Sciences, Curtin University and Faculty of Medicine, Dentistry & Health Sciences, The University of Western Australia; 2) Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton, UK; 3) NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK; 4) MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK; 5) Telethon Kids Institute, University of Western Australia, Perth, Australia; 6) Academic Unit of Human Health and Development, University of Southampton; 7) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia , Vancouver, BC, Canada; 8) School of Medicine and Pharmacology, University of Western Australia Perth, Australia; 9) Environmental and Genetic Epidemiology, Murdoch Childrens Research Institute, Melbourne Australia; 10) Translational Medicine Head of Gastroenterology & Hepatology, Fiona Stanley and Fremantle Hospitals Western Australia; 11) Singapore Institute for Clinical Sciences (SICS), Agency for Science Technology and Research (A*STAR), Singapore; 12) School of Women's and Infant Health, University of Western Australia, Perth, Australia; 13) Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia.

Non-alcoholic fatty liver disease (NAFLD) is closely linked to adiposity and insulin resistance. NAFLD itself is increasingly recognized as causing a substantial global health burden. Epigenetic modifiers are known to be involved in biological pathways involving lipid metabolism, insulin resistance, oxidative stress, and inflammation, which can induce hepatic fat accumulation and NAFLD. To investigate the association between DNA methylation and NAFLD in adolescence, we used the Illumina 450K Bead Chip to perform an epigenome-wide association study (EWAS) on DNA extracted from peripheral blood collected at age 17 years in 852 Western Australia Pregnancy Cohort (RAINE) participants. NAFLD was assessed using ultrasound as a quantitative trait with Hamaguchi's steatosis score, ranging from 0 (no NAFLD) to 6 (moderate to severe NAFLD), combined with daily alcohol consumption <10g for females and <20g for males. The association of NAFLD with genome-wide BMIQ (Beta Mixture Quantile dilation) normalized CpG probes was tested using linear mixed effects models for quantitative trait values and penalized logistic regression for binary values. Covariates included age, sex, ancestry principal components from GWAS and batch effects; independent models adjusting for either measured white cell count or estimated Houseman cell counts (CD8T, CD4T, NK, B cells, monocytes, granulocytes). Using a conservative Bonferroni correction (p=1.00E-07) for multiple testing we identified a genome-wide significant DNA methylation CpG probe for quantitative NAFLD (cg10144473, p=8.94E-09) located in a CpG island near the gene, *SLC2A1*. Genetic variants in this *SLC2A1* have previously been associated with NAFLD, and decreased expression of this gene has been shown in NAFLD liver biopsies. In addition, in vitro down regulation of siRNA *SLC2A1* using THEL2 cells has been shown to promote lipid accumulation and increased oxidative injury. Our most significant signal for NAFLD as a binary trait was in the gene *MTA3* (cg02679503, p=1.81E-07) that plays a role in maintenance of normal epithelial architecture but has not been previously associated with NAFLD. To our knowledge this represents the first EWAS for NAFLD in an adolescent cohort. We identify plausible known and novel candidate disease markers providing additional insight into the pathogenesis of NAFLD. .

1960W

Pro-inflammatory cytokines modulate the transcriptomic landscape of long non-coding RNAs in human islets of Langerhans. *S. Kaur[1], A.H. Mirza[1,2], J. Juan-Mateu[3], D.L. Eizirik[3], F. Pociot[1,2].* 1) Department of Pediatrics, Copenhagen Diabetes Research Center (CPH-DIRECT), Herlev University Hospital, Herlev Ringvej 75, Herlev DK-2730, Denmark; 2) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-1165, Denmark; 3) Center for Diabetes Research, Medical Faculty, Universite Libre de Bruxelles (ULB), Brussels, Belgium.

The pro-inflammatory cytokines released by the infiltrating immune cells in islets of Langerhans modulate the islet gene expression programming, resulting in destruction of insulin-producing β-cells and contributing to type 1 diabetes (T1D). Modulation of the expression of both protein-coding and non-coding RNAs (ncRNAs) contributes to this process. Recent studies have shown a pivotal role for long non-coding RNAs (lncRNAs) in various biological processes, including regulation of cell cycle and immune-mediated processes. Recently, the human islet transcriptome has been profiled using deep sequencing approaches. However, most of these studies have only focused on the protein-coding and small ncRNAs such as miRNAs facet of the islet transcriptome. We have presently used paired-end RNA sequencing to survey the transcriptomic landscape of ncRNAs in human islets from five donors under control conditions and following exposure to the pro-inflammatory cytokines IL-1β and IFN-γ. A total of 1,864 ncRNA genes and 12,194 protein coding genes were expressed in human islets (FPKM >1). Differential expression analysis comparing control vs cytokine exposed samples identified ~1,200 differentially expressed ncRNA genes, including lncRNAs and potential novel miRNAs (logFC ≥ 1.2, pval < 0.05). The expression of ncRNAs modulated by the pro-inflammatory cytokines suggests that they might be involved in regulating β-cell function. The precise molecular mechanisms through which these ncRNAs exert their function are presently under study in relevant cellular and sub-clinical models.

1961T

Novel and replicated DNA methylation loci associated with Type 2 Diabetes among sub-Saharan Africans – The RODAM Study. *K.A.C. Meeks[1], P. Henneman[2], A. Venema[2], A. Adeyemo[3], C. Agyemang[1].* 1) Department of Public Health, Academic Medical Center - University of Amsterdam, Amsterdam, The Netherlands; 2) Department of Clinical Genetics, Academic Medical Center - University of Amsterdam, Amsterdam, The Netherlands; 3) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda MD, The United States.

Introduction: Several Epigenome Wide Association Studies (EWAS) have found significant DNA methylation at several loci associated with type 2 diabetes (T2D), including *ABCG1*, *SOCS3* and *TXNIP*. However, data are lacking on DNA methylation in relation to T2D among sub-Saharan African populations. We therefore aimed to identify DNA methylation loci that are significantly associated with T2D among Ghanaians. **Methods**: The participants consisted of 713 Ghanaians (256 with T2D, 457 controls) from the RODAM (Research on Obesity and Diabetes among African Migrants) Study. The Illumina 450k DNA-methylation array was used to profile methylation on whole blood samples. Association analysis was performed for T2D and adjusted for age, sex, blood cell composition and technical covariates. Body mass index (BMI) was added to the model in order to find T2D associated loci independent of BMI. A replication analysis was done for previously reported loci identified through a systematic literature search for EWAS or methylation studies for T2D. **Results**: The mean fasting glucose was 9.1 mmol/L for T2D cases and 5.0 mmol/L for controls. After Bonferroni adjustment five 450k probes (*cg19693031*, *cg04816311*, *cg00574958*, *cg03078690*, *cg07988171*) were significantly associated with T2D ($p < 5 \times 10^{-7} \sim q < 0.05$). The major hit on the *TXNIP* gene ($p = 2.6 \times 10^{-19}$) showed hypo methylation in T2D cases compared to controls and is located on chromosome 1 in a 3'UTR open sea region. Three out of five hits, located on genes *TXNIP*, *C7orf50* and *TPM4*, remained significant after adjustment for BMI. In replication analysis adjusted for BMI, 15 significant loci were found. The probes on *TXNIP* and *C7orf50* were again identified as the 2 major hits. *TXNIP* is involved in beta-cell biology and known to be in the T2D pathway. *C7orf50* has been previously associated with T2D in Mexican-Americans and Europeans. **Conclusions**: Our study has identified DNA methylation loci associated with T2D in sub-Saharan Africans, including some that have been reported in other populations. These results provide insight into the epigenetic loci and pathways that underly the risk of T2D in sub-Saharan Africans.

1962F

Epigenetic signature of impaired fasting glucose in the Old Order Amish. *M. Montasser, J. O'Connell, Y. Cheng.* Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine.

Type 2 diabetes mellitus (T2D) is a common chronic disease with serious health consequences and substantial economic impact. Lifestyle changes can significantly alter the course of the disease if detected at an early stage. Beside genetics and environmental factors, epigenetic changes may also play a role in T2D development and progression. Recent studies found a strong association between DNA methylation and several T2D related traits. Also, different methylation sites were detected between T2D cases and healthy controls. These results suggest a DNA methylation signature may serve as a biomarker for early detection of increase T2D risk. The few studies examining the association between DNA methylation level and T2D are mostly cross sectional designs, thus cannot discriminate between changes in methylation that predate the illness and to those caused by it. To overcome this limitation we used a longitudinal study design to identify differentially methylated sites between 24 normoglycemic Old Order Amish (OOA) individuals who later developed impaired fasting glucose (IFG) (cases), and 24 OOA individuals who remained normoglycemic after an average follow up of 10 years (controls). Cases and controls were matched on age, sex, BMI, and baseline fasting glucose. Genome-wide DNA methylation profiling for the baseline whole blood was performed using the Illumina Human Methylation 450 bead chip. Out of 274,088 successfully tested probes, we identified 17,702 probes with p-value < 0.05 which is significantly higher than that expected by chance alone ($p=2.2E-16$, X^2 test). While no probe reached genome wide significance of 1.8E-7, we identified 34 probes with suggestive p-values <1.0E-4, 23 of which are located inside genes. Identified genes include *BTC* that plays a role in pancreatic cell proliferation and insulin secretion and *ITGA1*, a known bone mineral density (BMD) gene that was recently found to be associated also with T2D and related traits, providing a possible explanation of the link between T2D and BMD, as well as *RPTOR* and *TSC2* both are part of insulin signaling pathway. Out of 1,828 probes that were identified and replicated in previous studies, 1,219 were tested in our study and 80 of them had a p-value <0.05. If replicated, these results can shed light on the initiation of IFG and help to identify high risk individuals for early intervention.

1963W

Maternal diabetes results in aberrant buccal cell DNA methylation in infants with complex malformations. *K.V. Schulze[1], A. Bhatt[2], M.S. Azamian[1], N. Sundgren[2], G. Zapata[2], P. Hernandez[2], K. Fox[4], J.R. Kaiser[2,4], J.W. Belmont[1,2,3,5], N.A. Hanchard[1,3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) USDA/ARS/Children's Nutrition Research Center, Houston, TX; 4) Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 5) Illumina, Inc., San Diego, CA.

   Uncontrolled maternal diabetes is a well-known teratogen, which causes a wide and highly variable spectrum of clinical phenotypes in offspring, including congenital malformations that can often mimic other genetic syndromes. Yet, the teratogenic mechanism of maternal diabetes remains largely elusive. In the absence of known genetic and alternative environmental causes, the diagnosis "infant of a diabetic mother" (IDM) has thus become one of exclusion. As a mediator between environment and gene expression, DNA methylation is a strong candidate for involvement in the teratogenicity of maternal diabetes. We analyzed IDM methylomes to investigate whether differences in CpG methylation could serve as biomarkers for the diabetic exposure. In order to identify large effects of maternal diabetes, buccal epithelial cells were collected within ten days of birth from infants diagnosed with diabetic embryopathy (n=7). DNA methylation profiles of these newborns were compared to those of healthy infants born to healthy mothers (n=6) using targeted enrichment methylation sequencing (CpGiant, Roche Nimblegen). Differentially methylated CpG sites (MethylKit, p<0.001, ≥10X coverage, with ≤10/1,000 better random permutations) were binned into regions, if they shared the same direction of effect and were within 1kb of adjacent sites. This resulted in 244 differentially methylated regions (DMRs) with ≥3 high-confidence sites showing a difference of ≥10% methylation, and ≥1 significantly differentially methylated CpG identified with a secondary method (MethylSig, p<0.05). Sixty-nine percent of DMRs overlapped known RefSeq genes, some of which have been associated with syndromes that coincide with the phenotypic spectrum of diabetic embryopathy. The majority of DMRs (87%) were hypomethylated and resulted in highly methylated sites transitioning to sites of intermediate methylation. Principal component analysis of all CpG sites within DMRs revealed a clear separation of diabetic embryopathy IDMs and controls, in addition to a split of the former from phenotypically less severe (n=9) and unaffected IDMs (n=6). Our findings demonstrate the added resolution and potential utility of methylation sequencing in a clinical context, imply that there may be large effects of maternal diabetes on offspring DNA methylation in buccal cells, and suggest the potential to use methylation profiles to distinguish the most severe infant phenotypes of maternal diabetes.

1964T

Adipose tissue DNA methylome changes in development of new onset diabetes after kidney transplantation. *Z. Sun[1], S. Baheti[1], Y. Zhang[1,2], P. Singh[3], J. Evans[1], M.D. Jensen[4], V.K. Somers[3], J.P. Kocher[1], H. Chakkera[5].* 1) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA; 2) Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA; 3) Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA; 4) Division of Endocrinology, Mayo Clinic, Rochester, MN, USA; 5) Divisions of Nephrology & Hypertension, Mayo Clinic, Arizona, AZ, USA.

  Background. New onset diabetes after transplant (NODAT) is a complication that occurs in 25-30% of kidney transplant recipients in the first post-transplant year and reduces recipients' survival. The pathophysiology of NODAT is not well understood, but like type 2 diabetes mellitus (T2DM) it may be related to the insulin resistance and genetic and environment factors play a significant role. Epigenetic modifications such as DNA methylation are known to contribute to insulin resistance. The aim of this study was to investigate the role of DNA methylation changes in development of NODAT. **Materials and Methods.** Renal transplantation patients who developed NODAT and did not (control) were selected and matched on age, sex, BMI and other clinical characteristics. The DNA of fresh frozen adipose tissues was extracted for Reduced Representation of Bisulfite sequencing (RRBS). The genome-wide epigenome changes were compared between the NODAT and the control patients. Differentially methylated CpGs (DMCs) and differentially methylated regions (DMRs) were detected by beta binomial model and reginal summation. The involved genes were pursued for their functional relevance. They were also compared with known T2DM associated genes from GWAS studies. Single nucleotide polymorphisms (SNP) were detected from the RRBS data simultaneously and their potential association with NODAT was studied. **Results.** Compared to the control patients, the adipose tissue from NODAT patients had noticeable reduced DNA methylation in almost all the genomic regions except CpG islands (CGIs). DMC analysis showed 78,740 differentially methylated CpGs with p value < 0.05 and methylation difference >10%, among which 41,023 were hypo- and 37,717 hyper-methylated. Similar to overall methylation, these DMCs tended to be more hypermethylated in CGIs but hypo in other genomic regions. DMR analysis revealed 540 DMRs with 390 associated genes. These genes were significantly enriched in certain pathways known to be associated with insulin resistance. Furthermore, 5 of the DMR genes are known genes associated with the risk of T2DM development. Ten genes whose SNPs were associated with NODAT were also among the T2DM genes. Some SNPs were associated with DNA methylation. **Conclusions.** Genome-wide DNA methylation sequencing of adipose tissue revealed the role and loci of DNA methylation in the development of NODAT.

1965F

Insights into type 2 diabetes predisposition through transcriptomic characterization of 427 human pancreatic islet preparations. *M. van de Bunt[1,2], A. Viñuela[3,4,5], N. Oskolkov[6], P.E. MacDonald[7], M.L. Stitzel[8], S.C.J. Parker[9,10], InsPIRE Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom; 3) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Geneva, Switzerland; 5) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 6) Department of Clinical Sciences, CRC, Lund University, Malmö, Sweden; 7) Alberta Diabetes Institute, University of Alberta, Edmonton, Canada; 8) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 9) Department of Human Genetics, Medical School, University of Michigan, Ann Arbor, MI, USA; 10) Department of Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, USA.

Genetic and physiological data support a central role for human pancreatic islet dysfunction in type 2 diabetes (T2D) pathogenesis. Genome-wide association studies have implicated numerous non-coding variants with T2D-risk, but the genes through which these variants act remain poorly characterized. To identify effector transcripts at association loci for T2D, we performed RNA-sequencing and genotyping in human islets from 427 cadaveric donors. Through exon-level eQTL mapping we identified variants significantly affecting expression levels in *cis* at 6039 genes (FDR<1%, 35.0% of total expressed genes). Conditional analysis identified an additional 1705 significant secondary eQTL signals at 1288 genes. Each gene was assigned an islet-specificity score by comparing its expression in islets against that in 16 other tissues. Using islet regulatory state information, we found that eQTLs for islet-specific genes were enriched in stretch enhancers. The enrichment was even stronger for secondary eQTLs compared to primary, suggesting the genetic regulatory architecture of cell specificity is encoded in distal stretch enhancers. Integrating islet eQTL data with genetic information on 82 known T2D-associated loci identified 11 genes where a significant (q<0.01) primary or secondary islet *cis*-eQTL was in strong linkage disequilibrium (r^2>0.9) with the T2D index variant. At one such locus, near *DGKB*, two independent T2D association signals are both linked to independent and directionally consistent *cis*-eQTLs for *DGKB*. Diacylglycerol kinase - *DGKB* encodes the beta isotype - metabolises the secondary messenger diacylglycerol, which plays an important role in insulin secretion. This makes *DGKB* a strong candidate gene for mediating the T2D association. To assess tissue-specificity of the overlap between eQTLs and T2D association signals, we called exon *cis*-eQTLs in 38 tissues in the GTEx v6 data using our pipeline. Given the smaller sample sizes (n=81-361), the number of significant primary eQTLs per GTEx tissue (n=209-3976) was lower than in the islet data. Across all non-islet tissues, 6 genes were linked to T2D associated loci, one of which (*AP3S2*) was also identified in islets. This means islet eQTLs are significantly enriched for overlap with T2D association signals (3.9-fold, p=6.8x10^{-3}) compared to all other GTEx tissues. Our results demonstrate the power of transcriptomic analysis in disease-appropriate tissues to deliver molecular insights in T2D pathophysiology.

1966W

Epigenome-wide association study (EWAS) in metabolic syndrome and its components. *M.L. Nuotio[1,2], N. Pervjakova[1,3], V. Salomaa[2], K. Kristiansson[1,2], M. Perola[1,2,3].* 1) University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare (THL), Helsinki, Finland; 3) University of Tartu, Tartu, Estonia.

Background: Metabolic syndrome (MetS) is a controversial entity, which encompasses a cluster of risk factors that have been found to significantly increase the risk of cardiovascular disease and type 2 diabetes mellitus. Genetic and metabolomic studies about the metabolic syndrome have so far produced rather modest biological information about this condition, suggesting that biological processes behind the entity are very complex and not easily interpreted. Studying epigenetic differences between individuals with metabolic syndrome and their healthy counterparts, as well as the epigenetic differences in each component of metabolic syndrome separately, can shed additional light on the complex genetic and epigenetic processes underlying the phenotype and give a needed push for the challenging research on cardiovascular disease and type 2 diabetes mellitus. **Methods:** We have tested the association between the level of methylation in CpG methylation sites across the genome and MetS status in individuals defined as cases or controls for metabolic syndrome with epigenome-wide linear regression where the M-value is corrected for age, sex, smoking and alcohol use. The association between the level of methylation in CpG methylation sites and the individual components of metabolic syndrome has been tested similarly. This study includes eligible individuals from Finnish FINRISK survey conducted in 2007 (Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome, DILGOM, study collected as an extension to the FINRISK 2007 survey) forming a total sample size of 517 individuals from whom the whole-blood DNA Illumina Infinium HumanMethylation450 BeadChip methylation data is available. **Results:** When comparing MetS cases and controls, various hyper- and hypomethylated CpG methylation sites were found across the genome. The strongest site reached the significance level of P=5.1x10^{-8} for association. Significant associations were also observed with glucose and lipid traits. **Conclusions:** In our study setting, various CpG methylation sites across the genome show altered levels of methylation. Given the challenging nature of epigenetic studies, further research is needed to interpret the biological implication of the findings. We believe however, that these results potentially point out novel functional sites across the genome that associate with biological processes underlying metabolic syndrome as an entity as well as related pathologies.

1967T

Gene-by-environment interactions in diet-induced epigenetic profiles.

C.R. Edillor[1], Y. Blum[2], B.W. Parks[1,2,4], L.D. Orozco[5], M. Pellegrini[3], A.J. Lusis[1,2,4]. 1) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Department of Medicine, Cardiology Division at the University of California Los Angeles, Los Angeles, CA; 3) Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA; 4) Department of Microbiology, Immunology & Molecular Genetics, University of California Los Angeles, Los Angeles, CA; 5) Department of Bioinformatics, Genentech Inc., South San Francisco, CA.

This study examines gene-environment interactions in obesity through the lens of epigenetics. Gene-environment interactions measure the effect of an individual's genotype in response to extrinsic variables such as diet. These interactions contribute to the heritability of complex phenotypes by capturing the effect of environmental modulation on underlying genetics. Gene-environment interactions are difficult to detect in human genetic studies due to variability in the environment of different individuals. The Hybrid Mouse Diversity Panel (HMDP) leverages natural genomic variation between different inbred mouse strains to isolate regions associated with complex phenotypes, while controlling for environment. To interrogate gene-environment interactions in diet-induced obesity, strains from the HMDP consumed a calorie-rich, high-fat, high-sucrose diet for 8 weeks. Clinical traits related to metabolism and adiposity were measured, and relevant tissues sampled for genomic analysis. While the inter-strain variability of obesity phenotypes is in part explained by nucleotide variation, epigenetic differences as a function of environment and genetics provide further insight into the genomic etiology of obesity and metabolic disease. This study uses reduced-representation bisulfite sequencing to examine global DNA methylation patterns in whole livers of 40 HMDP strains fed obesogenic (32% kcal fat) or chow (6% kcal fat) diets. Many biological processes regulating metabolism and adiposity occur in the liver. We identified variably methylated sites between strains and diets, suggesting diverse networks of epigenetic regulation following dietary perturbation. We are able to identify regions that vary in methylation as a response only to environment/diet; those that vary as a function of genetic background; and those whose response to diet depends on the genotype, indicating gene by environment effects. Overlapping these data with previously published expression and chromatin data in genetically diverse strains provides a multifactorial snapshot of metabolic disease etiology. In summary, this study identifies hepatic epigenetic mechanisms involved in genotype-environment interactions of diet-induced obesity.

1968F

Mechanistic and functional properties of the *KLF14 trans*-eQTL network associated to risk of type 2 diabetes.

J.S. El-Sayed Moustafa[1], J. Fernandez[2], M. Civelek[3], C. Glastonbury[1], M. Todorcevic[4], A. Mahajan[2], M. Horikoshi[2], A. Nag[1], I. Yet[1], M. Simon[5], G. Thorleifsson[6], U. Thorsteinsdottir[6,7], K. Stefansson[6,7], J. Bell[1], A.L. Gloyn[2,4], R. Cox[5], A. Lusis[8], F. Karpe[4], M. McCarthy[2,4], K. Small[1]. 1) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Department of Biomedical Engineering, University of Virginia, Charlottesville, VA; 4) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK; 5) MRC Harwell, Oxford, UK; 6) deCODE genetics/Amgen, Reykjavik, Iceland; 7) Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; 8) School of Medicine, University of California, Los Angeles, CA.

Variants upstream of the transcription factor *KLF14* are associated with type 2 diabetes (T2D) and HDL cholesterol via adipose-specific *cis*-regulation of *KLF14* expression ($P=8.70\times10^{-37}$) and resultant *trans*-regulation of 385 genes. The *KLF14 trans*-network is remarkable both for its size and reproducibility, replicating in three independent adipose expression datasets ($N_{total}=2,826$), and thus provides a powerful model to investigate mechanisms of *trans*-regulation in humans and its role in disease pathogenesis. Mediation analysis supports a causal role for *KLF14* expression in *trans*-gene regulation. The dominant *trans*-regulatory mechanism appears to be direct interaction of KLF14 with *trans*-gene *cis*-regulatory elements; upstream regions of the *trans*-genes are enriched for KLF14 binding motifs (NES=4.17) and direct binding of KLF14 in Chip-Seq data ($P=2.0\times10^{-4}$). *Trans*-genes bearing KLF14 motifs are enriched for activity during adipocyte differentiation ($Q=3.85\times10^{-8}$), consistent with a proposed role for *KLF14* in progression through adipogenesis. A sub-network of 18 genes enriched for binding motifs of SREBF1, itself a *KLF14 trans*-gene, but not KLF14 (NES=4.22) were enriched for lipid metabolism pathways ($Q=3.24\times10^{-7}$) and may mediate the lipid GWAS phenotypes. The *KLF14 trans*-eSNP is also an adipose *trans*-meQTL; the *KLF14 trans*-network is enriched for *trans*-regulation of methylation levels by the *KLF14* eSNP ($N=542; P\leq2.2\times10^{-16}$). Consistent with the known function of *KLF14* as both a transcriptional activator and repressor, the *trans*-eQTL includes both positive and negative effects, however the *trans*-meQTL is biased towards negative effects, consistent with increased KLF14 occupancy. We find a direct link between a subset of *KLF14 trans*-genes and T2D: *cis*-eQTL variants at three *trans*-genes (*STARD10*, *c6orf57* and *CDK2AP1*) were coincident with genome-wide significant T2D variants; signals in some instances attributed to different genes. This suggests both *cis* and *trans* regulation of these genes can impact risk of T2D. Within the *KLF14* network, we also observe evidence of gene-gene interactions between the *KLF14 trans*-eSNP and *cis*-eQTL variants of the *KLF14 trans*-genes, and replicate four GxG *cis-trans* effects in an independent dataset ($N=770$). Results of in-depth exploration of functionality and regulatory mechanisms of the *KLF14 trans*-network thus provide a potential mechanistic basis for *KLF14 trans*-regulation of gene expression mediating its role in T2D pathogenesis.

1969W

variancePartition: Interpreting drivers of variation in complex gene expression studies. *G.E. Hoffman, E.E. Schadt.* Department of Genetics and Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

As large-scale studies of gene expression with multiple sources of biological and technical variation become widely adopted, characterizing these drivers of variation becomes essential to understanding disease biology and regulatory genetics. We describe a statistical and visualization framework, variancePartition, to prioritize drivers of variation based on a genome-wide summary, and identify genes that deviate from the genome-wide trend. Using a linear mixed model, variancePartition quantifies variation in each expression trait attributable to differences in disease status, sex, cell or tissue type, ancestry, genetic background, experimental stimulus, or technical variables. Analysis of large-scale transcriptome profiling datasets illustrates that variancePartition recovers striking patterns of biological and technical variation that are reproducible across multiple datasets. Our open source software, variancePartition, enables rapid interpretation of complex gene expression studies as well as other high-throughput genomics assays. variancePartition is available on Bioconductor: http://bioconductor.org/packages/variancePartition.

1970T

An epigenome wide association study of BMI percentile in a family cohort with children and adolescents identified novel epigenetic loci influencing metabolic syndrome. *Y. Zhang[1,2], D. Cerjak[1,2], J.W. Kent Jr.[3], R. James[1,2], J. Blangero[4], M.A. Carless[3], O. Ali.[5].* 1) Medicine, Medical College of Wisconsin, Milwaukee, WI; 2) Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI; 3) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX; 5) Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Background: Epigenetic mechanisms mediate the interaction between gene and environment and may play an important role in the obesity epidemic. Metabolic Syndrome (MetS), a clinical comorbidity of obesity, affects 30% of American population with a clustering of cardiovascular risk factors. Epigenetic mechanisms underlying MetS are largely unknown. *Methodology:* We conducted a comprehensive epigenetic study in which we assessed the relationship of DNA methylation in peripheral blood mononuclear cells (PBMCs) with body mass index percentile (BMI%), a clinical index for obesity in children and adults, at 485,000 CpG sites across the genome in multi-generational families of Northern European ancestry (8-90 years of age) using a discovery cohort (192 individuals) and a validation cohort (1,052 individuals). We also assessed gene expression at the most significant epigenetic site in a subset of the validation cohort consisting of 330 subjects. *Results:* After Bonferroni-correction ($p_{\alpha=0.05}$=1.31 x 10^{-7}) for genome-wide significance, we identified three loci, cg18181703 (*Suppressor of Cytokine signaling 3, SOCS3*), cg04502490 (*Zinc Finger Protein 771, ZNF771*) and cg02988947 (*LIM domain containing 2, LIMD2*) where methylation status was associated with BMI%, a clinical index for obesity in children and adults. These sites were also associated with multiple MetS traits including central obesity, fat depots, insulin responsiveness and plasma lipids. The *SOCS3* methylation locus was also associated with the clinical definition of MetS. In the validation cohort, *SOCS3* methylation status was found to be inversely associated with BMI percentile (p=1.75 x 10^{-6}), waist to height ratio (p=4.18 x 10^{-7}), triglycerides (p=4.01 x 10^{-4}) and MetS (p=4.01 x 10^{-7}) and positively with HDL-cholesterol (p=4.57 x 10^{-8}). Functional analysis in a sub cohort (333 individuals) demonstrated *SOCS3* methylation and gene expression in the PBMCs were inversely correlated (p=2.93 x 10^{-4}) and expression of *SOCS3* was positively correlated with status of MetS (p=0.012). *Conclusion:* We conclude that epigenetic modulation of *SOCS3*, a cytokine signaling regulator gene involved in leptin and insulin signaling, may play an important role in etiology of MetS and DNA methylation of this gene may have diagnostic, prognostic and therapeutic significance.

1971F

Identification of cis-eQTLs for co-expressed genes associated with body mass index. *E. Kim[1], M. Alvarez[1], K. Mohlke[2], M. Laasko[3], P. Pajukanta[1,4,5].* 1) Dept. of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Dept. of Genetics, University of North Carolina, Chapel Hill, NC; 3) University of Eastern Finland, Kuopio, Finland; 4) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 5) Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA.

Obesity affects one third of adults in the U.S., and is known to increase the risk of metabolic syndrome, heart disease, stroke, and diabetes. Complex diseases can be attributed to complex gene regulation and expression, which in turn is determined by expression quantitative trait loci (eQTLs). We hypothesized that sets of co-expressed genes in adipose tissue can influence development of obesity. To elucidate the expression patterns that contribute to obesity, we obtained RNA-sequence data from subcutaneous adipose samples as well as imputed genotypes and phenotype data from the Finnish METSIM (n=793) and GTEx (n=283) cohorts. We conducted weighted gene co-expression network analysis (WGCNA) on each cohort to obtain co-expression modules correlated with body-mass index (BMI). We performed eQTL analysis on the imputed genotype data (MAF>5%) from each cohort with Matrix-eQTL to discover cis-eQTLs in the co-expressed gene modules. We found significant co-expression modules for BMI (adjusted P=0.00063) in METSIM that are highly preserved between populations: 578 out of 720 genes in the METSIM network modules for BMI were found in GTEx modules for BMI, indicating that these genes may play an integral role in regulating BMI across populations. Among the shared cis-eQTL SNPs ($\leq$1 Mb, FDR <5%) that affect the same gene in the same direction in both populations, we observed 6,440 cis-eQTLs (on average 55 SNPs/gene) associated with 118 of the 578 genes (20%) from the BMI-correlated modules. The large number of variants reflects the tight linkage disequilibrium among the imputed SNPs in some regions. Functional pathway analysis of the 118 shared cis-regulated genes and the 578 overlapping genes in the BMI-modules showed significant KEGG pathways (adjusted P<0.05) using the Webgestalt server. The KEGG pathways for the 118 genes included salivary secretion and carbohydrate digestion and absorption, and for the 578 genes, multiple inflammatory pathways, supporting the claim that the co-expression of these genes serves a functional purpose in regulating BMI. In conclusion, we identified a set of co-expressed BMI genes with shared cis-eQTL regulators across populations.

1972W

Landscape of obesity pathophysiology across 44 human tissues: Evidence for disease-relevance and causality through transcriptome sequencing. *T. Tukiainen[1,2], T. Pers[2], F. Aguet[2], P. Chen[1], P. Luukkonen[3], A. Rissanen[4], C. Lindgren[5], K. Ardlie[2], K. Pietiläinen[4], S. Ripatti[1,6,7].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Minerva Foundation Institute for Medical Research, Helsinki, Finland; 4) Obesity Research Unit, Research Programs Unit, University of Helsinki, Helsinki, Finland; 5) The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 6) Department of Public Health, Faculty of Medicine, University of Helsinki, Finland; 7) Wellcome Trust Sanger Institute, Hinxton, UK.

Obesity is a world-wide epidemic with severe and diverse health consequences. To elaborate the mechanisms linking obesity to its many sequelae, we investigated the association of body mass index (BMI), a common measure for adiposity, with gene expression (GE) assessed via RNA-sequencing in 44 tissue types in 450 genotyped donors from the Genotype-Tissue Expression (GTEx) project. Together these tissues cover the majority of human body systems, thus providing an unprecedented opportunity to understand the obesity pathophysiology. We find BMI-GE-associations in several tissues (FDR<5%), including adipose and liver, but also skin and lung tissues. The BMI-associated genes are only partly shared between tissues (estimate for sharing, mean π_1=21%) indicating tissue-dependent processes. Gene set enrichment analyses further reveal the plethora of BMI-related alterations across human tissues: Besides replicating previously documented active inflammation and reduced mitochondrial function in adipose tissue ($P_{perm.}$<0.005), we, for instance, find that increased BMI associates with decreased sensory perception in skin ($P_{perm.}$=0.004) and with reduced male germ cell number in testis ($P_{perm.}$=0.034). Our analyses also reveal tissue-tissue interplay. E.g. most processes shared between adipose tissue and adrenal gland have an opposing direction of effect, suggesting possible compensatory mechanisms between these two endocrine tissues. Pointing to potential disease-relevance of these BMI-associations, we observe that genes implicated in genome-wide association studies (GWAS) are enriched among the genes highlighted in the BMI-GE analysis. E.g. GE-alterations in adipose tissue significantly overlap with GWAS loci for several autoimmune traits, including celiac disease and multiple sclerosis ($P_{perm.}$=0.001), underscoring the presence of inflammation in obesity and supporting epidemiological observations of suggested link between obesity and immune system function. To disentangle whether BMI is mediating these GE-changes or vice versa, we utilize Mendelian randomization, and find evidence for BMI causally driving several of the observed GE-changes (e.g. $P_{perm.}$<0.0001 in adipose tissue). In summary, these results highlight how widespread and diverse obesity-associated changes are in the human body. The new affected pathways and causal links identified provide potential avenues for the prevention and treatment of obesity complications.

1973T

Cardiometabolic GWAS variants colocalize with subcutaneous adipose tissue splice eQTLs (sQTLs). *C.K. Raulerson[1], Y. Wu[1], A. Ko[2], M. Alvarez[2], T.S. Furey[1,3], M. Laakso[4], P. Pajukanta[2,5], K. Molhke[1].* 1) Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Department of Biology, University of North Carolina, Chapel Hill, NC; 4) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Molecular Biology Institute at UCLA, Los Angeles, CA.

Cardiovascular and metabolic diseases lead to substantial public health burden and increased mortality rates in the US and worldwide. Genome-wide association studies (GWAS) have identified >200 loci for cardiometabolic traits such as type 2 diabetes (T2D), LDL and HDL cholesterol levels, triglyceride levels, body fat distribution and adiposity traits. However, the variants, genes, and mechanisms that underlie these signals are not well understood. Expression of different isoforms, as identified by RNA-sequencing reads spanning splice junctions, may contribute to the mechanisms of these GWAS signals. To investigate genetic variants that influence the expression of different splice junctions, called splice-expression quantitative trait loci (sQTLs), we used genotypes and subcutaneous adipose tissue RNA-seq data (50-bp paired-end, avg. ~45M reads) from 792 Finns from the METabolic Syndrome in Men (METSIM) cohort. We first aligned these samples to the hg19 reference, using STAR-2pass, and implemented the newly-described LeafCutter algorithm to identify both novel and known splice junctions (n=313,192) expressed in our dataset. We quantile-normalized these splice junctions, adjusted for batch and RIN value, and performed association tests between splice junctions and 8.9 M genetic variants (MAF>0.01) using Matrix-eQTL. We identified 105,055 splice junctions with at least 1 sQTL at a chromosomal FDR<1%. Among these, we focused on sQTL variants in LD with a set of 1,231 cardiometabolic GWAS loci from the NHGRI-EBI GWAS catalog. We identified 87 sQTL variants at 72 genes for which the sQTL variant is in LD $r^2>=0.8$ with the lead GWAS variant. For example, at the T2D locus *POU5F1-TCF19*, we identified an intronic SNP (rs3094226) associated with two known splice junctions of the *CCHCR1* gene (best p=9x10^{-11}), which act in the same direction. This result suggests a potential role in T2D risk for *CCHCR1,* which encodes coiled-coil alpha-helical rod protein 1. We further identify 22 sQTLs for genes without a reported gene-level eQTL in adipose, suggesting that splicing may contribute to the mechanisms. At 31 GWAS loci, only one splice junction has an sQTL; among these, 24 tag a unique isoform and 7 tag a splice junction shared among isoforms, which may represent the aggregate gene. Taken together, these results suggest that sQTL analysis of RNA-seq data can aid with the interpretation of target genes and mechanisms at GWAS loci.

1974F

Transcriptomic profiles of liver biopsies from obese individuals reveal 15 genes for fatty liver. *A. Ko[1,2], D. Kaminska[3], E. Nikkola[1], M. Laakso[4], J. Pihlajamäki[3,5], P. Pajukanta[1,2,6].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Molecular Biology Institute at UCLA, Los Angeles, CA; 3) Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland; 4) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland; 6) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA.

Obesity is a serious health problem, predisposing to cardiometabolic disorders and leading to non-alcoholic fatty liver disease (NAFLD) that may escalate to nonalcoholic steatohepatitis (NASH), cirrhosis, and cancer. However, not all extreme obese individuals exhibit NAFLD, suggesting genetic predisposition. To identify regulatory variants and genes affecting liver health, we genotyped 265 obese individuals (mean body mass index (BMI)=43.07) that underwent bariatric surgery and RNA-sequenced their liver biopsy samples. To estimate liver health, we performed a non-linear PCA on five different liver traits (inflammation, steatosis, fibrosis, liver cell degeneration, and clinical diagnosis) and used the first component as a meta-liver trait. By Weighted Co-expression Network Analysis, we identified 7 co-expressed gene modules associated with the meta-liver trait (P<0.0021) (referred as NAFLD-modules). The NAFLD-modules are enriched for natural killer cell mediated cytotoxicity (adjusted P=6.33x10^{-16}) and metabolic KEGG pathways (adjusted P=0.006). Differential expression (DE) between healthy liver (n=175) and NASH (n=43) based on clinical diagnosis revealed a significant upregulation of central metabolic functions in healthy liver, whereas in NASH, upregulation of autoimmune and cytotoxicity pathways prevails. NAFLD-modules are significantly enriched for DE genes vs. non-NAFLD modules (51.38% vs. 18.43%, P<0.0001), indicating a coordinated shift in transcriptional regulation between healthy liver and NASH. Finally, we carried out *cis*-eQTL (≤1MB) mapping (n=265) and identified 3,987 independent loci (R^2<0.1) regulating 3,128 genes. We compared the *cis* eQTL effect sizes in obese liver to non-obese liver from GTEx (n=96 subjects, mean BMI=26.7) and found 210 SNPs, differentially regulating 197 genes (adjusted P<0.05, Z-score), which are enriched for drug metabolism and metabolic pathways (adjusted P<0.05). Fifteen DE genes in 4 NAFLD-modules are also under this differential *cis*-eQTL effect, including *MTHFD1* (lead eQTL rs2987965), a key liver enzyme of folic acid metabolism, and *NR3C2* (lead eQTL rs6970169), a nuclear receptor associated with adverse cardiovascular outcomes. Overall, we discovered tight liver transcriptional control perturbed by NASH, and 15 genes that reside in key co-expressed gene networks associated with liver health and are distinctly *cis*-regulated in obese liver as well as differentially expressed between healthy liver and NASH.

1975W

RNA-sequencing reveals multiple sclerosis patients share similar pathways to Sjögren's syndrome anti-Ro negative patients. *I. Adrianto[1], J.A. Ice[1], A. Rasmussen[1], C.G. Montgomery[1], R.H. Scofield[1,2,3,4], G. Pardo[1], K.L. Sivils[1,2], R.C. Axtell[1], C.J. Lessard[1,2].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Department of Medicine, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 4) Medical Service, Department of Veterans Affairs Medical Center, Oklahoma City, OK.

   Multiple sclerosis (MS) is an inflammatory and degenerative disorder of the central nervous system characterized by damage to the myelin sheath. Sjögren's syndrome (SS) is an autoimmune disease characterized by autoantibodies to Ro and/or La proteins and lymphocytic infiltration into exocrine glands. Even though MS and SS have different clinical manifestations, genetic studies have suggested that the underlying etiology is common in both diseases. We used RNA-seq to compare the expression of protein-coding and non-coding transcripts in 15 MS patients to SS anti-Ro positive (n=27) and SS anti-Ro negative (n=30) patients as well as 27 healthy controls. Whole blood RNA samples were isolated using the NuGEN Encore kit. Sequencing was performed using the Illumina HiSeq 2000. Raw FASTQ files were aligned to the human genome using Tophat. The read counts per transcript were generated using easyRNASeq in R. Differentially expressed (DE) transcripts were determined using DESeq with a false discovery rate adjusted P-value (P_{adj}) of 0.05 and a fold change (FC) of >2 or <0.5. Among the 2203 DE protein-coding regions, GRIN3B was the most DE transcript in the MS case-control analysis (P_{adj}=7.40x10^{-32}, FC=4.30). Of the 1849 DE non-protein coding transcripts, we observed DE of a long non-coding RNA (lncRNA) at 5p15.1 (P_{adj}=2.57x10^{-22}, FC=0.29). The biological relevance of these transcripts to MS is unclear. Relevant pathways from MS DE transcripts included synaptic transmission, nervous system development, and cell differentiation among others similar to those of SS anti-Ro negative patients. Using DE transcripts relevant to distinguish two groups of SS patients, we also found MS patients clustered together with SS anti-Ro negative patients in the principal component analysis plot. In addition, we observed the expression of interferon-inducible genes and the B-cell cytokine, APRIL, were significantly higher in SS anti-Ro positive patients compared to MS patients and to healthy controls. Our analysis shows that RNA profiles of MS patients are more similar to SS anti-Ro negative patients than the SS anti-Ro positive patients. This suggests that MS treatments may be beneficial for SS anti-Ro negative patients. Furthermore, these data have identified several putative DE in both coding and lncRNA regions in SS and MS. Further comparisons of different MS and SS subtypes are needed to confirm and expand these findings.

1976T

Single cell expression quantitative trait loci (eQTL) analysis of established systemic lupus erythematosus (SLE)-risk loci in lupus patient monocytes. *Y. Ghodke-Puranik[1], Z. Jin[1], W. Fan[1], M. Jensen[1], J. Dorschner[1], D. Vsetecka[1], S. Amin[2], A. Makol[2], F. Ernste[2], T. Osborn[2], K. Moder[2], V. Chowdhary[2], T. Niewold[1].* 1) Dept of Immunology and Division of Rheumatology, Mayo Clinic, Rochester, MN; 2) Division of Rheumatology, Mayo Clinic, Rochester, MN.

  Purpose: While most of the confirmed SLE-risk loci are in or near genes with immune system function, a major unanswered question is how these loci influence diverse immune cell subsets. In this study we performed a single cell eQTL analysis in human monocytes to determine impact of some well-established SLE-risk loci in single human monocytes.**Methods:** CD14++CD16-classical monocytes (CL) and CD14dimCD16+ non classical (NCL) monocytes from SLE patients were purified by magnetic separation. The Fluidigm C1 System was used for single cell capture and target gene pre-amplification and equal numbers of classical and non-classical monocytes were studied. Real time PCR was used to quantify expression of 90 monocyte-related genes, and the same SLE patients were genotyped for 7 SLE-risk SNPs to enable eQTL analysis. Non-parametric analyses were used with the single cell data in CL and NCL populations separately. **Results:** We observed a large number of significant eQTL associations that surpassed the 5% FDR, supporting the idea that single cell gene expression data allows for robust eQTL discovery. The SLE-associated SNPs demonstrated more eQTLs in NCLs as compared to CLs (p=2.5x10^{-8}). For a given SNP, the eQTL associated transcripts differed between cell types (p<0.001 for all 7 SNPs for discordance), suggesting that the same SNP resulted in different cellular events between the two monocyte subsets. When comparing eQTL lists between the different SLE-associated SNPs, there was a greater degree of sharing observed in NCLs as compared to CLs. Loci which shared a significant proportion of eQTL associations with each other in NCLs included TNFAIP3, IRF5, IRF7, PTPN22, and SPP1. In CLs, TNFAIP3 shared a large number of eQTLs with SPP1 and ITGAM, although SPP1 and ITGAM showed more limited overlap with each other. Thus, SLE-associated risk loci exert coordinated effects on gene expression within individual human monocytes, and the risk loci interact in different ways in different cell types.**Conclusion:** This study emphasizes the strength of single cell gene expression strategy for eQTL discovery. Our study revealed striking differences in the occurrence and interaction between of SLE risk associated eQTLs within different but closely related cell types. This suggests pleiotropic effects from each locus across various immune cell types, and a high degree of complexity when considering how these loci impact the immune system. .

1977F

Stimulating human CD4+ memory T cells uncovers cell state-dependent allelic expression enriched in risk genes for rheumatoid arthritis. *M. Gutierrez-Arcelus[1,2], S. Hannes[1,2], N. Teslovich[1,2], H.J. Westra[1,2], K. Slowikowski[1,2], D.A. Rao[3], J. Ermann[3], M.B. Brenner[3], S. Raychaudhuri[1,2,4].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Smith Building, 1 Jimmy Fund Way, Boston, MA 02115, USA; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

A major challenge in autoimmune disease genetics is elucidating how genetic risk factors affect gene regulation in the specific cell-types and physiological states most relevant for each disease. Since allele specific expression is largely driven by genetic regulatory variation in *cis* (Buil *et al.* 2015), it is a powerful approach to study regulatory effects in multiple cellular states without the need of a very large sample size. Here we have analyzed the effects of genetic variation on gene regulation in varying cell states of CD4+ memory T-cells, a relevant cell-type for rheumatoid arthritis (RA) (Hu *et al*, 2011). As a pilot, we isolated CD3+ CD4+ CD45RO+ T-cells by negative selection with magnetic beads from the peripheral blood of a single healthy individual. We stimulated these cells with anti-CD3/CD28 beads and performed RNA-seq at 0, 4, 8, 12, 24 and 72 hours following stimulation. We then called heterozygous variants from RNA-seq data and measured allelic expression. We performed logistic regression to identify sites with significant allelic expression changes over time. In this high-resolution time course experiment, out of 3,578 tested heterozygous sites, we found 17 SNPs (FDR < 10%) with cell-state dependent allelic effects, overlapping 13 protein-coding genes. These effects reflect the activation or inactivation of genetic regulatory effects upon stimulation. We found a 2-fold enrichment of low P-values for SNPs in RA risk genes. We have now optimized the experimental set up to process 1M CD4+ memory T cells per time point coming from 80-100 mls of blood, and multiple genotyped individuals at a time. We are currently evaluating technical variation and we are in the process of expanding the experiment to 20 densely genotyped individuals. This will allow us to characterize disease-associated pathways affected by genetic regulatory variation at multiple stages during T-cell activation.

1978W

Polygenic burdens on cell-specific cytokine pathways are the risk of rheumatoid arthritis. *K. Ishigaki[1,2], Y. Kochi[1], A. Suzuki[1], Y. Okada[6], Y. Momozawa[3], Y. Kamatani[2], R. Yamada[5], K. Fujio[4], M. Kubo[3], K. Yamamoto[1,4].* 1) Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 2) Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 4) Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Japan; 5) Statistical Genetics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 6) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan.

The mechanisms underlying genetic components of complex diseases are still not well-elucidated and becoming the major target of genetic research in the post-GWAS era. Comprehensive characterization of human genome by ENCODE Project demonstrated that GWAS-identified risk alleles are enriched within cell-specific regulatory regions. Although many eQTL studies have been conducted, majority of them utilized unfractionated peripheral blood or a few cell populations from European descents. The functional variation of immune cell subsets and the complexity of population specific haplotype structure prompted us to design cell-type specific eQTL study using Asian samples. We collected peripheral blood from 105 Japanese healthy volunteers and further sorted them into five major immune cell subsets (CD4+ T cells, CD8+ T cells, B cells, NK cells and monocytes). We quantified gene expressions and splice isoforms by RNA-sequencing and tested its association with neighboring common variants. We identified 8,204 genes with significant cis-eQTL effect and 2,226 genes with significant splicing-QTL effect at least in one subset (q-value < 0.05). Given that dysregulated transcriptome caused by multiple risk alleles is probably the major drivers of immune related complex diseases, the analysis which directly examines the causal abnormality in the transcriptome should be a more straightforward approach to detect fundamental mechanism of diseases than GWAS that targets the genome. Furthermore, this approach has potential to expand the analysis from individual genes into activity of relevant pathways. Based on such motivation, we developed a three-step analytical pipeline comprising i) prediction of individual gene expression using our eQTL database, ii) gene-level association study and iii) prediction of causal pathway activity. By applying our pipeline to two independent Japanese rheumatoid arthritis GWAS datasets (2303 cases versus 3380 controls and 2342 cases versus 5017 controls), we identified multiple candidates of cell-specific causal genes (8 novel genes with Bonferroni significance) and causal cytokine pathways (upregulation of TNF in CD4+ T cells and other 8 cytokines). Our approach is an efficient way to understand the biological mechanisms of genetic contributions on complex diseases.

1979T

DNA methyltransferase 1 inhibitor 5-aza-2'-deoxycytidine modifies gene expression profiles at the asthma-associated *ZPBP2-ORMDL3-GSDMA* region in human airway epithelial cells. *S. Moussette[1], A. Al Tuwaijri[2], H.-R. Kohan-Ghadr[1], R. Farias[3], J. Bérubé[3], S. Rousseau[3], A.K. Naumova[1,2,4].* 1) The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) The Meakins-Christie Laboratories at the Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 4) Department of Obstetrics and Gynecology, McGill University, Montreal, Quebec, Canada.

Chromosomal region 17q12-q21 is associated with asthma and harbors regulatory polymorphisms that influence expression levels of three genes: *zona pellucida* binding protein 2 (*ZPBP2)*, gasdermin B (*GSDMB)* and ORMDL sphingolipid biosynthesis regulator 3 (*ORMDL3*). The 17q12-q21 genotype and asthma also influence promoter methylation levels of *ZPBP2* and gasdermin A (*GSDMA*). DNA methylation differences between asthmatic and non-asthmatic subjects are very small, which raises the question of whether such small changes in methylation levels are sufficient to cause significant effect on gene expression. Therefore, the functional significance of inter-individual variation in methylation levels in this region had to be established. To clarify the role of DNA methylation in the genetic control of gene expression in the 17q12-q21 asthma-associated region, we examined the impact of DNA methyltransferase 1 inhibitor 5-aza-2'-deoxycytidine (5-aza-dC) on DNA methylation of regulatory elements and gene expression in the human airway epithelium cell line NuLi-1. NuLi-1 cells were treated with 0.5 μM 5-aza-dC for 24h and harvested after 7 days in culture for DNA and RNA extraction. Expression levels of 5 genes, IKAROS family zinc finger 3 (Aiolos) (*IKZF3*), *ZPBP2*, *GSDMB*, *ORMDL3* and *GSDMA* residing in the asthma-associated region, were determined using quantitative RT-PCR. Allelic expression was characterized using RT-PCR followed by Sanger sequencing. Promoter DNA methylation was assessed using the sodium bisulfite sequencing assay. 5-aza-dC treatment reduced promoter methylation levels (11% to 30%) and upregulated *ZPBP2* (19-fold)*, GSDMB* (4-fold) and *GSDMA* (5.7-fold) expression. Therefore, our data show that moderate reduction in promoter methylation results in dramatic upregulation of gene expression. Analysis of allelic expression showed no change in allelic ratios of *IKZF3*, *GSDMB* and *GSDMA*; however allelic expression of *ZPBP2* and *ORMDL3* was modified. We conclude that modest changes in promoter methylation levels may have a significant impact on RNA levels. Furthermore, allelic expression of *ZPBP2* and *ORMDL3* depends on DNA methylation, whereas allelic expression of *IKZF3*, *GSDMB* and *GSDMA* does not.

1980F

Novel statistical and experimental approaches for quantifying the X chromosome inactivation landscape using RNA-seq data. *D. Liu[1], L. Carrel[2].* 1) Department of Public Health Sciences, Penn State University, Hershey, PA; 2) Department of Biochemistry and Molecular Biology, Penn State University, Hershey, PA.

X chromosome inactivation (XCI) is a cellular mechanism which randomly silences one female chromosome in each cell (i.e. the inactive X chromosome - Xi) and only express alleles on the active X (Xa) for most genes. However, a subset of genes is known to escape XCI, and expressed from both alleles. It is well known that XCI escape is variable across different individuals and may be genetically regulated. A key step to understand XCI regulation is to quantify the XCI status. Direct assignement of the Xa or Xi is complicated by the mosaic nature of XCI. In order to properly analyze broadly available RNA-seq dataset, we developed a novel empirical Bayes method to infer the XCI status. The method first calculates the XCI skewness using a training set of 177 genes known to be inactivated, and compares the genic allelic expression with the estimated skewness. Genes that have different allelic imbalance from the skewness are likely to escape XCI. In order to improve power, we also borrow strength across samples by assuming realistic parametric distribution on the read depths and allelic expression levels and estimate the posterior probability for each gene to escape XCI using a new approximate empirical Bayes method. Applying these methods, we analyzed the female samples from GEUVADIS. After rigorous quality control and the correction of reference bias, 670 genes are informative for allelic inference and contain at least 1 heterozygous SNPs. We identified 375 genes to be X inactivated, while 207 genes are classified as variable escape genes and bi-allelically expressed in >15% of the samples. We validated our approaches using multiple lines of evidence. First, pseudo-autosomal (PAR) genes are expected to escape XCI and importantly 10 of 11 PAR genes were correctly classified. Next, we also observe high concordance between escape genes identified by our method and by male/female differential expression. 72% of genes identified with significant Xi expression levels are replicated by differential expression analysis. Finally, to confirm or refine estimates of XCI skewing and inference of XCI escape in mosaic LCLs, we selected a mosaic LCL and isolated single-cell derived lines with complete non-random XCI of the maternal or paternal Xi. In these lines, the number of reads that over one X directly reflects XCI escape without the need to infer XCI skewing. The analysis of the experimental data is on-going, and will be used to confirm our findings and methods.

1981W

Regulatory fine mapping resolves shared and distinct associations to multiple diseases in the same locus. *P. Shooshtari[1,2], C. Cotsapas[1,2,3].* 1) Department of Neurology, Yale University, New Haven, CT; 2) Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of MIT-Harvard, Cambridge, MA; 3) Department of Genetics, Yale University, New Haven, CT.

Genome-wide association studies (GWAS) have identified hundreds of loci mediating risk to autoimmune and inflammatory diseases (AID). These diseases are co-morbid, and many risk loci harbor associations to multiple AID. However, causal variant identification is confounded by linkage disequilibrium (LD), so it remains difficult to conclude that these associations represent true pleiotropic effects. As most GWAS associations are driven by alterations to gene regulation, we still lack systematic approaches to identify pathogenic genes. To address this problem, we have developed a systematic approach to identify regulatory regions mediating

risk in associated loci. We first fine-map each association to identify credible interval (CI) variants likely to be causal, then overlay these on to regulatory regions identified by DNase I hypersensitivity (DHS) in the Roadmap Epigenomics Project to calculate the posterior probability of a DHS mediating risk. We depart from standard practice by first clustering DHS peaks across 56 replicated tissues, and identifying clusters of DHS peaks that statistically replicate across samples. We then correlate the accessibility of these peaks in each tissue to expression levels of nearby genes. By combining the correlation between DHS and gene expression and the posterior probability of a DHS mediating risk, we calculate a posterior probability of association for each gene in the region. We use this method to dissect 24 loci with strong evidence that risk of at least two AID diseases is mediated by DHS. Within these loci, we found that the most highly associated SNPs are often different between disease associations, but our model identifies the same genes as causal (Fisher's exact test p=0.014). One striking example is the BACH2 locus on chromosome 6q15, where our analysis prioritizes MDN1 for autoimmune thyroid disease, multiple sclerosis, type 1 diabetes, and possibly for celiac disease, but a distinct effect on MAP3K7 for IBD, despite the strongest associations being present in an intron of the BACH2 gene. Thus, our approach can overcome the lack of resolution inherent in genetic mapping to resolve shared associations across traits by considering the likely functional effects in a locus.

1982T

Insights into differentiation and diseases of blood cells from (epi)genomics. *R.C. Hardison[1], Y. Zhang[1], C.A. Keller[1], B. Giardine[1], L. An[2], F. Yue[2], E. Hueston[3], D. Bodine[3].* 1) Departments of Biochemistry and Molecular Biology and of Statistics, Pennsylvania State University, University Park, PA; 2) Department of Biochemistry and Molecular Biology, Pennsylvania State University, Hershey, PA; 3) Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Statement of purpose: Differentiation of multipotent stem and progenitor cells into mature cells of different lineages proceeds by the precisely regulated expression of distinct cohorts of genes defining each cell type. Genetic variants in the DNA sequences that regulate gene expression (cis-regulatory modules or CRMs), such as enhancers and promoters, are strongly associated with complex traits including disease susceptibility. Thus it is important to identify CRMs reliably and connect variants in them to phenotypic effects. **Methods used:** We have generated genome-wide datasets for transcriptomes (RNA-seq) and the regulatory landscape (chromatin accessibility revealed by ATAC-seq) for mouse primary hematopoietic stem and multilineage progenitor cells. These features plus specific transcription factor (TF) binding and histone modifications were determined in mature cells and informative cell lines. All these data plus complementary data from other laboratories and consortia (including chromatin interaction frequencies) are being integrated using formal statistical modeling. One approach identifies chromatin states in two dimensions, along chromosomes and across cell types, using our new Integrative and Discriminative Epigenome Annotation System (IDEAS). This method substantially improves the identification of cell type-specific changes in inferred regulatory activity. **Summary of results:** Further analysis of the primary and integrated data gives new insights into relationships among cell types, global predictions of common and cell type-specific CRMs, prediction of gene targets for regulation, and identification of groups of transcription factors acting at a common site. Two examples of insights gleaned are described here. (1) We find limited changes in the regulatory landscape among stem and progenitor cells, whereas the landscape and transcriptional profile changes dramatically in the erythroid lineage, perhaps during commitment. Megakaryocytes are surprisingly close to multilineage progenitor cells. (2) Experimental tests of CRMs predicted by different criteria emphasize the importance of binding by key transcription factors or co-activators, whereas more general features such as histone modification profiles and chromatin accessibility give more false positives. Ongoing work aims to define quantitative, multi-CRM models for regulation that apply globally.

1983F

Non-coding sequence variants within cell type restricted DNA hypomethylated regions are significantly associated with tissue specific phenotypes. *E. Hodges[1], R.T. Levinson[1], D.C. Rinker[1], B. Chen[1], D.C. Samuels[1], A.D. Smith[2], J.A. Capra[1].* 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Molecular and Computational Biology section, Division of Biological Sciences, University of Southern California, Los Angeles, CA.

The use of functional genomic data is crucial to drawing biological links between non-coding sequence variants identified in Genome-wide association studies (GWAS) and their associated phenotypic states. In particular, the disruption of gene regulatory sequences has been experimentally shown to contribute to the mis-expression of genes in certain cellular contexts, implying that SNPs located around gene regulatory elements become functionally relevant within a limited cell-type context. However, these expression changes can be subtle and are often subject to developmental timelines, making the links between sequence disruption and phenotype difficult to detect. We have shown that patterns of non-coding DNA methylation, and specifically discrete regions of hypomethylation, distinguish the identities of different cell-types. Moreover, comparative profiling of intergenic and intronic hypomethylation (iHMRs) may be used not only to index putative enhancer elements, but, unlike DNase hypersensitivity and other measures of chromatin state, may also reveal the most functionally restricted subset of enhancers for the cell type in question. We hypothesized that SNPs within these regions are more likely to exhibit non-pleiotropic links to phenotype, compared to SNPs in iHMRs broadly shared among tissues. As proof of concept, we performed comparative genome-wide methylation profiling to identify a highly specific set of iHMRs in primary human lymphocytes, and intersected these regions with the NHGRI catalogue of GWAS SNPs. Among all HMRs identified, the lymphocyte specific iHMRs were highly enriched for GWAS SNPs (1-Binomial Dist., 0.0006) and were associated with immunological and inflammatory disorders. We then rationalized that SNPs located in lymphocyte iHMRs were more likely to reveal associations with immunological phenotypes than other traits. Thus, taking the inverse approach to GWAS, we performed Phenome-wide association studies (PheWAS) for all genotyped SNPs proximal to lymphocyte iHMRs, and used aggregated ICD9 codes from the electronic medical record to approximate phenotypes for ~5500 individuals. Our approach revealed numerous novel, high ranking SNPs (p-value <0.0005) associated with immune-related phenotypes, including lymphocytic neoplasms and deficiencies in humoral immunity. We demonstrate the power of combining PheWAS with epigenetic information to detect biological links between non-coding SNPs and their associated phenotypes.

1984W

Insights into DNA methylation changes in placental tissue and cells during chorioamnionitis. *C.R. Konwar[1,2], M.E. Price[1,2], M. Penaherrera[1,2], A. Beristain[2,3], W.P. Robinson[1,2,3].* 1) Dept. of Medical Genetics, University of British Columbia, Vancouver, British Columbia , Canada; 2) Child and Family Research Institute, British Columbia, Vancouver; 3) Dept. of Obstetrics and Gynecology, University of British Columbia, Vancouver BC.

Introduction: One of the major causes of spontaneous preterm birth is inflammation of the placenta and associated membranes, known as chorioamnionitis (CA). Identification of DNA methylation (DNAm) signatures in the placenta that reflect in utero conditions may provide candidate markers for earlier diagnosis of inflammation. Inflammation in the placenta is associated with an increase in placental-specific macrophages, Hofbauer cells (HCs). DNAm signatures differ strikingly between different cell lineages and can change with i) gestational age and ii) placental pathology. We proposed that increases in HC number might be detectable in placentas affected by CA by measuring placental DNAm. Methods: Pooled placental samples (chorionic villi) isolated from multiple biopsies taken from the fetal side of the placenta from 12 CA cases and 12 preterm births without other pathology (controls) were run on the Illumina HumanMethylation450 BeadChip array (450K array). Additionally, HC cells isolated from 5 independent term samples were profiled using the 450k array. Results: We identified 18 differentially methylated sites associated with CA status based on i) statistical significance (False Discovery Rate, <5%); and ii) biological significance (magnitude of methylation difference, Δβ>0.1). The majority of these sites were associated with immune-system genes including *Orosomucoid 1* which is known to increase in response to infection. When HCs were compared to chorionic villi we identified 10,739 differentially methylated sites. Three of the 18 CA-associated sites were also differentially methylated in the same direction in this chorionic villi to HC comparison. Additionally, reference-free cell type correction algorithm, an R package tool, identified 7 of 18 CA-associated sites as cell type-associated DNAm signatures. Conclusion: DNAm profiling can be useful to understand the aetiology of placental inflammation. CA-associated placentas showed altered DNAm signatures which are not observed in the absence of inflammation. These changes are consistent with changes in innate immune function and may reflect both increased numbers of immune cells as well as altered function.

1985T

Application of chromatin profiling by ATAC-seq to understand the functional genomic landscape of psoriasis. *A. Lledo Lara[1], A. Sanniti[1], I. Pulyakhina[1], M.H. Al-Mossawi[2], D. Gutowska-Owsiak[3], P. Bowness[2], A. Berlanga-Taylor[1,4], G. Ogg[3], J.C Knight[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) The Botnar Research Centre, NDORMS, University of Oxford, Oxford, United Kingdom; 3) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 4) St Marys' Hospital, Imperial College London, London, United Kingdom.

Psoriasis is a chronic inflammatory skin disease which may be associated with chronic arthritis manifesting as psoriatic arthritis (PsA). Both are complex diseases with strong heritability and genetic association, notably involving HLA-C*06:02. Genome-wide association studies (GWAS) have also identified more than 40 additional loci, the majority involving non-coding regulatory variants. Understanding the functional role and impact of these variants in disease aetiology still remains challenging, requiring application of bioinformatics and experimental approaches in a disease setting and cell-specific relevant context to resolve functional causal variants. We have applied transposase-accessible chromatin (ATAC-seq) profiling to keratinocyte-enriched cell populations obtained from epidermal sheets of uninvolved and lesional psoriatic skin biopsies. ATAC-seq results show nucleosomal periodicity in fragment size distribution and enrichment in overlap with DNase-I hypersensitivity ENCODE data from primary cultured keratinocytes relative to other cell types. However, we find high background limiting further application of this approach in this setting despite optimisation. We are now extending the approach to generate complementary ATAC-seq data for different immune cell types (including CD14+ monocytes, CD4+, CD8+ and CD19+) derived from synovial fluid and/or peripheral blood of psoriasis vulgaris and PsA patients. We conclude that the low sensitivity of ATAC-seq in skin currently impairs the use of the data for comparison of regulatory landscapes across lesional skin and different immune cell types. Integration of chromatin and transcriptional profiles generated in primary immune cells will be potentially informative in revealing cell and context specific differences in the regulatory landscape and it may help to prioritise GWAS variants associated with disease. Overall, this novel approach could contribute to better understand the mechanisms of action of non-coding variants in a disease and cell specific manner.

1986F

***In utero* tobacco smoke exposures and DNA methylation in a study of asthma in Latino children.** *A.M. Neophytou[1], S.S. Oh[2], D. Hu[2], C. Eng[2], S. Huntsman[2], F. Lurmann[3], L.N. Borell[4], S. Sen[5], H.J. Farber[6], K. Meade[7], P.C. Avila[8], D. Serebrisky[9], W. Rodriguez-Cintron[10], R. Kumar[11,12], J.R. Rodriguez-Santana[13], J.R. Balmes[1,2], E.A. Eisen[1], E.G. Burchard[2,14].* 1) SPH, Environmental Health Sciences , University of California Berkeley, Berkeley, CA; 2) Department of Medicine, University of California San Francisco, San Francisco, CA; 3) Sonoma Technology Inc., Petaluma, CA; 4) Department of Health Sciences, , Lehman College, CUNY, New York, NY; 5) Department of Biostatistics, University of California San Francisco, San Francisco, CA; 6) Department of Pediatrics, Baylor College of Medicine, Children's Hospital, Houston, TX; 7) Children's Hospital and Research Center Oakland, Oakland, CA; 8) Department of Medicine, Northwestern University, Chicago, IL; 9) Pediatric Pulmonary Division, Jacobi Medical Center, Bronx, NY; 10) Veterans Caribbean Health Care System, San Juan, Puerto Rico; 11) Department of Pediatrics, Northwestern University, Chicago, IL; 12) The Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 13) Centro de Neumologia Pediatrica, CSP, San Juan, Puerto Rico; 14) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA.

Maternal tobacco smoking during pregnancy has been linked to several health problems in children, including asthma. Minority children are not only disproportionately affected by asthma, but are also more likely to be exposed to asthma risk factors such as *in utero* tobacco smoke exposures. We investigated the relationship between self-reported maternal smoking during pregnancy (*in utero* smoking) and DNA methylation at 26 specific CpG sites previously linked to *in utero* tobacco smoke exposures, and further examined the association between methylation at these sites and asthma status in a sub-sample of 573 Latino children from the Genes-environments & Admixture in Latino Americans study (GALA II). We assessed associations between percent methylation and self-reported exposures in linear regression models adjusted for age, sex, ethnicity, recruitment region and cell-type heterogeneity, and used logistic regression models to examine associations between asthma status and methylation adjusting for the same covariates. The relative methylation difference between exposed and unexposed at 25 out of 26 CpGs was in the same direction as previously reported, and was statistically significant at 14 of the CpG sites. Methylation was associated with asthma status for six CpGs; five of the six sites were also a subset of the 14 sites where self-reported maternal smoking was found to be a significant predictor. The largest magnitude association was within the *EXT1* gene, which encodes a protein found in the Golgi apparatus. A 1% increase in methylation at a CpG site within this gene corresponded to an 18% reduced odds for asthma in our population (odds ratio = 0.82, 95% confidence interval: 0.75– 0.90). Methylation of another gene involved in the regulation of detoxification of compounds found in tobacco smoke, *AHRR,* had three different locations within the gene that were also associated with asthma status in our study population. A 1% increase in methylation at any of these three *AHRR* CpG sites corresponded to an approximate 10% reduced odds for asthma. The above results were similar in a sensitivity analysis adjusting for global genetic ancestry. Our results suggest that DNA methylation is likely an important mechanism for the adverse effects of *in utero* tobacco smoke exposures on asthma.

1987W

Epigenetic regulation of innate immune responses. *A. Pacis[1,2], L. Tailleux[3], F. Mailhot-Léonard[1,2], V. Yotova[1], A. Dumaine[1], J.C. Grenier[1], L.B. Barreiro[1,4].* 1) CHU Sainte-Justine Research Center, Department of Genetics, Montreal, Quebec, Canada; 2) Université de Montréal, Department of Biochemistry, Montreal, Quebec, Canada; 3) Institut Pasteur, Mycobacterial Genetics Unit, Paris, France; 4) Université de Montréal, Department of Pediatrics, Montreal, Quebec, Canada.

The first immune mechanisms recruited to defend against invading pathogens are those associated with innate immune cells, such as dendritic cells (DCs) or macrophages. Once they sense an intruder, these cells rapidly induce sophisticated transcriptional programs, which are regulated through a series of epigenetic changes related to histone modifications. Yet, we still know remarkably little about other epigenetic mechanisms, notably DNA methylation, in controlling transcriptional responses to infection. This is due to the belief that methylation marks are highly stable, and unlikely to respond to environmental perturbations on a short time scale. Recently, previous work from our laboratory, has shown that DNA methylation patterns can rapidly and actively change (in contrast to passive changes during cell replication) in response to infection and, are strongly associated with differentially expressed genes induced by infection. However, it still remains unclear whether changes in methylation are a cause or consequence of changes in gene expression during infection. Here, we used a combination of genomic and statistical approaches to study the causal relationship between dynamic changes in DNA methylation and changes in gene expression, during infection. Specifically, we infected DCs with a pathogenic strain of Mycobacterium tuberculosis (MTB), the causative agent of tuberculosis (TB) in humans, at different time-points (2, 18, 24, 48 and 72 hours). Following infection, we collected DNA methylation data across thousands of target regions, in non-infected and MTB-infected DCs, using capture-based bisulfite sequencing. In parallel, we collected RNA sequencing data to be able to characterize genome-wide patterns in gene expression. We observe that changes in gene expression tend to occur *prior* to detectable changes in DNA methylation. Moreover, we observe that loss of methylation in a handful of CpG sites remain stable even if expression of the associated gene reverts back to basal state. This work will yield unprecedented insights on the role of active changes in DNA methylation in the control of transcriptional responses to infection and ultimately, may lead to the implementation of novel preventive strategies against infectious diseases and other immune-related disorders.

1988T

Defining the cell specific transcriptional and regulatory landscape in patients with ankylosing spondylitis. *A. Sanniti[1], I. Pulyakhina[1], A. Lledó Lara[1], K. Plant[1], G. Scozzafava[1], K. Doig[2], J. O'Donoghue[2], P. Bowness[2], J. Knight[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Botnar Research Centre, NDORMS, University of Oxford, United Kingdom.

Ankylosing Spondylitis (AS) is an inflammatory disease with high genetic heritability, the strongest genetic association being with HLA-B27. Genome Wide Association Studies have now identified at least 43 independent associations throughout the genome, but the understanding of how these associations relate to genes involved in disease aetiology are currently largely undetermined. We have begun to answer this question by recruiting AS patients through the GENExpressID study (Genetics of Gene Expression in Immune-related Diseases) for RNA-seq and ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) analysis. We present data for 10 AS patients and controls, sequenced for mRNA and miRNA in up to 6 immune cell types. This data defines transcriptional activity in CD14+ monocytes, CD4+, CD8+, CD19+, CD16+ and NK cells, the role of which in AS remains unresolved. We integrate this transcriptomics data with chromatin profiling for AS patients, defining the disease specific regulatory landscape compared to healthy matched controls demonstrating the effects before and after starting biologic therapy. This integrated approach will aid in prioritising variants associated with disease, and advance our understanding of the underlying mechanisms of action.

1989F

Converging disease genes in ICF syndrome: *ZBTB24* **controls expression of** *CDCA7* **in mammals.** *S.M. van der Maarel[1], H. Wu[1], P.E. Thijssen[1], E. de Klerk[1], K.K.D. Vonk[1], J. Wang[1,2], B. den Hamer[1], C. Aytekin[3], L. Daxinger[1].* 1) Human Genetics; S4-P, Leiden University Medical Center, Leiden, Netherlands; 2) Institutes of Biology and Medical Sciences, Soochow University, Suzhou 215123, China; 3) Department of Pediatric Immunology, Dr Sami Ulus Maternity and Children's Research and Educational Hospital, Ankara, Turkey.

For genetically heterogeneous diseases a better understanding of how the underlying gene defects are functionally interconnected will be important for dissecting disease etiology. The Immunodeficiency, Centromeric instability, and Facial anomalies (ICF; OMIM 242860/614069) syndrome is an autosomal recessive disorder characterized by immunodeficiency, developmental delay and facial anomalies. Based on the underlying genetic defects, ICF syndrome can be divided into at least five different clinically indistinguishable subgroups (ICF1, ICF2, ICF3, ICF4 and ICFX). About half of the ICF patients carry mutations in the *de novo* DNA methyltransferase 3B gene (*DNMT3B*), which is referred to as ICF1. ICF2 patients carry mutations in the zinc-finger- and BTB domain containing 24 gene *ZBTB24* and account for ~30% of the ICF cases. We have recently shown that mutations in cell division cycle associated 7, *CDCA7*, and helicase, lymphoid-specific, *HELLS/LSH*, are causative for the ICF3 and ICF4 subtypes respectively, and together with few unexplained cases (ICFX) account for the remaining ICF cases. While DNMT3B and HELLS are cooperatively necessary for establishment of DNA methylation during early embryogenesis, the functional connectivity between the remaining ICF genes remains unknown. We generated a Zbtb24 BTB domain deletion mouse and found that loss of functional Zbtb24 leads to early embryonic lethality. Transcriptome analysis identified *Cdca7* as the top down-regulated gene in *Zbtb24* homozygous mutant mESCs, which can be restored by ectopic ZBTB24 expression. We further demonstrate enrichment of ZBTB24 at the *CDCA7* promoter suggesting that ZBTB24 can function as a transcription factor directly controlling *Cdca7* expression. Finally, we show that this regulation is conserved between species and that *CDCA7* levels are reduced in patients carrying *ZBTB24* nonsense mutations. Together, our findings demonstrate convergence of the two ICF genes *ZBTB24* and *CDCA7* at the level of transcription. We suggest that ICF1 and ICF4 might represent disorders of a failure in DNA methylation establishment, while ICF2 and ICF3 might be caused by a failure in DNA methylation maintenance.

1990W

Discovery of gene-networks governed by two Ikaros transcription factors highlight their function as master regulators of lymphocyte homeostasis, and suggest a pivotal role in SLE pathology. *C. Odhams[1], A. Cortini[1], L. Chen[1], A.L. Roberts[1], D.L. Morris[1], T.J. Vyse[1,2], D.S. Cunninghame Graham[1,2].* 1) Division of Genetics and Molecular Medicine, King's College London, London, UK; 2) Division of Immunology, Infection and Inflammatory Disease, King's College London, UK.

Recent genome-wide association and ensuing functional studies document three Ikaros family transcription factors (TFs) (*IKZF1*, Ikaros; *IKZF2*, Helios; *IKZF3*, Aiolos) as susceptibility genes in Systemic Lupus Erythematous (SLE) a complex autoimmune disease with undefined aetiology. Expression quantitative trait loci (eQTL) analysis show that SLE-associated variants disrupt transcription of these genes in *cis*, and it is postulated that this may be a pathogenic mechanism as the encoded proteins act as master controllers of gene expression throughout lymphocyte development and signalling. In order to understand more fully the underlying biology of the Ikaros TFs and elucidate their regulatory networks, we performed, to ENCODE quality standards, ChIP-Seq of Helios and Aiolos in lymphoblastoid cell lines (GM12878). We integrated genome-wide binding positions with RNA-Seq expression data derived from independent siRNA mediated knockdown of the two genes to predict whether binding of each factor had either an activating or repressing gene expression activity at each direct or indirect target. For Helios and Aiolos respectively, we defined 3,736 and 7,067 significant peaks (*IDR*<0.02), and 249 and 557 differentially expressed genes (*FDR*<0.01) from knock-down relative to control. Coupling gene-network analysis showed significant enrichment of direct target genes in autoimmune relevant pathways such as TLR signaling (*P*=4.5E[-10]) and FcγR mediated phagocytosis (*P*=4.9E[-10]). Isolation of discrete pathways verified the importance of Helios and Aiolos as master regulators in lymphocytes. In particular, we observed that knock-down of Aiolos increased expression of *BTK* (Bruton tyrosine kinase), a known SLE susceptibility gene involved in pro-inflammatory cytokine production through TLR9 signaling, and antibody maturation in B-cells. Using our Aiolos ChIP-Seq data, we identified a binding site in the *IBTK* promoter the natural inhibitor of *BTK*, suggesting decreased expression of *IKZF3* diminishes cellular *IBTK* which leads to aberrant TLR9 and B-cell receptor signaling. By mimicking the functional effect of SLE-associated variants through gene knockdown of *IKZF2* and *IKZF3* coupled with ChIP-Seq/RNA-Seq, we have identified the target genes and regulatory networks governed by these TFs in lymphoblastoid cells. The biological significance of these data will be the subject of targeted follow-up studies as potential markers of therapeutic intervention.

1991T

Identification of sex-biased expression and expression quantitative trait loci (eQTLs) in innate and adaptive immunity. *M. Oliva[1,2], C. Czysz[1,2], B.E. Stranger[1,2].* 1) The University of Chicago, Institute for Genomics and Systems Biology, Chicago, IL; 2) The University of Chicago, Department of Medicine, Section of Genetic Medicine, Chicago, IL.

Human males and females exhibit sexual dimorphism in a wide range of phenotypes. Several diseases display sex biases in prevalence, clinical features and prognosis, e.g. autoimmune diseases are more common in females and infectious diseases in males. Moreover, disease-related phenotypes behave in a sex-dependent manner. Herein, we characterize the influence of sex in determining inter-individual transcriptional variation in cells representing the adaptive and innate branches of the immune system in cohorts of different ancestry. We aim to: a) identify genes and regulatory networks differentially expressed (DE) between sexes, b) identify and characterize sexually dimorphic expression quantitative trait loci (eQTLs), c) pinpoint differences in sex-biased eQTL profiles between autosomal and sex chromosomes and d) determine the variation of sex biases across immune cell types and their contribution to disease. We have analyzed genotype and gene expression data from naïve T lymphocytes and monocytes in healthy Caucasian-American, Asian-American and African-American populations (Immune Variation project, Raj et al. 2014), to discover genes and pathways regulated differently between healthy males and females. We identified 6% and 7% genes that exhibit significant (FDR<5%) sex-biased expression in naïve T-cells and monocytes, respectively. Sex-biased gene sets are enriched in X-linked genes (F-test p<2×10-16) and also in pathways related to apoptosis, immunity, infectious and autoimmune diseases. Several immune pathways mediated by hormones (e.g. prolactin) exhibit sex bias, highlighting how sex-specific hormone profiles explain differences between male and female immunity. Next, to identify genetic variation that regulates gene expression differently between sexes, for each cell type and population we have modeled transcript abundance as a function of sex, genotype and the interaction of both, testing for significance of the latter as sex-biased eQTLs. To increase the statistical power of the association analysis, we meta-analyzed populations for each cell type. We are characterizing: a) whether genes that are DE between sexes demonstrate sex biases in the genetic basis of their regulation and b) differences in sex-biased regulation profiles between sex and autosomal chromosomes. We observe a strong X chromosome enrichment for sex-biased regulation. Finally, we are exploring the contribution of sex-biased eQTLs to disease through integration with GWAS results.

1992F

Hierarchical relationships among susceptibility genes, transcription factors and differentially expressed genes for systemic lupus erythematosus and signaling modules revealed by a systems approach. *T.Y. Wang, Y.F. Wang, Y. Zhang, J.J. Shen, M. Guo, J. Yang, Y. Lau, W. Yang.* Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Hong Kong SAR, China.

Systemic lupus erythematosus (SLE) is a prototype autoimmune disease with complex pathogenesis and manifestations. Recent advancement in genome-wide association studies (GWAS) has markedly improved our understanding of the genetic basis of SLE, but the details of the disease mechanisms remain elusive. An increasing amount of evidence supports the important role of gene expression in the etiology of the disease. Understanding the functional relationship between genetic findings and gene expression changes may help us better fathom the complicated disease mechanisms. Making use of public data from NCBI Gene Expression Omnibus (GEO) on a number of different studies on SLE expression profiles, we have identified 440 differentially expressed genes (DEGs) in three different cell types through gene expression meta-analysis, which is so far the most comprehensive analysis of gene expression in SLE. Functional analysis of these genes showed that up-regulated genes were mainly involved in immune and inflammatory responses, type I interferon-mediated signaling and responses to viral infection. Through integrative analyses of state-of-the-art, high-throughput biomedical data, such as expression quantitative trait loci (eQTLs), transcription factor binding sites(TFBS), and protein-protein interactions(PPIs), we identified a hierarchical regulatory system, consisting of disease-associated genes (DAGs) in the top layer, involved TFs in the middle layer and DEGs in the bottom layer. Furthermore, modularization analysis found pathways such as cell cycle regulation, apoptosis and mRNA surveillance are closely implicated in SLE pathogenesis. Our results provided a general framework to bridge GWAS findings and gene expression changes, and expanded our understanding of gene expression regulation and its roles in SLE pathogenesis.

1993W

SNPs in *NRM* and *GPANK1* miRNA target regions are associated with risk of pediatric-onset multiple sclerosis. *L.F. Barcellos[1], B. Rhead[1,2], X. Xhao[1], H. Quach[1], D. Quach[1], R. Metlapally[3], J. Hart[4], J. Graves[4], E. Eggers[4], H.C. Von Buedingen[4], A. Waldman[5], T. Lotze[6], T. Schreiner[7], B. Greenberg[8], B. Weinstock-Guttman[9], G. Aaen[10], J.M. Tillema[11], J. Ness[12], M. Candee[13], L. Krupp[14], M. Gorman[15], L. Benson[15], T. Chitnis[16], S. Mar[17], I. Kahn[18], C. Casper[19], S. Roastadt[19], J. Rose[19], C. Schaefer[20], E. Waubant[4].* 1) Genetic Epidemiology and Genomics Laboratory, UC Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, UC Berkeley, Berkeley, CA; 3) School of Optometry, UC Berkeley, Berkeley, CA; 4) Department of Neurology, University of California, San Francisco; 5) Children's Hospital of Philadelphia, Philadelphia, PA; 6) Texas Children's Hospital, Houston, TX; 7) University of Colorado School of Medicine, Aurora, CO; 8) University of Texas Southwestern, Dallas, TX; 9) SUNY Buffalo, Buffalo, NY; 10) Loma Linda University, Loma Linda, CA; 11) Mayo Clinic, Rochester, MN; 12) University of Alabama at Birmingham, Birmingham, AL; 13) University of Utah, Salt Lake City, UT; 14) SUNY Stony Brook, Stony Brook, NY; 15) Boston Children's Hospital, Boston, MA; 16) Brigham and Women's Hospital, Boston, MA; 17) Washington University St. Louis, St. Louis, MO; 18) Children's National Medical Center, Washington, DC; 19) Network of Pediatric Multiple Sclerosis Centers Data Coordinating and Analysis Center, University of Utah, Salt Lake City, UT; 20) Kaiser Permanente Division of Research, Oakland, CA.

Multiple sclerosis (MS [MIM 126200]) is a disabling chronic autoimmune disease that affects the central nervous system, with onset occurring in children in as high as 10% of cases ('pediatric-onset MS'). Epigenetic influences are strongly implicated in MS pathogenesis, including the contribution of micro (mi) RNAs, which are small non-coding RNAs that can alter gene expression. Further, genetic polymorphisms (SNPs) can affect miRNA functions in different ways depending on where they are located. These miRNA-related SNPs (miR-SNPs) can impact miRNA expression levels or influence the miRNA-target interaction. Resulting gene dysregulation may affect both susceptibility and phenotype. We investigated the role of miRNAs in ~36,000 controls drawn from UCSF and the Kaiser Permanente Northern California (KPNC) membership and 700 pediatric-onset MS cases recruited through the U.S. Pediatric MS Network, UCSF, and the KPNC MS Research Program. We tested 144 SNPs in regions encoding miRNAs (from the mirBase v21 database) and 28,764 SNPs in the 3' UTRs of target genes (from the MirSNP and polymiRTS 3.0 databases) for association with pediatric-onset MS in White, non-Hispanic cases (N=394) and controls (N=10,875). No SNPs in regions encoding miRNAs were associated. SNPs in miRNA target genes were tested in two separate models: the first for 28,473 SNPs outside of the major histocompatibility complex (MHC), adjusted for genetic ancestry, and the second for 291 SNPs within the MHC, adjusted for genetic ancestry and *HLA-DRB1*15:01*, the strongest genetic risk factor for MS. Two target genes in the MHC were significantly associated with pediatric-onset MS after correcting for multiple tests: *NRM* in the MHC I region on ch. 6p21 (OR=2.7, 95% CI 1.8-4.2, p=3.1e-06) and *GPANK1* in the MHC III region on ch. 6p21 (OR=2.2, 95% CI 1.4-3.3, p=3.4e-04). In addition to miR-SNP analysis, we examined expression levels of miRNAs in four sorted peripheral immune cell types from 8 treatment-naive MS female cases and 8 controls matched on age and race/ethnicity. Expression of 1,979 miRNAs was measured using the mirBase 19 multispecies microRNA microarray (Microarrays INC; Huntsville, AL). Notably, one miRNA, miR-4668, demonstrated the highest fold-change (3.7) in cases compared to controls in CD4+ T cells. CD4+ T cells are strongly implicated in MS. Our results provide the first evidence for miRNA regulation in pediatric-onset MS.

1994T

Genetic variants underlying histone modifications explain a large fraction of autoimmune disease heritability. *B. van de Geijn[1], H. Finucane[1], S. Gazal[1], A. Gusev[1], X. Liu[1], Y. Reshef[1], A. Price[1,2,3].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.

Chromatin state, marked by histone modifications, is an important driver of gene regulation. However, the genetic control of histone modifications is not well understood. We partitioned the local heritability (+/- 10kb from peak) of several histone modifications, including H3K27ac, H3K4me3 and H3K4me1, by extending stratified LD score regression to jointly analyze a large number of histone modification phenotypes, using data previously collected in 65 Yoruba lymphoblastoid cell lines (LCLs). We find the most striking enrichment for heritability in the overlap of H3K27ac and DNase sensitivity peaks, where 1.6% of the SNPs in the genome explain 52% of histone modification SNP-heritability across 136,264 H3K27ac peaks (33x enrichment, P = 2.1 x 10^{-23}). These likely represent active regulatory regions, as marked by H3K27ac, which are bound by transcriptional machinery, leading to DNase sensitivity.   We investigated whether these enrichments for histone modification heritability extend to complex traits by applying stratified LD score regression to GWAS summary statistics for 15 diseases (average N=30,789). The overlaps of H3K27ac and DNase sensitivity peaks were highly enriched for autoimmune diseases where LCLs are expected to be relevant, including celiac disease, Crohn's disease, lupus, and rheumatoid arthritis (1.6% of SNPs explain 33% of SNP-heritability on average; 21x enrichment, P = 1.3 x 10^{-8}). These enrichments were even larger than those for coding or conserved regions, making them vital for unraveling the causes of these diseases. We observed no significant enrichment for this annotation in non-autoimmune diseases, likely due to LCLs being the incorrect cell type for diseases that are not linked to the immune system. We go on to use histone modification heritability explained in LCLs, where most types of data are available, to prioritize combinations of functional data for use in cell types where only a subset have been collected. Finally, we use these combinations to test a wide range of cell type specific annotations for enrichment of disease heritability.

1995F

miRNA expression in patients with HIV/AIDS with antiretroviral resistance to Tenofovir/Emtricitabine/Efavirenz. *J. Marquez-Pedroza[1,2,6], J. Cárdenas-Bedoya[2,3], C. Morán-Moguel[3], M. Escoto-Delgadillo[2], E. Vázquez-Valls[2,4], M. Pérez-Ríos[5], M. Rosales-Reynoso[6], B. Torres-Mendoza[2,3].* 1) Doctorado en Farmacologia, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Laboratorio de Inmunodeficiencias y Retrovirus Humanos, Centro de Investigación Biomédica de Occidente, Centro Médico Nacional de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 3) Doctorado en Genética Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jalisco, México; 4) Dirección de Educación y de Investigación, U.M.A.E. Hospital de Especialidades, Centro Médico Nacional de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 5) Hospital Regional 110, Instituto Mexicano del Seguro Social, Guadalajara, México; 6) División de Medicina Molecular, Centro de Investigación Biomédica de Occidente, Centro Médico Nacional de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México.

   Background: The micro-RNAs (miRNAs) are a class of small non-coding RNA that modulates gene expression post-transcriptionally by any of the heretofore, known mechanisms: mRNA destabilization or translational repression. MiRNAs have been proposed as therapy for the treatment of diseases in which the expression of proteins plays a key role in sustaining and disease progression, as in the case of HIV-1 and to the which it is directed the action of some antiretrovirals. The failure of antiretroviral therapy affects control of HIV infection, this is secondary to the administration of antiretrovirals and HIV resistance primary or transmitted in patients who had never received antiretroviral therapy (treatment-naïve). The success of the highly active antiretroviral therapy (HAART) with TDF/FTC/EFV in HIV patients could be related to an expression of miRNAs when compared with HIV-1 patients with and without resistance to HAART. The aim of this study was to compare the expression of miRNAs in HIV/AIDS resistance or not to the scheme TDF/FTC/EFV. **Methods:** A cross-sectional study was made using 13 medical records and 13 plasma samples from Northwestern Mexican patients with HIV/AIDS. They were involved in three groups: antiretroviral resistance to TDF/FTC/EFV (5); antiretroviral without resistance (5); and healthy subjets without HIV-1 infection as control (3). RNA viral was used to genotyping antiretroviral resistance test, with Stanford University criteria. To obtain miRNAs from plasma, first total RNA was extracted, the cDNA synthesis was performed, and the expression of 84 miRNAs detectable and differentially expressed in plasma in each subjects. The Human Serum and Plasma miScript miRNA PCR Array Qiagen® was used. The stably expressed normalization candidates for the evaluation by qRT-PCR were hsa-miR-106b-5p, hsa-miR-17-5p, hsa-miR-18a-5p, hsa-miR-148-3p and was made using a synthetic spike-in control. Student's t-test was used to examine differential expression of miRNAs between groups. **Results:** The analysis of MiRNAs expression in HIV infected treated patients with resistance versus non-resistance shown that miRNAs hsa-miR-15b-5p, hsa-miR-126-3p, hsa-miR-20a-5p, hsa-miR-26a -5p and hsa-miR-16-5p were differentially expressed (p < 0.05). **Conclusion:** miRNA should be considered in future studies as potential biomarkers of resistance to HAART. .

1996W

miRNA and mRNA profiling during disease development in an animal model of multiple sclerosis. *S.G. Gregory[1,2], S.K. Siecinski[1], C.C. Zhao[1], W.R. Morgenlander[1], D. Corcoran[3], L. Kwee[1], S. Arvai[1].* 1) Medicine, Duke Molecular Physiology Institute, Durham, NC; 2) David H Murdock Research Institute, Kannapolis, NC, USA; 3) Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC, USA.

   Purpose: Multiple sclerosis (MS) affects ~400,000 individuals in the U.S. and 2.5 million worldwide. MS is triggered by autoimmune dysfunction that causes demyelination and neuronal damage. Here, we describe the use of a mouse model of MS, experimental autoimmune encephalomyelitis (EAE), to profile microRNA (miR) and messenger RNA (mRNA) expression during EAE development. **Methods:** EAE was induced in C57BL/6J wt mice and severity quantified according to clinical scores (CS) of caudal to rostral paralysis. miR/mRNA from whole blood were profiled at increasing CS from baseline, 0.5-1.5, 2-2.5, 3-3.5. miRNAs were profiled using 50bp single end sequencing, processed using Bowtie alignment to GRCm38, and quantified after log2 transformation. cRNA libraries were hybridized to Agilent Mouse GE v2 arrays with GeneSpring used to normalize data and calculate differential expression. ANOVA was used to identify differentially expressed miR and mRNA across all clinical scores, and a moderated t-statistic in *limma* was used to identify expression changes between groups. **Results:** mir-149-5p (FDR p-val 0.002), which is down regulated by IFN-β therapy and CD8+ T-cells of MS patients and modulated by TNFα in macrophages, was upregulated as each time point. KEGG identified differentiation of miRs between early and late EAE with key MS pathways (Wnt, TGF-Beta, axon guidance) emerging as EAE progressed. mRNA analysis identified cytokine regulation, immune and defense responses as significant, with cellular and tissue components showing a gradual transition into neuronal and myelination associated genes. Analysis of mRNA between baseline and 0.5-1.5 identified connective tissue breakdown of the blood brain barrier, inflammatory/immune response. Multiple MS GWAS study genes were differentially expressed, including IKZF3, which is implicated in lymphocyte development and T-cell mediated apoptosis. Interestingly, the LXR/RXR pathway activation was significant, the importance being the recent association of LXR (NR1H3) with PP-MS. Correlation analysis at the same time points identified differential expression of 9 miRs and 14 validated mRNA targets implicated cytotoxic T-cell mediated apoptosis pathways, with 9 known miRNAs dysregulated in EAE. **Conclusions:** Our analysis identified novel quantitative changes in miR/RNA expression in etiologically relevant pathways throughout the course of EAE and describe a dynamic interrelationship of miR/mRNA expression in a model of MS.

1997T

Characterizing regulatory variation and asthma-associated SNPs in memory Th2 cells. *G. McVicker[1], J. Zapardiel[2], J. Cayford[2], B. White[2], G. Seumois[2], P. Vijayanand[2], B. Peters[2].* 1) Salk Institute for Biological Studies, La Jolla, CA; 2) La Jolla Institute for Allergy and Immunology, La Jolla, CA.

Genome-wide association studies (GWAS) have identified thousands of single nucleotide polymorphisms (SNPs) that are associated with autoimmune and inflammatory diseases. Since most of these GWAS hits lie outside of protein coding regions of the genome, they are often assumed to affect gene regulation in disease-relevant cell types. However, the mechanisms and cell types through which these SNPs act is typically unclear. We conducted a study to examine the role of regulatory genetic variation in memory T cells of the Th2 subset, which are known to be involved in autoimmune responses such as allergy and asthma. We genotyped 76 donors, and performed RNA-seq and H3K27ac ChIP-seq on Th2 memory T cells obtained from these individuals. We identified a total of 751 gene expression quantitative trait loci (eQTLs), several of which are also GWAS hits for asthma. A subset of these hits overlap with H3K27ac peaks that have an allele-specific bias toward one allele. For these GWAS hits, the eQTLs implicate specific genes and the H3K27ac peaks suggest specific enhancer sequences that are likely to be affected by the causal SNP. In summary we provide the first catalog of functional regulatory variation in Th2 memory T cells and identify enhancers that are likely to play a functional role in asthma.

1998F

Allele-dependent binding of a viral protein to autoimmune disease-associated genetic variants. *M. Weirauch[1,6], D. Miller[1], L. Kottyan[1], I. Ibarra[1], S. Syed[2], X. Chen[1], E. Zoller[1], A. Lynch[1], C. Schroeder[1], J. Lee[1], A. Magnussen[1], A. Yang[3], T. Hughes[3], J. Park[4], C. Vinson[2], J. Harley[1,5].* 1) Center for Autoimmune Genomics (CAGE), Cincinnati Children's Hospital, Cincinnati, OH; 2) Center for Cancer Research, National Cancer Instititue, Bethesda, MD; 3) Department of Molecular Genetics, University of Toronto, Canada; 4) Divisions of Urology and Developmental Biology, Cincinnati Children's Hospital, Cincinnati, OH; 5) US Department of Veterans Affairs Medical Center, Cincinnati, OH; 6) Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital, Cincinnati, OH.

Risk factors are known for many diseases, but the etiologies of most autoimmune diseases remain unknown and are idiopathic. Pathogenesis of disease likely involves complex interplay between genetic and environmental risk factors. Specifically, Epstein Barr virus (EBV) has suggestive associations with many autoimmune diseases, and EBV infection is nearly ubiquitous in adults. The molecular mechanisms underlying these associations, however, remain unclear. We tested the hypothesis that some autoimmune variants might act by altering the binding of the EBV-encoded transcription factor ZTA, consequently resulting in downstream changes in gene expression. To this end, we comprehensively characterized the DNA binding of ZTA to both methylated and unmethylated DNA sequences using protein binding microarrays (PBMs). Based on these data, we identified plausible causal variants for multiple sclerosis (MS), systemic lupus erythematosus (SLE), and juvenile idiopathic arthritic (JIA) predicted to alter ZTA binding. From among these, we identified variants located within likely regulatory regions in EBV-infected B cells using publically available functional genomic datasets. We screened these candidate variants using electrophoretic mobility shift assays (EMSAs) to identify general differential binding of nuclear factors, and validated differential ZTA binding using EMSA-supershift and DNA Affinity Precipitation Assays coupled with Western blots (DAPA-Westerns). These experiments revealed three genetic variants, associated with MS, SLE, and JIA, respectively, exhibiting stronger ZTA binding to the risk allele. We provide data showing that each of these variants is associated with genotype-dependent expression in EBV-transformed B cell lines. Using luciferase reporter assays, we further demonstrate that the autoimmune risk alleles result in greater promoter activity. Collectively, these data demonstrate for the first time that differential binding of a viral protein to a disease-associated genetic variant can result in altered levels of host gene expression in ways that are predicted to influence autoimmune disease risk of MS, SLE, and JIA. Since ZTA is a viral protein, and is expressed throughout human life subsequent to EBV infection, but only in virus infected cells, these results offer a potential therapeutic target for multiple autoimmune diseases.

1999W

Bloom's syndrome study by RNA-seq. *M.M. Montenegro[1], D.A.M. Silvestre[2], A. Rangel[3], G.M. Novo-Filho[1], E.A. Zanardo[1], A.T. Dias[1], A.M. Nascimento[1], T.V.M.M. Costa[1], F. Madia[1], Y. Gasparini[1], C.A. Kim[4], L.D. Kulikowski[1].* 1) Department of Pathology, Cytogenomic Laboratory, FMUSP, Sao Paulo, Brazil; 2) Department of Bioinformatics Interunit, Mathematic Institute of São Paulo University, São Paulo, Brazil; 3) Laboratory of Medical Investigation / LIM36, ICr-HC-FMUSP, São Paulo, Brazil; 4) Genetic Unity, Department of Pediatrics, Children Institute, HC-FMUSP, São Paulo, Brazil.

Bloom's syndrome (BS) is a rare autosomal recessive chromosome instability disorder principally characterized by pre and post natal growth deficiency, microcephaly, photosensibility, immunodeficiency and predisposition to develop neoplasias at an early age. It is caused by mutations in the BLM gene (RECQL3) which in normal conditions encodes the DNA repair protein BLM helicase, that functions in homologue recombination via. Gene expression changes in the functional absence of BLM are poorly understood. In this sense, the transcriptome profiling of mRNA could be a powerful approach to identify new transcripts, gene regulation, SNPs and indels variants that are relevant genomic markers that could be associated to Bloom's syndrome. We performed RNA-Seq profiling using the Illumina's HiSeq 2500 platform of samples derived from two patients with BS and three unaffected controls. The raw data analysis was generated and analyzed using specialized softwares (Subread, SAMtool, BEDtools, DESeq2's, VEP and DAVID database). The RNA-Seq assay revealed the precise location of transcription limits, with resolution of a single nucleotide and high level of efficiency, showing high genetic complexity. The differentially expressed (DE) analysis disclosed 216 genes in the group with BS, being most hipoexpressed genes related to immune and infection system via. The variant analysis detected 295.408 variants, being 186.239 (63%) substitutions and 109.169 (37%) SNVs for Patient 1 and 376.490 variants, being 231.947 (61,6%) substitutions and 144.543 (38,4%) SNVs for Patient 2. Indeed, we found high impact variants detected in some of DE genes. Our results suggested that gene expression network in BS could interfere in the regulation of the pathways associated with the immunological systems regulation probably caused by disturbance of DNA repair mechanisms. Furthermore, the study of the transcriptome using RNA-Seq may help to a breakthrough in the pathogenesis of BS.

2000T

SNPs from single-variant credible sets in inflammatory bowel disease can show allelic differences in expression in a reporter assay. *C.J. Cardinale, P.M. Sleiman, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Work by the International IBD Genetics Consortium (Huang et al.) has used dense genotyping on the Immunochip to elucidate 18 loci that could be fine-mapped to single variants with greater than 95% posterior probability of causality. Not surprisingly, 8 of these SNPs were coding variants due to their strong phenotypic impact. However, other loci were identified in which the causal SNP has no known motif or feature that suggests its mechanism of action, such as being contained within a transcription factor binding site, a histone modification region, or an expression quantitative trait locus. We have taken 8 loci, approximately 700 bp in size surrounding the credible SNP, and cloned both the reference and alternate allele into a luciferase reporter vector with a minimal promoter. Six of the loci did not show any expression of luciferase above the background level of the empty vector. However, both *JAK2* and *PRDM1* region SNPs showed significantly above-baseline expression as well as a difference in the expression level of the reference and alternate alleles. The *JAK2* SNP lies within the promoter region 556 bp from the transcriptional start site and showed a 20–30% allelic difference. These results suggest that credible SNPs can mediate differences in gene expression even when known functional annotations are lacking.

2001F

A set of human transcription factors (TFs) bind genomic DNA at multiple Sjögren's syndrome (SS) genetic loci. *J.B. Harley[1,2], M.T. Weirauch[1,3], X. Chen[1], M.J. Pujato[1], C.J. Lessard[4], K.L. Sivils[4], L.C. Kottyan[1].* 1) Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) US Department of Veterans Affairs Medical Center, Cincinnati, Ohio; 3) Divisions of Bioinformatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; 4) Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma.

   Purpose Dry eyes and mouth are the primary symptoms of Sjögren's syndrome, a chronic autoimmune disorder, mainly found in older women who often develop systemic complications. Genome wide association studies (GWASs) of Sjögren's syndrome (SS) have provided at least 17 convincing genetic associations, many of which suggest that variants involving inflammatory pathway genes are important in the mechanisms of pathogenesis. Since the vast majority of GWAS loci are in gene regulatory regions, we asked whether specific TFs were known to bind DNA at multiple genetic loci. **Methods** Logistically, we defined a locus as those variants which had disequilibrium at $r^2 > 0.8$ with the most significant marker at the locus. We assembled the chromatin immunoprecipitation with next generation sequencing (ChIP-Seq) datasets available in the public domain. We developed a simulation test for statistical significance. The null distribution is generated with 2000 records in which the linkage disequilibrium and allele frequency structure of all 17 SS loci were randomly assigned to genomic locations. The Bonferroni corrected probabilities (p_c) were calculated for each TF evaluated. We confined our attention to TF data sets that bound DNA in ≥4 SS genetic loci and that had $p_c < 0.00001$. **Results** Of the ~2000 TF datasets available from many cell types, 11 satisfied these criteria with p_c to ~10^{-15} and relative risk of 6.6 to 28.8 for the intersection of SS loci with TF binding peaks. Of the 11 TFs multiple members influence the NFκB pathway and interferon production. All 11 of these TFs datasets are from transformed lymphocyte cell lines and tend to involve the same subset of SS loci. **Summary** *These data are consistent with a common transcriptional gene control mechanism operating across a subset of SS risk loci in a shared intracellular environment that defines the site of differential allele action of causal variants.* The GWAS loci have small odds ratios (usually <1.2). If there are coordinated mechanisms across loci that alter disease risk with larger effect sizes (SS relative risks, 6.6 to 28.8), then shared gene regulatory mechanisms are relatively important in generating disease risk. In addition, the action of multiple risk variants may act across loci in the same cell type or types. That all 11 TF relationships are found in the transformed B cell nominates this cell type as a possible site for coordinated differential gene expression and gene action.

2002W

Identifying subject-specific regulatory networks in asthma. *D.C. Croteau-Chonka, K. Glass, B.A. Raby, Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE) Consortium.* Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

   RATIONALE: We have previously demonstrated that differential gene expression profiles of poor asthma symptom control reflect highly relevant biology, such as signaling pathways driving granulocytic inflammation. Regulatory networks are frameworks for integrating across multiple levels of biological information, including gene expression, which could provide further insights into disease mechanisms. **METHODS:** We measured whole blood gene expression using Illumina Human HT12 expression arrays (13,761 probes) in a subset of 245 asthmatic adults from Asthma BRIDGE. We used LIONESS, or Linear Interpolation to Obtain Network Estimates for Single Samples, an extension of the Passing Attributes between Networks for Data Assimilation (PANDA) method, to construct subject-specific regulatory networks. We then identified genes differentially targeted by transcription factors (TFs) between subjects with poor versus good asthma control, and determined via gene set enrichment analysis (GSEA) whether these genes clustered in specific biological processes. **RESULTS:** We constructed a collection of subject-specific regulatory networks and compared their properties among subjects with poor asthma control versus those with good control. Among the individual genes that showed the most significantly increased targeting by TFs during poor control were *TSLP* (rank #2) and *IL18R1* (rank #19), which are both well-known asthma and allergy susceptibility genes and pharmacologic candidates in asthma. Other notable genes included *TLR4* and *CASP4*, both critical components of lipopolysaccharide (LPS) signaling. Moreover, we found increased TF targeting of 140 genes (including *TSLP*) related to activation of TREM-1 signaling in monocytes – a key pathway promoting granulocytic inflammation, including in response to LPS exposure. Among TFs targeting these genes, the most influential were SP1 and EGR1, each implicated in airway remodeling and targeting ~99% of the TREM-1 genes (Fischer's Exact P ~ 1E-05). **CONCLUSION:** Comparing networks between disease states reveals important biological information about differences in gene regulation. Our identified regulatory networks helped refine previous differential expression profiles of asthma control and focus our attention on a specific pathways related to LPS exposure and TREM-1 signaling. **FUNDING:** This work was supported by grants from the National Institutes of Health (K01 HL127265, RC2 HL101543, P01 HL105339, R01 HL111759, R01 HL086601, and R01 HL118455).

2003T

Genome wide analysis of DNA methylation identifies a novel locus associated with asthma among African Americans. *Y. Li[1], M.P. Boorgula[2], R.A. Mathias[2], K.C. Barnes[3], Z.S. Qin[1], Y.J. Hu[1], Consortium on Asthma among African-ancestry Populations in the Americas(CAAPA).* 1) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 2) Department of Medicine, Johns Hopkins University, Baltimore, MD; 3) Department of Medicine, University of Colorado at Denver, Denver, CO.

   Background Asthma is a common chronic inflammatory disease, affecting more than 32 million people in the United States. Previous research suggested that asthmatics of African descent tend to have more severe clinical symptoms than non-African populations. In addition to genetic factors, asthma has also been linked to environmental exposures, which are likely mediated through epigenetic mechanisms.   **Objective** We conduct an epigenome-wide association study (EWAS) to identify differentially methylated CpG sites between asthmatics and non-asthmatics in the African American population. We also perform a methylation quantitative trait locus (meQTL) study to enhance our understanding of the interaction between DNA methylation and genetic factors. **Methods** The samples we used are part of NHLBI-supported Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). DNA Methylation data were collected from 79 asthmatics and 87 non-asthmatics using the Illumina Infinium Human Methylation 450K BeadChip, which covers more than 480,000 CpG sites genome-wide. We performed data filtering, quality control, and normalization. In EWAS, we used a linear regression model at each CpG site to examine the association between asthma status and methylation level after adjusting for age, gender, batch, cell type composition, and principal components that capture unmeasured sources of variation in the methylation data. In the meQTL study, we used a similar linear model examining the association between each single nucleotide polymorphism (SNP) and methylation level with the same adjustment as in EWAS.   **Results** Our analysis identified cg00326161, a CpG site on chromosome 6, to be significantly associated with the asthma trait (p value = 1.53×10^{-7}). Interestingly, we found that the neighborhood of this CpG site showed evidence of cell specific in vivo binding of a transcription factor *FOXA1* that is known to be involved in lung development, suggesting that methylation of the CpG site might be involved in the regulation process of *FOXA1*.

2004F

DNA methylation profiles of polycystic ovarian syndrome in Chinese women: A case-control study. *S. Li[1], H. Duan[2], D. Zhu[3,4], A. Ren[3,4], D. Glintborg[5], M. Andersen[5], V. Skov[6], M. Thomassen[1], T. Kruse[1], Q. Tan[1,7].* 1) Unit of Human Genetics, Department of Clinical Research, University of Southern Denmark; 2) Department of Medicine, Kolding Hospital, Kolding, Denmark; 3) Center of Reproductive Medicine, Linyi People's Hospital, Linyi, China; 4) Department of Obstetrics and Gynecology, Shandong Medical College, Linyi, China; 5) Department of Endocrinology, Odense University Hospital, Odense, Denmark; 6) Department of Hematology, Roskilde Hospital, Roskilde, Denmark; 7) Epidemiology, Biostatistics, and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark.

   As a universally common endocrinopathy in women of reproductive age, the polycystic ovarian syndrome is characterized by composite clinical phenotypes reflecting the contributions of reproductive impact of ovarian dysfunction and metabolic abnormalities with widely varying symptoms resulting from interference of the genome with the environment through integrative biological mechanisms including epigenetics. We have performed a genome-wide DNA methylation analysis on polycystic ovarian syndrome using Illumina's HumanMethylation450 BeadChip array. We identified a substantial number of genomic sites differentially methylated in the whole blood of 30 PCOS patients and 30 healthy controls (52 sites, false discovery rate<0.05 and corresponding p value<5.68e-06), highly consistently replicating biological pathways extensively implicated in immunity and immunity-related inflammatory disorders (false discovery rate<0.05) that were reportedly regulated in the DNA methylome from ovarian tissue under PCOS condition. Most importantly, our genome-wide profiling focusing on PCOS patients revealed a large number of DNA methylation sites and their enriched functional pathways significantly associated with diverse clinical features (levels of prolactin, estradiol, progesterone and menstrual cycle) that could serve as novel molecular basis of the clinical heterogeneity observed in PCOS women. .

2005W

Transcriptomics of liver tissue in alcoholic hepatitis. *T.M. Norden-Krichmar, Southern California Alcoholic Hepatitis Consortium (SCAHC).* Dept. of Epidemiology, University of California, Irvine, Irvine, CA.

Alcoholic Hepatitis (AH) is an alcohol-induced inflammatory liver disease with high mortality and morbidity rates. The factors that determine the progression to AH and the subsequent prognosis are not well understood. Gene expression in AH has previously been studied mainly using microarrays and qPCR. To our knowledge, this is one of the first studies using RNA sequencing for gene expression profiling of human liver tissue from patients with AH. The liver biopsies in this study were collected by the Southern California Alcoholic Hepatitis Consortium (SCAHC) from AH patients (n=9), and from patients diagnosed with decompensated liver (DL) disease (n=3) due to alcohol abuse with no episodes of AH. As normal liver tissue controls, RNA sequencing data of normal liver biopsy samples (n=3) from the Human Protein Atlas (www.proteinatlas.org) were incorporated into this analysis. RNA was extracted from the AH and DL liver biopsies and sequenced at 2x100 paired-end on an Illumina HiSeq2500. The RNAseq data was aligned to the human genome, followed by differential expression (DE) analysis with the Cufflinks software, using upper quartile normalization between the data files. Visualization and principal components analyses (PCA) were performed with cummeRbund using the entire gene expression dataset. PCA revealed that the global gene expression patterns of AH and DL clustered together, while the normal samples were clearly distinguishable. Normalized DE between the groups yielded 528 total genes with fold change >= 2, FPKM>=5, and were significant at FDR-adjusted p-value <= 0.05. Of these, there were 16 DE genes between AH and DL, 304 between AH and normal, and 471 between DL and normal. The gene expression profiles of AH and DL were similar, and distinct from the normal samples. Many of our results confirm those reported in previous microarray studies. That is, relative to normal samples, AH and DL show elevated expression of secreted phosphoprotein 1, selenoprotein, and chemokines. Additionally, genes related to drug metabolism and detoxification, such as an aldo-keto reductase and a glutathione s-transferase, were elevated in AH and DL compared to normal. The most significantly DE genes between AH and DL were related to immune function. These results suggest that there is a characteristic gene expression profile of alcoholic liver disease that may play a role and may point to treatment options for alcoholic hepatitis. (Supported by U01AA021838).

2006T

Long-range interactions organize the 11p13 cystic fibrosis modifier locus around *EHF.* *L.R. Stolzenburg[1,2], R. Yang[1,2], J. Kerschner[1,2], S. Fossum[1,2], M. Xu[1,2], S. Wachtel[1,2], S.H. Leir[1,2], A. Harris[1,2].* 1) Human Molecular Genetics, Lurie Children's Research Center, Chicago, IL; 2) Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL.

Although cystic fibrosis (CF) is caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, lung disease phenotypes cannot be predicted by *CFTR* mutation alone. Through genome wide association studies (GWAS), additional genetic loci associated with CF disease severity were identified by others. One such locus maps to chromosome 11p13 in an intergenic region between Ets homologous factor (*EHF,* an Ets family transcription factor) and Apaf-1 interacting protein (*APIP,* involved in apoptosis); however, the functional significance of this region is not understood. To identify *cis*-regulatory elements within the intergenic region, we used chromatin immunoprecipitation and DNase I digestion followed by high-throughput sequencing (ChIP-seq and DNase-seq). These assays identified peaks of H3K4me1/H3K27ac and DNase hypersensitive sites (DHS) across the 11p13 locus in primary lung epithelial cells and cell lines. Luciferase reporter assays in transformed bronchial epithelial cells confirmed that at least two of these DHS could function as strong enhancer elements for the *EHF* and nearby *ELF5* (another Ets family transcription factor expressed in glandular epithelium) promoters. None of the elements assayed showed significant activity with the *APIP* promoter. Next, circular chromosome conformation capture (4C-seq) was used to measure interactions between DHS and promoters across the region. The two enhancer elements showed reciprocal interactions with the *EHF* promoter, which also interacted with several other elements that appear to organize a topologically associated domain (TAD) across the region and are enriched for CTCF. No strong interactions were observed with the *APIP* promoter, which lies outside of the main TAD encompassing the GWAS signal. These results highlight the complex 3D chromatin architecture at 11p13 and suggest an important role for *EHF* in modifying CF lung disease severity.

2007F

Genome-wide regulatory roles of the C2H2-type zinc finger protein ZNF764 responsible for multiple steroid hormone resistance in a case with 16p11.2 microdeletion on the transcriptional activity of the glucocorticoid receptor. *T. Kino[1,2], A. Fadda[3], R. Mackeh[1], A. Marr[1].* 1) Experimental Genetics, Sidra Medical and Research Center, Doha, Qatar; 2) Program in Reproductive and Adult Endocrinology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 3) Biomedical Informatics, Sidra Medical and Research Center, Doha, Qatar.

The zinc finger protein (ZNF) 764 is a transcription factor harboring one KRAB domain and seven C2H2-type zinc finger motifs respectively in its N- and C-terminal half, and we previously reported that haploinsufficiency in its encoding *ZNF764* gene may be responsible for tissue resistance to multiple steroid hormones including glucocorticoids observed in a patient with 16p11.2 microdeletion (*J Clin Endocrinol Metab*, **98**: E1557, 2012). To evaluate the mechanisms underlying ZNF764-associated multiple steroid hormone resistance, we examined genome-wide ZNF764 actions on the glucocorticoid receptor (GR), a ligand-dependent transcription factor mediating diverse actions of glucocorticoid hormones, using HeLa cells as a model system. We found that ZNF764 enhances GR-induced transcriptional activity in a dose-dependent fashion in reporter assays employing the glucocorticoid-responsive luciferase reporter and in mRNA expression of the representative glucocorticoid-responsive genes. In chromatin immunoprecipitation assays followed by high through-put sequencing (ChIP-Seq), ZNF764- and GR-binding sites demonstrated similar distribution in various genomic features. They positioned predominantly around 50-500 kbs from the transcription start site of their targeting genes, closely locating with each other with overlapping in ~37% of cases. In ZNF764 knockdown experiments, ZNF764 demonstrated differential on/off effects on GR/DNA binding and subsequent induction of glucocorticoid-responsive genes: while some genes did not require ZNF764, others were absolutely dependent on the presence or absence of ZNF764. Pathway analysis revealed that these 3 gene groups in terms of ZNF764 requirement were enriched for distinct cellular actions. Finally, ZNF764 and GR physically interacted with each other *in vitro* and *in vivo*, and their mutual binding contributed to the regulation of representative ZNF764-sensitive glucocorticoid-responsive genes. These results suggest that ZNF764 functions as a cofactor that directs GR transcriptional activity toward specific biological pathways by changing GR binding and transcriptional activity on glucocorticoid-responsive genes. The results also provide important insight to the actions of ZNF764 on other steroid receptors and the mechanisms underlying multiple steroid hormone resistance observed in our case with 16p11.2 microdeletion.

2008W

The gene expression signature associated with rheumatoid arthritis is altered during pregnancy. *D. Jawaheer[1], E. Purdom[2], A. Mittal[1], L. Pachter[2], J.L. Nelson[3], J. Olsen[5], M. Smed[4], V. Zoffmann[4], M.L. Hetland[6], H. Kjaergaard[4].* 1) Children's Hospital Oakland Research Institute, Oakland, CA; 2) University of California, Berkeley, CA; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Juliane Marie Center, Copenhagen, Denmark; 5) Aarhus University, Aarhus, Denmark; 6) Glostrup Hospital, Copenhagen, Denmark.

Objective: To identify an RA-associated expression signature among women at the pre-pregnancy stage, and determine whether pregnancy-induced changes in gene expression alter the RA-associated expression signature by the third trimester. **Methods**: Blood was drawn from 9 women with RA and from 5 healthy women before pregnancy and at the third trimester. Total RNA was isolated and used to prepare cDNA libraries which were sequenced on an Illumina HiSeq2500 instrument at an average depth of 60 million reads (100bp). The pre-processed RNA sequencing (RNA-seq) data was pseudo-aligned to the reference Human transcriptome and gene expression levels were quantified using kallisto. Genes differentially expressed between groups (between RA and healthy women, or between time-points within each group of women) were identified with edgeR using a fold-change cutoff of 2 and a significance threshold of p<0.05 (FDR corrected). Functional category enrichment analysis was performed using Webgestalt. **Results**: Before pregnancy, 257 genes were differentially expressed between RA and healthy women (p<0.05), with 60 of these genes showing greater than 2-fold change in expression between the two groups. These genes were enriched in various immune-related pathways including those for defense response, innate immune response, inflammatory response and cytokine-mediated signaling. In contrast, in the third trimester of pregnancy, differential expression between the same set of RA and healthy women was limited to only 11 of the 60 baseline signature genes (>2-fold change in expression between the groups). Among the genes that lost their differential expression were OLFM4, CAMP, MMP8 and ORM1, which have been implicated in RA or in pregnancy. Interestingly, the dilution of the pre-pregnancy RA gene expression signature during pregnancy was the result of temporal changes in expression among RA and/or healthy pregnancies such that expression levels of these genes that showed RA-associated expression before pregnancy became comparable in the two groups of women by the third trimester. **Conclusions**: These findings have implications for both healthy and RA pregnancies. They suggest that, during healthy pregnancy, the immune system in the maternal periphery is active. In RA, temporal changes in expression that occur during pregnancy may be involved in the pregnancy-induced improvement of clinical symptoms of the disease.

2009T

Whole-genome methylation analysis for osteoporosis using reduced representation bisulfite sequencing (RRBS). *H. Shen, C. Qiu, H.W. Deng.* Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA.

Osteoporosis is a complex disease mainly characterized by a reduced bone mineral density (BMD) and increased risk of fractures. Peripheral blood monocytes (PBM) constitute 3-8% of human leukocytes in the blood and may act as precursors of osteoclasts, the bone resorption cells, and also produce cytokines important for osteoclast activity, and thus represent major systemic target cells for bone metabolism. Alterations in DNA methylation has been implicated as a key regulatory mechanism in the pathogenesis of various human complex diseases, such as cancers, cardiovascular disease and obesity *etc.* However, the DNA methylation profiles in humans underlying osteoporosis risk, especially at base-resolution level, are largely unknown. In this study, we performed a genome-wide DNA methylation analysis in primary human PBMs from 108 unrelated subjects, including 54 subjects with extremely high hip BMD and 54 with low BMD. By using reduced representation bisulfite sequencing (RRBS), we generated over 5.5 billion uniquely mapped reads for all the 108 samples with average uniquely mapped rate achieve to 69.7%. Genome-wide DNA methylation signals of each sample were normalized and quantified using methylKit analysis package. Differentially methylated regions (DMRs) were identified by comparing the genome-wide methylation profiles between the high and the low BMD subjects. The PBM methylome profile shows similar patterns as that in other somatic cells. For instance, most of CpG sites in the promoters are unmethylated or partially methylated, especially for promoters with high CpG density. Among the 2.04 million tested CpG sites with at least 10x coverage in promoter and CGI regions, a total of 327 CpG sites in 67 genes showed significant difference (q-value <0.01, %methylation difference >25%) in methylation levels between the high and the low BMD groups, including a number of known bone-related genes, such as the RUNX2 gene (p=1.4E-4, %methylation difference =0.47) and the BMP2 gene (p=4.3E-5, %methylation difference =0.52). We also identified non-CpG site methylation profiles and it appears to be spatially correlated with CpG methylation. For example, 34.7% of differential methylation in non-CpG sites overlapped with DMRs of CpG sites. In summary, we conducted the comparative whole genome DNA methylation analysis for PBMs and identified a number of novel epigenetic-regulated candidate genes for osteoporosis risk.

2010F

Roadmap to identifying craniofacial enhanceropathies. *J. Cotney, A. Wilderman.* Genetics and Genome Sciences, UConn Health, Farmington, CT.

Defects in embryonic patterning resulting in craniofacial abnormalities are common birth defects affecting more than 1 in 750 live births. The genetic causes of these defects have been difficult to determine, but all current evidence suggests defective gene regulation during embryonic development underlies these birth defects. Most of the individuals affected by congenital craniofacial abnormalities do not have defects in other tissues or organ systems, thus are relatively tissue specific and considered non-syndromic. The regulatory programs that build and shape the craniofacial complex require precisely controlled spatiotemporal gene expression. Much of the information that controls these gene expression programs is thought to be encoded in the large expanses of the genome between protein-coding genes and within intronic sequences. Efforts by large projects such as ENCODE and the Roadmap Epigenome Project have established associations of many biochemical features with functional portions of the genome and state of activity. These efforts have revealed that enhancers are typically tissue or timepoint specific and regulate one or a small number of genes over very large genomic distances. The tissue specific nature of enhancers coupled with the enrichment of a variety of disease phenotype associations from genome-wide association studies in these regulatory sequences suggest that defects in enhancers are at fault in many common disorders. The non-syndromic nature of most craniofacial defects indicate they are potentially enriched for defects in enhancer activity and likely to be "enhanceropathies". To date early stages of human craniofacial development have not been interrogated with modern functional genomics techniques preventing large scale integrative analysis of genetic associations and tissue-specific chromatin states. We have generated comprehensive chromatin state maps with chromatin immunoprecipitation sequencing (ChIP-Seq) for critical stages in human craniofacial development ranging from 4.5 post conception weekds (pcw) to 8 pcw. We have targeted selected tissue specific enhancer for unbiased long range interaction mapping using circularized chromosome conformation capture sequencing (4C-Seq). These chromatin profiles and enhancer-promoter contact maps have allowed identification of potential causative genetic changes in human patients with craniofacial abnormalities and develop mouse models through targeted genome editing.

2011W

An evolutionary understanding of DNA methylation patterns associated with osteoarthritis. *G. Housman[1,2], L.M. Havill[3], E. Quillen[3], A.C. Stone[1,2].* 1) School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA; 2) Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA.

Background: According to the WHO, the degenerative joint disease of osteoarthritis (OA) is present in 9.6% of men and 18.0% of women ages 60 or older world-wide. Of those affected, 80% have movement limitations and 25% are unable to perform major daily activities of life. The CDC further notes that OA of the knee joint is especially prevalent in the USA. Thus, it is important to understand what factors contribute to the development of this disease. Epigenetic factors, such as DNA methylation, play an influential role, and while these molecular mechanisms have been heavily studied in humans, few efforts have taken an evolutionary perspective. **Objective:** This project assesses the relationship between DNA methylation patterns and knee OA development in a baboon primate model (*Papio anubis*) and compares these findings to previously published data for humans in order to study the evolution of OA epigenetics. **Methods:** Skeletal tissue DNA methylation patterns were assessed in cartilage and trabecular bone of the right distal femora from pedigreed, age- and sex-matched adult baboons using the Illumina HumanMethylation 450K BeadChip. **Results:** Preliminary data, collected from 12 individuals (6 with and 6 without knee OA), reveal that several loci in each tissue were significantly differentially methylated between healthy and OA individuals. Specifically, out of over 450,000 positions, approximately 2.06% were differentially methylated between the OA and control groups, 1.94% between tissue types, and 1.32% among the four combinations of tissue type and OA status. Further, genes containing differentially methylated positions, such as *HOXA9*, *PLOD2*, *IRX2*, and *MSX2* overlap with those previously identified in human OA DNA methylation studies. Analysis of cartilage and trabecular bone from an additional 56 individuals (28 with and 28 without knee OA) is currently in progress. **Discussion:** From an evolutionary perspective, these findings give us an appreciation of DNA methylation variation in skeletal tissue from one primate species and two skeletal tissues. They also give us insight into the degree to which a common skeletal condition (OA) affects that variation. Analyses of an expanded sample set will further inform our understanding of epigenetic regulation and complex trait evolution in primates.

2012T

Identification of methylome alterations associated with disease progression in osteoarthritic subchondral bone. *Y. Zhang[1,2], N. Fukui[3,4], M. Yahata[2,5], Y. Katsuragawa[6], T. Tashiro[7], S. Ikegawa[8], M. Lee[1,2,9].* 1) Lab 119, Weis Center, Geisinger Genomic Medicine Institute, Danville, PA; 2) Laboratory for International Alliance on Genomic Research, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Clinical Research Center, National Hospital Organization Sagamihara Hospital, Kanagawa, Japan; 4) Department of Life Sciences, Graduate School of Arts and Sciences, the University of Tokyo, Tokyo, Japan; 5) Laboratory for Pharmacogenomics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 6) Department of Orthopedic Surgery, Center Hospital of the National Center for Global Health and Medicine Center Hospital, Tokyo, Japan; 7) Department of Orthopaedic Surgery, Tokyo Yamate Medical Center, Tokyo, Japan; 8) Laboratory for Bone and Joint Diseases, Center for Integrative Medical Sciences, RIKEN, Tokyo, Japan; 9) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Methylome alterations were rarely studied in subchondral bone of osteoarthritis (OA) due to the difficulties associated with accessing to the tissue and isolating DNA with adequate quantity and quality. In this study, we employed a unique system, which allows us to access three regions of the tibial plateau that could reflect the early, intermediate and late stages of OA. The methylome of subchondral bone from 3 regions in 12 primary OA knee joints were examined using HumanMethylation450 BeadChip. Of the significant differential methylated probes (DMPs), more than half were hypermethylatedin both intermediate (72 DMPs, 22 genes) and late (397 DMPs, 134 genes) stages of OA when compared to the early stage. These DMPs showed altered composition of genome features. The differential methylated genes (DMGs) were enriched in morphogenesis and development of skeletal system. DMGs that associated with OA and bone remodeling were identified (eg: *COL4A1*, *FAM5C*). A clustered of transcription factors were also identified, particularly the HOX transcription factors, of which many harbored multiple DMPs. For example, *TBX15* had 26 hypermethylated probes, and was significantly down-regulated in the intermediate and late stage. The expression was significantly negatively correlated to the methylation (spearman's r = -0.5228, p=0.0005). Different patterns were identified in subchondral bone when compared to the findings of the site-matched cartilage. However, shared DMP and DMGs with the same trend of methylation changes were also identified, which had even higher enrichement in the morphogenesis and development of skeletal system. Shared DMGs also included genes that participate in the crosstalk between cartilage and subchondral bone. Together, our data suggested DNA methylation contributed to the OA disease progression and highlighted the common pathways and crosstalk between the cartilage and subchondral bone.

2013F

Genome-wide analysis of differential DNA methylation and methylation quantitative trait loci (meQTLs) in systemic lupus erythematosus. *J. Imgenberg-Kreuz[1], J. Carlsson Almlöf[1], D. Leonard[1], G. Nordmark[2], M.-L. Eloranta[2], L. Padyukov[2], I. Gunnarsson[3], E. Svenungsson[3], C. Sjöwall[3], L. Rönnblom[4], J.K. Sandling[1,2], A.-C. Syvänen[1].* 1) Molecular Medicine and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 2) Rheumatology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Rheumatology Unit, Department of Medicine, Karolinska Institut/Karolinska University Hospital, Stockholm, Sweden; 4) Rheumatology/AIR, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden.

Systemic lupus erythematosus (SLE) is a rheumatologic autoimmune disease with complex etiology. More than 50 genetic loci that contribute risk for SLE have been identified, but a large proportion of genetic susceptibility still remains unexplained. Therefore we investigated the impact of DNA methylation on SLE susceptibility and on the phenotypic variability in SLE. In an initial discovery phase we determined genome-wide DNA methylation patterns in blood cells from 347 Swedish SLE patients and 400 healthy controls using the Illumina HumanMethylation 450k array. The Minfi R package was used for quality control, normalization of intensity data and calculation of methylation beta values at 385,851 CpG sites that passed quality control. Differential blood cell count estimations, age and sex were included as covariates in the association model and a p-value of <1.3x10^{-7} for the association was considered significant (Bonferroni correction α=0.05). For independent replication, 201 SLE patients and 188 healthy controls were included.   We found that DNA methylation is perturbed at thousands of CpG sites in blood cells from SLE patients compared to healthy controls. The largest differential methylation between patients and controls was observed at interferon (IFN) regulated genes, which are hypomethylated in SLE. Analysis of DNA methylation of the X-chromosome of female SLE patients identified differential methylation at genes involved in activation of immune cells, such as *TLR8* and *VSIG4*. In addition, we identified methylation patterns associated with disease duration and clinical manifestations, such as lupus nephritis. By mapping methylation quantitative trait loci (meQTLs) we observed genetic regulation of methylation levels by SNPs in regulatory regions identified by genome-wide association studies (GWAS) of SLE. For example at the *IRF5-TNPO3* locus the SLE-associated SNP rs4728142 was associated with methylation levels at the nearby CpG site cg04864179 (p=5x10^{-9}), both located in the *IRF5* promoter. Differential analysis of meQTLs in SLE patients and controls found additional disease specific meQTLs in the patients. Our results suggest that several risk variants for SLE exert their influence on the phenotype through alteration of DNA methylation levels at regulatory regions of target genes in blood cells. Analysis of meQTLs in SLE patients may be a powerful approach for identification of genetic regulation of methylation relevant for SLE and its phenotypes.

2014W

Cell-free transcriptomics reveal dysregulation of collagen and extracellular matrix organization in synovial fluid after ACL/meniscal injury. *A.J. Griswold[1], K. Nuytemans[1], L. Wang[1,2], T. Strong[1], D.D. Vance[3], M.K. Smith[4], J.M. Vance[1,2], M.A. Pericak-Vance[1,2], L.D. Kaplan[4].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School Medicine, Miami, FL; 3) Department of Orthopedic Surgery, New York Presbyterian Hospital, Columbia University Medical Center, New York, NY; 4) UHealth Sports Performance and Wellness Institute, University of Miami, Miller School of Medicine, Miami, FL.

Anterior cruciate ligament (ACL) and meniscal tears are two of the most common activity-induced orthopedic knee injuries and among the leading causes of subsequent osteoarthritis (OA) of the knee. While OA usually becomes symptomatic years after the injury, we hypothesize that molecular changes, initiated soon after the initial trauma, activate pathways eventually leading to increased susceptibility to OA. This study set out to identify gene expression biomarkers by performing RNAseq of cell-free RNA obtained from synovial fluid aspirated from knees at the time of repair surgery. Our initial study cohort consisted of 12 total samples from seven patients (ages 19-47) undergoing ACL and/or meniscal repair (ranging 6-24 weeks post injury). Five patients had samples from an injured knee plus the contralateral uninjured knee (control knee), and two patients had a sample from only an injured knee. Radiograph studies found no existing OA at the time of surgery. Total RNA was extracted from 0.5-1mL of synovial fluid, prepped with the NEBNext Small RNA Library Set, and sequenced 10M single end 50bp reads on an Illumina HiSeq2500. These reads were mapped to the genome and quantified using the STAR software. General linear models implemented in the EdgeR software were used in a paired analysis to identify genes differentially expressed between injured and uninjured knees while adjusting for differences between individuals. In total, 443 genes were nominally differentially expressed (p ≤ 0.05, fold change ≥ 1.5): 122 upregulated and 321 downregulated in injured compared to control knees. Ingenuity Pathway Analysis indicated that the top enriched molecular pathway of the differentially expressed genes is "cellular assembly and organization." These include filament and fibril formation genes, including collagen type I genes (*COL1A1* and *COL1A2*) and microfibrillar-associated proteins that were upregulated by knee injury. Conversely, cytoskeletal organization genes including cortactin (*CTTN*) and tissue inhibitors of metalloproteinases (*TIMP3*) were downregulated. Both collagen-related genes and MMPs have been previously implicated in the OA disease process, but this is the first evidence that the dysregulation of these molecular pathways can be identified as synovial extracellular RNA biomarkers early after injury. Understanding the role of these biomarkers, may ultimately allow early intervention and potential therapeutics to limit damage due to injured-induced OA.

2015T

Genome-wide DNA methylation analysis identifies novel hypermethylated genes associated with hidradenitis suppurativa (acne inversa). *N.K. Jain[1], D.D. Jhala[2], M. Raveendrababu[2], N.M. Saiyed[3], R.S. Rawal[4], B.C. Gorijala[5], F.M. Al-Ali[6], L.I. AlOtaibi[6], E.A. Mohamed[6], M. Naveed[6], U. Ratnamala[7], U. Radhakrishna[8].* 1) Life Sciences, School of Sciences, Gujarat University, Ahmedabad, India; 2) Department of Zoology, School of Sciences, Gujarat University Ahmedabad, India; 3) Biotechnology, Nirma Institute of Science, Nirma University, Ahmedabad, India; 4) Department of Dermatology, V.S. Hospital, Smt NHL Municipal Medical College, Ahmedabad, India; 5) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, India,; 6) Dermatology Centre, Rashid Hospital, Dubai Health Authority, Dubai, UAE; 7) Department of Pharmacology, Creighton University, Omaha, NE, USA; 8) Department of Obstetrics and Gynecology, Oakland University-William Beaumont School of Medicine, Royal Oak, MI, USA.

  Hidradenitis Suppurativa/Acne Inversa (HS/AI) is a chronic skin disease that usually presents after puberty with painful, deep-seated inflamed lesions in the apocrine gland-bearing areas of the body. In some families HS is hereditary and correlates with specific γ-secretase mutations. However, disease etiology is heterogeneous as these mutations are involved in the pathogenesis of only some familial and/or sporadic HS cases, while the majority of HS occupancies are sporadic. The diagnosis of HS is typically based on clinical characteristics, such as history and physical exam, rather than tissue biopsy. HS is frequently misdiagnosed or overlooked, resulting in a delayed diagnosis only when the disease is in an advanced stage. We hypothezed that epigenetic modification may play a crucial role in the development of HS; thus we conducted a genome-wide DNA methylation analysis using the Illumina Infinium HumanMethylation450 BeadChip array in a cohort of 24 HS individuals previously excluded from γ-secretase mutations and 24 controls matched for age, sex, and ethnic origin. Differential methylation analysis identified significantly altered CpG methylation at 304 sites (at least 2.0 fold hypermethylation) in 304 genes in HS subjects as compared to controls (False discovery rate (FDR) = ≤ 0.00001). All 304 identified CpG sites genes have ROC AUC ≥0.75 and had ≥5% methylation difference between HS and controls. Further gene ontology analysis identified biological processes and functions for these genes including Notch pathway, histone and DNA methylation, chromatin remodeling, immunological, and inflammation pathways. Many important significant novel CpG sites that might be related to HS development were detected including cg24332422 (*APC*), cg18584561 (*GREB1*), cg21308111 (*TNFRSF10A*), cg12506373 (*CAD*), cg02209504 (*EGR2*), reinforcing the potential importance of epigenetic modification in HS.

2016F

Early alterations in the NF-kB signalling pathway may underlie psoriasis pathogenesis. *B. Rhead[1,2], P. Ghai[3], L.F. Barcellos[2], A.M. Bowcock[3].* 1) Computational Biology Graduate Group, UC Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Laboratory, UC Berkeley, Berkeley, CA; 3) National Heart and Lung Institute, Imperial College, London, United Kingdom.

  Psoriasis (MIM 177900) is a chronic inflammatory disease of the skin affecting ~2% of the European population. It is accompanied by hyper-proliferation of the resident skin cell, the keratinocyte, infiltration of immune cells into the dermis and epidermis, and a disorganized skin barrier. The disease is also accompanied by increased vascularization due to proliferation and differentiation of resident endothelial cells. GWAS have revealed over 50 loci associated with disease. Many of these affect immune cell activation pathways and activation of the NF-kB pathway. We have also previously identified rare gain of function mutations in the gene encoding caspase recruitment domain protein 14 (*CARD14*) that lead to enhanced NF-kB signalling. *CARD14* is expressed in keratinocytes and endothelial cells, and we hypothesize that alterations in NF-kB signalling in both cell types, accompanied by genomic changes, are the early events in psoriasis pathogenesis. To investigate this hypothesis, we examined altered CpG methylation in both keratinocytes and endothelial cells following TNFα stimulation, which activates a number of signalling pathways including those of NF-kB. We measured transcriptional response to TNFα in 40 human umbilical vein endothelial cell (HUVEC) and 35 normal human epidermal keratinocyte (NHEK, 24 fetal and 11 adult) primary cell lines. Genome-wide genotyping and gene expression profiling via microarray, before and 24 hours after stimulation with TNFα, were performed on all 75 cell lines. Genes with differential expression in response to TNFα across all cell types were identified. Genome-wide methylation profiles were generated with Illumina HumanMethylation450 (450k) and EPIC BeadChips on 8 cell lines before and 24 hours after TNFα stimulation (16 profiles) and 21 HUVEC lines at 6 timepoints (126 profiles). Analysis of changes after 24 hours in 3 HUVEC, 3 fetal NHEK, and 2 adult NHEK lines shows significant increases in global mean methylation in all 3 cell types, with shifts in mean β values from 0.46 to 0.47, 0.43 to 0.44, and 0.45 to 0.46, respectively (p-values <0.001). Correlation between methylation response and transcriptional response to TNFα stimulation was observed, most notably: an increase in β value of 5-9% for 4 CpGs upstream and in the 5' UTR of *COL21A1* corresponded with decreased expression upon TNFα stimulation in HUVECs. Work to further elucidate early changes in the NF-kB signalling pathway in these cell types is underway.

2017W

Functional intronic *MAPT*-haplotype single nucleotide polymorphisms regulate *MAPT* exon 3 alternative splicing through differential binding of splice factors. *T.M. Caffrey[1,2], M.C. Lai[1,2], A.L. Bechy[1], F. Denk[1], M. Gavriliouk[1], B.J. Ryan[1,2], R. Wade-Martins[1,2].* 1) Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Parkinson's Disease Centre, University of Oxford, Oxford, United Kingdom.

Genome wide association studies have identified H1 microtubule-associated protein tau (*MAPT*) haplotype single nucleotide polymorphisms as leading common risk variants for Parkinson's disease, progressive supranuclear palsy and corticobasal degeneration. In the absence of protein coding changes, differences in *MAPT* expression and splicing of alternate exons are proposed to contribute risk or protection to these diseases. Expression studies performed in both cell lines and post-mortem brain tissue have demonstrated greater haplotype-specific expression of *MAPT* exon 3-containing transcripts from the protective H2 haplotype compared to the H1. Here, we investigate the regulation of *MAPT* exon 3 alternative splicing by common, non-coding, haplotype-specific single nucleotide polymorphisms (SNPs) through a novel use of whole-genomic locus expression vectors. Firstly, we generate *MAPT*-PAC expression vectors bearing 143 kb of either the H1 or H2 *MAPT* loci. These vectors contain 8 kb of upstream promoter, all exonic and intronic sequence, as well as 5 kb of 3' untranslated region. Together, the *MAPT*-PACs capture 86% of the documented variation within the gene. Secondly, we interrogate the polymorphic differences between these haplotype vectors, initially in silico, followed by generation and analysis of haplotype-hybrid vectors in a neuronal culture model. We identify rs17651213, located near the exon 3 5' splice site, as a polymorphism that regulates the haplotype-specific expression of exon 3. Thirdly, using RNA:electrophoretic mobility shift assays (RNA-EMSA), we show differential RNA:protein complex formation at the H1 and H2 sequence variants of SNP rs17651213 and identify candidate trans-acting splicing factors interacting with rs17651213 sequences by RNA:protein pull-down experiment and mass spectrometry. Finally, by paring knockdown of the mass-spectrometry identified candidate splice factors with a novel allele-specific expression assay, we demonstrate a role for hnRNP F and hnRNP Q in the haplotype-specific regulation of exon 3 inclusion. In this study, we dissect the *MAPT* locus to fine map sequences regulating alternative splicing of exon 3 and provide a starting point for mechanistic insights into how common non-coding H1/H2 *MAPT* haplotype-specific SNPs may contribute to the risk of neurodegeneration at a complex gene locus.

2018T

Investigating the role of *FAST-1* in Friedreich ataxia. *H. Mikaeili, M. Sandi, A. Bayot, S. Al-Mahdawi, M.A. Pook.* Life Sciences, Brunel University London, Uxbridge, Middlesex, United Kingdom.

Friedreich ataxia (FRDA) is an autosomal recessive neurodegenerative disorder caused by a GAA trinucleotide repeat (TNR) expansion within the first intron of the *FXN* gene, leading to severe deficiency of *FXN* transcript. FRDA patients have disease-related epigenetic changes, which may be the underlying cause of *FXN* gene silencing. Furthermore, it has previously been shown in other TNR diseases that increased antisense RNA expression can induce heterochromatin formation and epigenetic silencing of the corresponding sense gene. The frataxin antisense transcript, *FAST-1*, is overexpressed in FRDA patient-derived fibroblasts, associated with depletion of CTCF, a chromatin insulator protein, and heterochromatin formation. We have overexpressed *FAST-1* in both HEK293 and HeLa cell lines and we have identified a corresponding 30-70% decrease of *FXN* expression levels compared to control cells. Additionally, we identified a positive correlation between *FAST-1* copy number and *FAST-1* expression ($R^2 = 0.7$, P=0.04) and a negative correlation between *FAST-1* copy number and *FXN* expression level ($R^2 = -0.58$, P=0.04). Chromatin immunoprecipitation (ChIP) of *FAST-1* overexpressing HeLa cells showed reduced occupancy of CTCF at the 5'UTR of the *FXN* gene. It is plausible that increased *FAST-1* transcription is involved in the displacement of CTCF from the 5'UTR, resulting in heterochromatin formation and *FXN* gene silencing. Therefore, inhibition of *FAST-1* expression may be an effective approach to pursue for future FRDA therapy.

2019F

Identification of chemicals that mimic transcriptional changes associated with autism, brain aging and neurodegeneration. *J. Simon, B. Pearson, E. McCoy, G. Salazar, G. Fragola, M. Zylka.* University of North Carolina at Chapel Hill, Chapel Hill, NC.

Environmental factors, including pesticides, have been linked to autism and neurodegeneration risk using retrospective epidemiological studies. Here we sought to prospectively identify chemicals that share transcriptomic signatures with neurological disorders, by exposing mouse cortical neuron-enriched cultures to hundreds of chemicals commonly found in the environment and on food. We find that rotenone, a pesticide associated with Parkinson's disease risk, and certain fungicides, including pyraclostrobin, trifloxystrobin, famoxadone and fenamidone, produce transcriptional changes *in vitro* that are similar to those seen in brain samples from humans with autism, advanced age and neurodegeneration (Alzheimer's disease and Huntington's disease). These chemicals stimulate free radical production and disrupt microtubules in neurons, effects that can be reduced by pretreating with a microtubule stabilizer, an antioxidant, or with sulforaphane. Our study provides an approach to prospectively identify environmental chemicals that transcriptionally mimic autism and other brain disorders.

2020W

DNA methylome of human brain cortex and susceptibility of Parkinson's disease. *X. Chen[1], C. Kani[2], S. Tashiro[2], K. Dashtipour[2], C. Wang[1].* 1) Center for Genomics, Basic Science Department, Loma Linda University, Loma Linda, CA; 2) Division of Movement Disorders, Department of Neurology & Movement Disorders, Loma Linda University, Loma Linda, CA.

The current diagnosis of Parkinson's disease (PD) is primarily based on clinical signs and symptoms, and only an autopsy can confirm diagnosis. As such, identifying early pathological changes is crucial to enable therapeutic interventions before a major neuropathological damage occurs. As PD is a multifactorial disease where environmental and genetic factors are intricately associated, epigenetic modifications are the key interface between the environment and genome and may play a major role in PD susceptibility. These changes accumulate over time in response to lifestyle and environmental exposures, subsequently affecting gene expression. The association between epigenetic modifications and some diseases such as cancers has been recently established, but the link between aberrant DNA methylation and neurodegenerative disorders is only beginning to be explored. The objective of our current pilot study is to determine and compare DNA methylation patterns between patients with PD and the age- and sex-matched non-PD controls. We hypothesized that the methylation status of CpG islands are different in PD patients from non-PD individuals and a global differential methylation is related to the susceptibility of PD. DNA was extracted from frozen human brain cortex (superior frontal gyrus at the level of genu of the corpus callosum) from twelve PD patients and twelve controls. The human brain tissues were obtained from the Banner Sun Health Institute Brain and Body Donation Program. A genome-wide DNA methylation profiling was performed using the Illumina HumanMethylation 450k BeadChip array. We identified 2,795 differentially methylated CpG sites (DMSs) between PD cases and controls with a FDR ≤ 0.01 and 328 DMSs with a FDR ≤ 0.001. The majority of differentially methylated regions were hypomethylated suggesting overall hypomethylation in the brain of PD cases compared to controls. Unsupervised hierarchical clustering analysis based DMSs showed a distinct separation between PD and non-PD subjects, and there was a pattern of robust hypermethylation of synphilin-1, α-synuclein interacting protein (SNCAIP) gene in PD cases (P = 4.93×10^{-7} and $\Delta\beta = 0.60$). Our results suggest that a global differential DNA methylation is associated with the susceptibility of PD and there is a link between SNCAIP methylation and PD risk. The methylation level of SNCAIP may be a biomarker for PD diagnosis and treatment.

2021T

Allele-specific brain expression and association with CNS disease in 419 individuals. *Z. Zhang[1,2], S. Mostafavi[4], D.J. Philip[2,3], M. Kellis[1,2].* 1) EECS, Massachusetts Institute of Technology, Cambridge, MA; 2) The Broad Institute of Harvard and MIT; 3) Brigham & Women's Hospital and Harvard Medical School; 4) University of British Columbia.

Systematic differences in gene expression between the two alleles of the same individual can yield important insights in gene regulation. Here, we study allele-specific gene expression in human cortex across 419 individuals from two longitudinal aging cohorts (ROS and MAP). We find evidences of imprinting in 28 genes with mono-allelic expression across individuals, including 6 novel imprinted genes in the aging brain: PM20D1, AC091814, LTF, MIMT, COPG2, and NAA60. Some of these 28 genes also show evidence of regulatory region imprinting in their histone modification patterns, with consistent directionality. Interestingly, we show genomic imprinting could happen in a small regulatory region instead of the whole gene body using H3K9ac ChIP-seq and DNA methylation data in the same set of individuals. Also, to our best knowledge, our result is the first evidence to show that the imprinting regulatory region controlling allelic expression of their target genes. For one third of allelic genes, we predict 4701 driver cis-regulatory variants based on strict evidence of activity-correlated variants in separate haplotype blocks. Moreover, we develop a novel statistical method which utilizes that correlation between TF expression and target gene allelic expression on the heterozygous binding sites could be a good indicator for identifying driver regulators in the brain, including BHLHE40, ZBTB7A and BCL6. This result is supported by the published ChIP-seq data and the independent cohort GTEX data. Further, we note that many susceptibility variants for CNS diseases have potential regulatory effect, which can cause allelic expression in our data. Finally, we identified 18 genes whose allelic expression are affected by brain pathologies (i.e., Amyloid Beta, Neurofibrillary Tangle, Lewy Bodies and Cognitive Decline Rate) through cis-variants. Our approach links cis-regulatory variants, disease pathology and target gene and paves a new way for understanding disease mechanism through allelic activity. .

2022F

Dopaminergic neuron chromatin signatures refine a novel Parkinsonian-associated interval and establish a pipeline for informing future genetic studies of neurological disease. *S.A McClymont[1], P.W. Hook[1], M.A. Beer[1,2], O.A. Ross[3], A.S. McCallion[1].* 1) Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD; 2) Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD; 3) Mayo Clinic College of Medicine, Department of Neuroscience, Jacksonville, FL.

Non-coding sequence variation plays an important role in complex disease risk. However, the search for biologically appropriate regulatory sequences in which causative variation may act is not facile. It relies on the selection and isolation of appropriate cell types, and on the scrutiny of the chromatin landscape at appropriate developmental stages or following exposure to the correct stimuli/insults. In an effort to explore variation implicated in disorders of the basal ganglia, including Parkinson disease (PD), we have used ATAC-seq to define regions of open chromatin in subsets of lineage-identified dopamine (DA) neurons. We generated multiple enhancer catalogs, each from 50,000 FAC-sorted *ex vivo* DA neurons either from the forebrain (n=3) or midbrain (n=3) of E15.5 (Tg(Th-EGFP)DJ76Gsat) transgenic mice. Each ATAC-seq library yielded ≤28 million high quality mapped reads, producing catalogs with strong evidence of functional sequence constraint (PhastCons>0.3) and significant enrichment for processes and functions appropriate to neuronal function and dysfunction by GREAT/GO. These high quality catalogs facilitate the development of a regulatory vocabulary (computational classifier) of DA neurons using support vector machine learning. With preliminary filtering alone, our data yield a classifier with high predictive power (auROC=0.823) that is already providing insight into transcription factors active in DA neurons and the nature of the regulatory variation that might influence DA enhancer function. To improve the classifiers, we are now applying more stringent filters. Importantly, our catalogs can inform our understanding of the biology at GWAS-implicated intervals. Our data has already highlighted a putative enhancer in an intron of *SNCA*, overlapping a GWAS hit for Lewy body dementia. In preliminary zebrafish transgenic reporter assays of enhancer function, over 90% of mosaic embryos exhibit reporter activity in the CNS, driven by this putative enhancer sequence. Transgenic reporter mice assaying this enhancer are currently being evaluated. Beyond this locus, we are establishing massively parallel reporter assays to both functionally validate the enhancer catalogs and assess our ability to predict the function of regulatory sequence variation, particularly at GWAS-implicated loci. We will report progress in our efforts to assay the impact of PD and related movement disorders GWAS-implicated variation on DA neuron function and disease pathology.

2023W

Transcriptional programs defining functional domains in the human brain. *A.C. Watts, K.T. Siebenthall, J. Halow, R. Sandstrom, S. Ibarrientos, K. Lee, A. Castillo, A.K. Johnson, R. Kaul, J.A. Stamatoyannopoulos.* University of Washington, Seattle, WA.

Functional domains of the human brain have been mapped at a fine level, but the transcriptional programs by which these regions develop and are maintained are poorly understood. Knowledge of these transcriptional programs can provide insight into the pathology of neurological disorders and cancers. Human post-mortem brain tissue from 13 different brain regions was obtained from cognitively and pathologically normal individuals who participated in the Adult Changes in Thought Study at the University of Washington. Additionally, neuronal and glial cells were isolated from the same tissue to obtain paired cell-type specific data. We generated reference maps of accessible chromatin utilizing the DNase I assay, and transcription using RNA-sequencing. Thus far we have identified clusters of DNase I hypersensitive sites (DHSs) that are present in a specific sub-structure (e.g. occipital lobe) as well as in larger nuclei (e.g. basal ganglia). We are currently dissecting the clusters of DHSs to identify transcription factor motifs, as well as incorporating data from neurons and glial cells. We will further characterize the functionality of these DHSs utilizing genome editing technology to introduce mutations within them in human-specific model systems, including cerebral organoids, followed by comparative analysis of cell morphology, spatial organization, chromatin accessibility and gene expression. This research has the potential to elucidate the regulatory programs underlying region-specificity and functionality in the human brain, and provide insight into the molecular pathways involved in neurological disorders.

2024T

EWAS shows DNA methylation association with differences in sleepiness among African Americans. *R. Barfield[1], B. Cade[2,3], R. Saxena[2,3,4], D. Yan[1], H. Chen[1], C. Wang[1,5], S.R. Patel[7], S. Sunyaev[8], X. Zhu[9], S. Gharib[10], W.C. Johnson[11], J. Rotter[12], Y. Liu[13], X. Lin[1,14], S. Redline[2,6].* 1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 2) Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA; 3) Department of Medicine, Harvard Medical School, Boston, MA, USA; 4) Department of Anaesthesia, Massachusetts General Hospital, Boston, MA, USA; 5) Computational and Systems Biology, Genome Institute of Singapore, Singapore; 6) Division of Pulmonary, Critical Care & Sleep Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA; 7) Department of Medicine, University of Pittsburgh, Pittsburgh PA, USA; 8) Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA; 9) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 10) Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, WA, USA; 11) Department of Biostatistics, University of Washington, Seattle, WA, USA; 12) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 13) School of Medicine, Wake Forest, Winston-Salem, NC, USA; 14) Department of Statistics, Harvard University, Cambridge, MA, USA.

The epigenetics of normal and disordered human sleep are largely unknown. The Epworth Sleepiness Scale (ESS) is a widely-administered clinical questionnaire used to assess sleepiness. Excessive daytime sleepiness is a frequent symptom of Obstructive Sleep Apnea (OSA) and other sleep disorders, and increase risk for motor vehicle crashes and work related injuries and reduce quality of life. Daytime sleepiness is most prevalent in African Americans, and this association is not completely explained by measured sleep disorders or sleep duration. We performed the first epigenetic-wide association study of ESS scores and DNA methylation marks from purified monocytes in the Multi-Ethnic Study of Atherosclerosis (MESA). 619 individuals of African, European, and Hispanic ancestry, ages 54 to 91, with both ESS data and epigenomic results were analyzed using 64,785 high-varying probes on the M-value scale (inter decile range greater than 0.1). Models were adjusted for age, sex, cell-type enrichment, and site of enrollment. Technical covariates were adjusted prior to analyses via ComBat. An African-American analysis (n=132) found one Bonferroni-significant probe and an additional 12 probes with q-values less than 0.05. Mean ESS was 6.7 +/- 4.5. Two genes were associated with 6 of the 13 probes, *RAI1* and *KIFC3*. The Bonferroni-associated probe was found at *KIFC3* (p-value of 4.18E-07) and was associated with a -.6 decrease in the ESS score for every .1 unit increase in DNAm. cg14720773 and cg27208169 near *RAI1,* a gene associated with disruptions in sleep-wake cycles and recently found to be associated with OSA, were associated with a -.6 and -.7 decrease in ESS score for every .1 unit increase in the DNAm M-value (p-values of 1.41E-06 respectively). An increase in DNAm is thus associated with a decrease in sleepiness. These results were not impacted by the inclusion in the model of socio-economic or other sleep variables. This pattern was found for all significant or suggestive probes except for two: one near *COQ2* and another near *MEIS2*. There were no significant probes found in the other ethnic groups or in a multi-ethnic analysis. This study provides evidence that altered methylation patterns associate with daytime sleepiness in African Americans and suggest the potential impact of epigenetic factors in influencing this trait.

2025F

Genome-wide methylation patterns are associated with Parkinson's disease. *Y. Chuang[1], S. Horvath[2,3], K. Paul[1], J. Bronstein[4], Y. Bordelon[4], B. Ritz[1,4,5].* 1) Epidemiology, UCLA, LA, CA., USA; 2) Human Genetics, UCLA, LA, CA., USA; 3) Biostatistics, UCLA, LA, CA., USA; 4) Neurology, UCLA, LA, CA., USA; 5) Environmental Health, UCLA, LA, CA., USA.

   Background: There is growing interest in understanding the role of DNA methylation in Parkinson's disease (PD). Yet, genome-wide studies of methylation in PD are still rare. Weighted co-methylation network analysis (WGCNA) is a system biology method to build robust genetic network and to identify trait-associated co-methylation modules. Using genome-wide DNA methylation data from whole blood and saliva, we aim to identify gene-specific CpG sites and co-methylation modules that would implicate mechanisms or pathways in PD etiology or could possibly serve as blood based biomarkers. **Methods:** We used Illumina 450k microarrays to investigated whole blood methylation profiles in 335 PD and 237 controls and saliva methylation profiles in 128 PD and 131 controls in a population-based study in Central California that enrolled subjects of European and Hispanic ancestry. We used correlations analysis to identify individual CpGs associated with PD and conducted a meta-analysis across ethnic groups. Moreover, using WGCNA we assessed the relation of modules of co-methylated CpGs with PD. The gene set enrichment analysis tool -DAVID- provided us with gene functional annotations for PD-associated genes and modules. **Results:** In blood samples, we identified 82 CpGs significantly associated with PD at the genome-wide level. PD-associated CpG with the strongest association was on the mitochondrial gene. Employing WGCNA, three out of 6 PD-associated modules remained significant after Bonferroni correction and were significantly enriched for genes related to immune response. After adjusting for blood cell composition, PD-associated modules were enriched for genes related to neuron projection morphogenesis and the Wnt signaling pathway. We replicated our blood based findings - with and without adjustment for blood cell composition- in saliva samples from an independent case control sample. Moreover, in saliva samples PD-associated CpG was found on the ferritin gene. **Conclusions:** Our study provides the first evidence for epigenetic differences related to mitochondrial and immune-related genes in PD also identify underappreciated pathways of importance in PD like the cytoskeleton, the Wnt signaling pathway, and brain iron and ferritin. It is standard to adjust for blood cell composition in DNA methylation studies using blood samples. However, our results suggest that this approach may miss important signals when blood cell composition and immunologic factors play a role as in PD.

2026W

Variegated hypermethylation of the *FXN* CpG island shore in Friedreich ataxia. *Y.K. Chutake[1], C.C. Lam[1], G. Wiley[2], S.I. Bidichandani[1].* 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Oklahoma Medical Research Foundation, Oklahoma City, OK.

   Friedreich ataxia (FRDA) is caused by an expanded GAA triplet-repeat (GAA-TR) mutation in intron 1 of the *FXN* gene that results in epigenetic silencing of the *FXN* promoter. Bisulfite treatment and deep sequencing revealed that while the promoter is located within an unmethylated CpG island (CGI), the CpG island shore (CGI-shore), situated immediately downstream of the CGI in intron 1, is hypermethylated in FRDA but remains largely unmethylated in the non-disease state. The CGI-shore, which functions as a FRDA-specific differentially methylated region (FRDA-DMR), is hypermethylated due to enhanced spreading of DNA methylation from the expanded GAA-TR mutation in FRDA. The hypermethylated FRDA-DMR was detected in patient-derived lymphoblastoid cells, primary fibroblasts and peripheral blood samples. Analysis of individual DNA molecules revealed a variegated pattern of DNA methylation, with apparently stochastic methylation of individual CpG dinucleotides within the FRDA-DMR, the magnitude and extent of which was dependent on the length of the expanded GAA-TR mutation. Treatment with 5-aza-2'-deoxycytidine enhanced the ability of a class I histone deacetylase inhibitor that is known to ameliorate epigenetic promoter silencing in FRDA, to increase *FXN* transcript levels in patient-derived cells, indicating that DNA methylation contributes to *FXN* silencing in FRDA. Given that variable CGI shore methylation determines differential expression of genes in normal and disease states, the FRDA-DMR represents a novel and tractable model of variegated silencing in the context of a human genetic disease.

2027T

Abnormal methylome is present in an animal model of temporal lobe epilepsy. *D.B. Dogini[1,3], A.S. Vieira[1,3], A.H.B. Matos[1,3], W. Souza[1,3], R. Gilioli[2], I. Lopes-Cendes[1,3].* 1) Departments of Medical Genetics, School of Medical Sciences, UNICAMP, Campinas, Sao Paulo, Brazil; 2) Laboratory of Animal Quality Control (CEMIB), School of Medical Sciences, UNICAMP, Campinas, Sao Paulo, Brazil; 3) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN).

Epigenetics summarizes alterations to the chromatin template that collectively establish and propagate different patterns of gene expression without changes in DNA sequence. To better understanding the role of epigenetic changes in epilepsy, we determined the methylation profile of an animal model of temporal lobe epilepsy in comparison with control animals. We used laser capture microdissection to obtain tissue from hippocampus of rats treated with pilocarpine (*n*=2) as well as control animals. DNA was extracted using proteinase K protocol with modifications and it was converted by bisulfite with EZ DNA Methylation-Lightning™ Kit (Zymo Research). After conversion libraries were generated using TruSeq DNA Methylation and sequencing was performed in a Illumina HiSeq 2500. Bioinformatics analysis found 18 differently methylated regions along the 20 rat chromosomes. In these regions we identified several genes, including Mef2a (myocyte enhancer factor 2) – a transcription factor; Atp10d (ATPase, class V, type 10D) – involved with ATP binding and Puf60 (poly-U binding splicing factor 60) – involved with RNA/DNA binding. This is a pilot study and the candidate regions are still under analysis in order to characterize all genes within the abnormally methylated regions. However, it is already clear that the epileptogenic insult induced by pilocarpine injections in rats resulted in significant methylation changes, which may be implicated in the mechanisms underlying epilepsy.

2028F

Epigenetics of plantation work and exposure to organochlorines in the Japanese male Kuakini Honolulu Heart Program Cohort with Parkinson's Disease. *R.C. Go[1,2,7], A.K. Maunakea[4], M.J. Corley[4], G.W. Ross[2,5,6], H. Petrovich[2,5,6], K.H. Masaki[1,5], Q. He[1,5], M. Tirikainen[3].* 1) Kuakini Health Systems, Honolulu, HI; 2) Pacific Health Research and Education Institute, Honolulu, HI; 3) University of Hawaii Cancer Center, University of Hawaii Manoa, Honolulu, HI; 4) Department of Native Hawaiian Studies, John A. Burns School of Medicine, University of Hawaii Manoa, Honolulu, HI; 5) Department of Geriatric Medicine, John A. Burns School of Medicine, University of Hawaii Manoa, Honolulu, HI; 6) Neurology, Veterans Affairs Pacific Islands Health Care System, Honolulu, HI; 7) Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL.

Parkinson's Disease (PD) is a disease with multiple etiologies. Epidemiological studies have provided evidence that exposure to agricultural related occupations or agrichemicals elevate a person's risk for PD. However, the mechanism(s) underlying this association remains unknown. Here, we sought to test the possibility that exposure impacts epigenetic processes, in particular DNA methylation, that may consequently influence risk for development of PD. To this end, we determined whether working on a plantation on Oahu, HI and/or exposure to organochlorines are associated with differentially methylated loci (DML) between PD cases with high and low exposures, utilizing a prospective cohort assessed for years of work on a plantation and exposures to 21 OGC in occipital lobe tissues. Blood buffy coat and temporal lobe DNAs were extracted and the Illumina Infinium Human Methylation 450K Bead Chip platform was used to assess DNA methylation. Cases were tested for DML across 480K+ loci using ANCOVA. The comparison of 10+ to 0 years of plantation work exposure, detected 8 DML and 129 DML at a significance level of 0.0001 in brain and blood respectively. The comparison of cases with 4+ to 0-2 detectable levels of OGC, identified 11 and 18 DML at a p<0.0001, for brain and blood respectively. Hierarchical clustering and Principal Components Analyses of DML showed that the highest were separable from the lowest exposure groups. Glial origin of these signals and pathway analyses revealed links to key PD neuropathological pathways.

2029W

Variants in chromatin factors such as SMCHD1 and DNMT3B modify the epigenetic repression of the D4Z4 repeat and disease penetrance in facioscapulohumeral dystrophy. *R.J.L.F. Lemmers[1], M.L. van den Boogaard[1], J. Balog[1], P.J. van der Vliet[1], K.R. Straasheijm[1], R.F.P. van der Akker[1], M. Kriek[1,2], K.B.M. Hanson[2], M. Wohlgemuth[3], B.G.M. van Engelen[3], M. Auranen[4], S. Kiuru-Enari[4], S. Mitsuhashi[5], I. Nishino[5], R. Tawil[6], S.J. Tapscott[7], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 2) Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands;; 4) Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 5) Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan; 6) Department of Neurology, University of Rochester Medical Center, Rochester, NY; 7) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA.

Facioscapulohumeral dystrophy (FSHD) is mainly characterized by progressive weakness and atrophy of the facial, shoulder and upper arm muscles. FSHD is associated with partial loss of chromatin condensation, marked by CpG hypomethylation, at the D4Z4 repeat locus on chromosome 4 and transcriptional derepression of the D4Z4-encoded *DUX4* gene in skeletal muscle. In most cases FSHD is caused by a contraction of the D4Z4 repeat array to a size between 1-10 units (FSHD1), while the normal size ranges between 8-100 units. In rare cases (FSHD2), the disease can be caused by mutations in the chromatin repressor SMCHD1 that is required for the maintenance of CpG methylation at D4Z4. Variants in SMCHD1 can also modify the disease severity in FSHD1 families. Since not all FSHD2 cases can be explained by *SMCHD1* mutations, we performed whole exome sequencing in SMCHD1 mutation-negative FSHD families and identified heterozygous mutations in *DNMT3B* segregating with D4Z4 hypomethylation and increased penetrance of FSHD. DNMT3B is one of the de novo DNA methyltransferases that is necessary for the establishment of CpG methylation at specific loci during development. Recessive *DNMT3B* mutations were previously shown to cause Immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome. *DNMT3B*-associated FSHD patients have no immunodeficiency, but ICF patients and their parents are at risk of expressing DUX4 and developing FSHD. Similar to what is observed for *SMCHD1* mutations, the effect of *DNMT3B* mutations on DUX4 expression and disease severity depends on the residual activity of the D4Z4 chromatin repressors and the size of the D4Z4 repeat. While overexpression of SMCHD1 can reverse D4Z4 chromatin decondensation and silence DUX4 expression in muscle cells, overexpression of DNMT3B fails to do so, perhaps because of a mismatch with the developmental window in which DNMT3B is normally functional. These results suggest that a combination of genetic and epigenetic factors that act on the D4Z4 repeat array determine the probability of DUX4 expression in skeletal muscle and disease penetrance and progression. FSHD thereby facilitates the study of chromatin factors that control accurate establishment and maintenance of chromatin structure and transcriptional activity from repeat DNA.

2030T

Genome-wide meta-analysis identifies SNPs associated with epigenetic age acceleration across multiple brain regions. *A. Lu[1], E. Hannon[2], M. Levine[1], E. Crimmins[3], K. Lunnon[2], J. Miller[2,4], D. Geschwind[5], S. Horvath[1,6].* 1) Human Genetics, UCLA, Los Angeles, CA; 2) University of Exeter Medical School,Exeter, UK; 3) Davis School of Gerontology, University of Southern California, Los Angeles, CA; 4) Institute of Psychiatry, King's College London, London,UK; 5) Department of Neurology and Center for Autism Research and Treatment, Semel Institute, UCLA, Los Angeles, CA; 6) Biostatistics, School of Public Health, UCLA, Los Angeles, CA.

We recently developed a biomarker of aging based on DNA methylation (DNAm) levels that is referred to as "epigenetic clock". This biomarker of aging applies to all human tissues, organs and cell types and results in a highly robust estimate of age. By contrasting the resulting age estimate (DNAm age) with chronological age, one can define a measure of epigenetic age acceleration. We have shown that epigenetic age acceleration in postmortem prefrontal cortex samples relates to pre-mortem estimates of cognitive decline and neuropathological measures even after adjusting for chronological age and other confounders. We propose that epigenetic age acceleration can also be used in a QTL analysis to identify SNPs that affect the epigenetic aging rate of the brain. Using DNA methylation data from cerebellar samples, we identified a mTOR associated gene *MLST8* and the RNA helicase gene *DHX57* in a genome-wide meta analysis. Here we extend these studies to non-cerebellar brain regions as well. We analyzed 1,796 brain tissue samples from 1,163 individuals across 7 studies. GWAS was performed on ~ 7M genotyped and imputed markers for epigenetic age acceleration in cerebellum, (pre)frontal cortex, pons and temporal cortex. The association results were combined across brain regions and studies using fixed effects meta analysis weighted by inverse variance. We identified a total of 7 SNPs associated at genomewide significance level of 5.0×10^{-8}. These GWAS SNPs are close to a calcium signaling associated gene. The risk allele of the leading SNP increases the epigenetic age estimate by 1.4 years in the frontal cortex, by 1 year in prefrontal cortex, temporal cortex and pons, and by 0.8 years in the cerebellum. These association signals are corroborated by 210 surrounding SNPs that are significant at a suggestive significance threshold of 5.0×10^{-7}. A large-scale *cis*-eQTL analysis (n=3,943) implicates the respective genes ($P=1.5 \times 10^{-19}$) for the underlying biological process. A MAGENTA enrichment analysis of the GWAS results implicated GTPase inhibitor activity ($P_{MAGENTA} = 2.0 \times 10^{-4}$). Furthermore, the genes implicated in brain age acceleration exhibit suggestive overlap with those found by GWAS of dementia ($P=5.8 \times 10^{-3}$) and Alzheimer's disease ($P=4.9 \times 10^{-3}$). Overall, our study suggests that it will be useful to use epigenetic age acceleration as quantitative phenotype for studying the genetic determinants of brain aging. .

2031F

Molecular mechanisms of pathogenicity: Why truncating variants in *KIAA2022* cause epileptic encephalopathy in some (but not all) females. *C. Saunders[1,2,3], J.B. LePichon[4], E. Farrow[1,3], S. Soden[1,3,4], S. Parikh[5], I. Thiffault[1,2,3].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospitals, Kansas City, MO, USA; 3) University of Missouri–Kansas City School of Medicine, Kansas City, MO, USA; 4) Department of Pediatrics, Children's Mercy Hospitals, Kansas City, MO, USA; 5) Center for Pediatric Neurology, Cleveland Clinic, Cleveland, OH, USA.

   X chromosome inactivation is a mechanism that modulates the expression of X-linked genes in females to achieve a proper balance between X-linked and autosomal genes. In humans 15-25% of X-linked genes have been shown to 'escape' from X-chromosome inactivation (XCI), being expressed to some extent from the inactive X chromosome. A significant subset of such escapees are genes associated with intellectual disability. The expression of X-linked genes in females lies on a continuum between those subjected to inactivation, and those that escape. Recent genomic studies have yielded contradictory results and fueled debate over the possibility that genomic imprinting is not limited to autosomes and the X-inactivation center. We report two females with intractable epilepsy caused by *de novo* loss of function variants in *KIAA2022*, an X-linked gene associated with severe intellectual disability in males. This is perplexing since the carrier mothers of affected males are unaffected. Random patterns of XCI have been observed in both affected females and unaffected carrier mothers so is not sufficient to explain these differences. We offer indirect evidence that *KIAA2022* escapes XCI, review what is known about *KIAA2022* and argue parent-of origin mono-allelic expression as a possible mechanism for the apparent dichotomous phenotypic expression of *KIAA2022*.

2032W

Multiple system atrophy: Are epigenetics to blame? *E.N. Scott[1], I. Guella[1], A. Rajput[2], A.H. Rajput[2], L. Parkkinen[3], M.S. Kobor[4], M.J. Farrer[1].* 1) Centre for Applied Neurogenetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Royal University Hospital, Saskatoon, Saskatchewan, Canada; 3) Oxford Parkinson's Disease Centre, University of Oxford, Oxford, United Kingdom; 4) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada.

   Multiple system atrophy (MSA) is an adult-onset neurodegenerative disorder with prominent autonomic dysfunction, poorly levodopa-responsive parkinsonism, and alpha-synuclein (*SNCA*)-immunopositive oligodendroglial pathology. The two predominant subtypes are MSA-P with clinical parkinsonism and MSA-C with cerebellar ataxia. Clinical heterogeneity in MSA has been a major stumbling block in the identification of pathogenic variants. Certain clinical features overlap with similar disorders, and, due to the lack of a 'hallmark' clinical presentation, the rate of misdiagnosis is high. Given the fact that it is a rare, sporadic disease, large sample sets are needed to have enough power to detect a true disease-causing variant. Whole exome sequencing has become a fundamental tactic for the identification of pathogenic variability in many rare diseases. However, if using only sporadic MSA cases, hundreds of shortlisted variants may be identified. Prioritization of variability in genes previously implicated in MSA, involved in neurodegeneration, and found within relevant pathways are only a few of the approaches taken to narrow down the list of candidates. Genome-wide genotyping, investigating copy number variability and regions of homozygosity, may implicate loci but even the largest meta-analyses will be underpowered. We and others have employed many of these methods, but have failed to identify any disease-causing variability in MSA. We hypothesize genetic, epigenetic, and expression changes in *SNCA* influence the distribution and burden of alpha-synuclein pathology in the brains of patients with MSA. Using striatal tissue from brains with pathologically-confirmed MSA, Parkinson's disease, dementia with Lewy bodies, and unaffected control subjects (n=32, 8/group), we pilot genome-wide SNP, methylation, and gene expression analyses. We will report on differential methylation and expression between tissues, patients, and control subjects, adjusting for SNP variability and considering the burden of inclusion pathology. Understanding the pathogenesis of MSA may provide insight into the propagation of alpha-synucleinopathies and inform the development of novel therapeutics.

2033T

Aberrant DNA methylation of the non-coding *VTRNA2-1* (*MIR886*) gene in Parkinson disease. *S.K. Sivasankaran[1,2], A. Ali[1,2], A. Mehta[1,2], D.C. Mash[1,2], G. Beecham[1,2], E.R. Martin[1,2], J. Vance[1,2], L. Wang[3], J.I. Young[1,2].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 2) John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL; 3) Department of Public Health Sciences, Division of Biostatistics, Miller School of Medicine, University of Miami, Miami, FL.

Parkinson disease (PD) is a common neurodegenerative disorder pathologically defined by the presence of Lewy bodies and extensive loss of dopaminergic neurons in the substantia nigra. Epidemiological studies have shown that there are many environmental exposures that modify the risk of PD, but the mechanism of these interactions is not known. As DNA methylation is one the known mechanisms through which environment can influence biological process, we investigated whether methylation differences exist between PD and controls. However, since it is well established that methylation profiles in the brain are region specific, we utilized brain regions known to be involved at different stages of PD. Previously, the majority of PD cases have been shown to have the earliest signs of Lewy body formation in the olfactory bulb and dorsal motor nucleus of the vagus (DMV), which have direct connections to the environment, through the olfactory and vagus nerves. In this study, we investigated genome-wide DNA-methylation profiles in the DMV of autopsy-confirmed PD cases and controls using Illumina 450K beadchip. A total of 34,942 CpGs were differentially methylated (FDR<0.05) in our analysis. Methylation patterns have been shown to be highly correlated, leading to the identification of differentially methylated regions (DMR), highly correlated CpG clusters that are less subject to identification artifacts than single site methylation. We identified 1,775 DMR in the DMV (adjusted P<0.05), including some in previously known PD associated genes like *MAPT* and *EIF4G1*. Of particular interest is a DMR associated with *VTRNA2-1*, which contains 20 differentially methylated CpGs including 9 CpGs in the promoter region. *VTRNA2-1* was previously reported found to lie within the 1-LOD score peak of a PD linkage region and is reported to be upregulated in "early" PD, suggesting it may contribute to the disease mechanism. Further, multiple studies have reported it is both imprinted and environmentally susceptible to methylation. This finding supports the hypothesis that environmental risks factors for PD may act early through the DMV and that the *VTRNA2-1* gene is a candidate gene for involvement in early PD.

2034F

iPSC-derived neurons as a model system for studying cis and trans factors regulating *SNCA* gene expression and their role in synucleinopathies. *O. Chiba-Falek[1,2], L. Tagliafierro[1,2].* 1) Neurology, Duke Univ, Durham, NC; 2) Center for Genomic and Computational Biology, Duke Univ, Durham, NC.

Accumulating evidence has suggested that *SNCA* expression levels are critical for the development of Parkinson's Diseases (PD), Dementia with Lewy Bodies (DLB) and other synucleinopathies. Common and distinct regulatory mechanisms of *SNCA* gene expression might be involved in the etiology of the different synucleinopathies. To study the *cis* and *trans* genetic factors that regulate *SNCA* expression in PD compared to DLB we established a model system of induced Pluripotent Stem Cells (iPSCs)-derived neurons from a normal subject and a patient with *SNCA*-triplication. These iPSC lines were differentiated into dopaminergic and cholinergic neurons to model PD and DLB, respectively. Each neuronal type was characterized throughout the stages of differentiation: iPSCs, Neural Precursor Cells (NPCs), final neurons, and aged neurons. *SNCA*-mRNA and protein expression in the *SNCA*-triplication cells exhibited two-fold increase compared to the cells from normal control for each neuronal type across differentiation, recapitulating the observations in human tissues and demonstrating the suitability of our model system for studying up-regulation of *SNCA* expression. TargetScan analysis identified five conserved miRNA binding sites in *SNCA*-3'UTR. Noteworthy, previously three of these miRNAs showed direct interactions with *SNCA*-3'UTR and down-regulated *SNCA*-mRNA and protein levels. The miRNAs demonstrated distinct expression patterns in the dopaminergic compared to the cholinergic neurons and across the differentiation stages of each neuronal type. Remarkably, miR-7 was expressed exclusively in mature dopaminergic neurons, while miR-223 was expressed predominantly in cholinergic neurons. Also, the overall lower levels of the miRNAs in final neurons compared to NPCs corresponded to higher *SNCA*-mRNA levels. Transcription factors previously reported to regulate *SNCA* transcription were also differentially expressed during differentiation and between the dopaminergic and the cholinergic neurons. For example, GATA2 levels were elevated in mature neurons and were 2-fold increased in cholinergic compared to dopaminergic neurons. Genome editing experiments (CRISPR/Cas9) are underway to examine how *cis*-variants modulate *SNCA* expression via their interactions with the *trans*-acting factors. In conclusion, we developed and are using an isogenic iPSC-based model system to discover the *cis* and *trans* acting factors that contribute to the regulation of *SNCA* expression in the context of PD and DLB.

2035W

Exploring the human genome for functional, non-coding variation reveals a far-upstream regulatory SNP at *TUBB2B***.** *W. Law[1,2], A. Vester[1], A. Antonellis[1].* 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD.

   Sequence variations within transcription factor binding sites (TFBS) can result in dysregulation of gene expression and directly cause or alter the severity of human disease. The implicated variants are often single-nucleotide polymorphisms (SNPs) within TFBS that alter DNA binding affinity. We hypothesize that such regulatory SNPs (rSNPs) impact the function of genes important for peripheral nerve development and function, and represent excellent candidate modifier loci for peripheral nerve disease. To address this hypothesis, we developed a computational pipeline to identify and functionally characterize novel rSNPs impacting the peripheral nervous system. Specifically, we computationally identified candidate TFBS across the human genome based on: (1) conservation of 5 base pair (or longer) genomic segments between human, mouse, and chicken; (2) overlap with a known, validated SNP; and (3) localization to non-protein-coding sequences. These efforts revealed 6,164 conserved SNPs in the human genome. We further refined our computational pipeline to identify rSNPs in a transcription-factor centric manner focusing on SOX10, which is essential for peripheral nerve function. We filtered the above 6,164 SNPs to find those that reside within a SOX10 consensus binding site or a SOX10 ChIP-Seq binding peak. These efforts revealed 22 putative SOX10-relevant rSNPs. To evaluate the regulatory potential of our predicted rSNPs, a genomic segment surrounding each rSNP was cloned upstream of a luciferase reporter gene and transfected into immortalized Schwann cells. Luciferase assays were performed to identify genomic segments with at least a 5-fold increase in activity over a control vector. This revealed four novel genomic segments with enhancer activity, of which two showed a significant difference in luciferase activity between the major and minor alleles. We deeply characterized one region using transgenic LacZ reporter mice, which revealed tissue-specific enhancer activity in SOX10 relevant tissues. We further assessed this element by deleting the region in cultured Schwann cells using CRISPR/Cas9 followed by RNA sequencing to identify a novel candidate target gene: *Tubb2b*. Both the identified rSNP and *Tubb2b* represent excellent candidate modifiers of peripheral nerve-related diseases. Additionally, the computational and functional pipelines developed here will be applicable to additional transcription factors.

2036T

APOE **and** *TOMM40* **gene regulation.** *L. Bekris[1,3], Y. Shao[1], J. Leverenz[2].* 1) Genomics Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, OH; 2) Cleveland Clinic, Neurological Institute, Cleveland Clinic Lou Ruvo Center for Brain Health, Cleveland, OH; 3) Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH.

  Background: The *APOE* gene is the strongest genetic risk factor for late-onset Alzheimer's disease (AD). Multiple regulatory elements have been described within and surrounding the *APOE* gene. Our group, and others have described a statistical, as well as mechanistic, relationship between *TOMM40* regulatory region genetic variants and *APOE* expression. *TOMM40* is located directly adjacent to and upstream from *APOE*. Interestingly, a potential regulatory element has been described within the distal region of *APOE* suggesting 3'UTR regulation. However, little is known about the influence of miRNA post-transcriptional regulation at this locus. MicroRNA (miRNA) play an essential role in post-transcriptional gene regulation in the brain. Genetic variants located across the *APOE* locus are the strongest risk factors for late-onset AD described to date. **Objective:** The aim of this investigation was to determine if the *APOE* 3'UTR and miRNA play a role in regulation of *APOE* or *TOMM40*. **Method:** *APOE* exon 4 and the 3'UTR were inserted into a luciferase reporter vector upstream from the poly-A. A library of reporter constructs containing either the *TOMM40* promoter or the *APOE* promoter and in combination with *APOE* exon 4 - 3'UTR haplotypes were produced. These reporter haplotypes were transiently transfected into several different human cell lines. MiRNA were tested for their influence on regional gene regulation, including regulation of the *TOMM40* promoter. **Results:** *APOE* exon 4 - 3'UTR ε2 and ε3 haplotypes, but not the ε4 haplotype, inhibit both the *APOE* and *TOMM40* promoters in the human neuronal SHSY5Y cell line. A miR-744 transiently transfected mimic further inhibits the *APOE* exon 4 - 3'UTR ε3 haplotype influence on the *TOMM40* promoter, but not the *APOE* promoter, suggesting that an interaction between a 3' regulatory element and the *APOE* promoter does not involve miR-744 in SHSY5Y cells. **Conclusion:** These results suggest that a complex regulatory structure within and surrounding *APOE* may be fine-tuned by miRNA post-transcriptional modulation.

2037F

A novel method for achieving single cell resolution of epigenomic status.
L. Kurihara[1], C. Luo[2,3], E.A. Mukamel[4], R. Castanon[3], J. Lucero[5], J.R. Nery[3], C.L. Keown[4], C. Schumacher[1], T. Harkins[1], M.M. Behrens[5], J.R. Ecker[2,3]. 1) Swift Biosciences, Inc., Ann Arbor, MI; 2) Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA; 3) Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA; 4) Department of Cognitive Science, University of California, San Diego. La Jolla, CA; 5) Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA.

When performing whole genome bisulfite sequencing (WGBS), highly efficient conversion of DNA fragments into library molecules is necessary when input quantity is limited. To meet this need, we developed an efficient library preparation using Adaptase for NGS adapter ligation to single-stranded, bisulfite-converted DNA fragments. This method significantly improves library complexity compared to existing commercially available methods. Since comprehensive methylome coverage was achievable from low DNA inputs, this method was modified and applied to single cells for classification of mammalian brain cell types based on methylation pattern. Approximately 20% of the mouse genome contains differential methylation that allows neuronal cell types to be distinguished by low pass WGBS. For example, >200,000 differentially methylated regions were identified among three cortical excitatory and inhibitory neuron types. Starting from single neuronal nuclei isolated from mouse frontal cortex that had undergone lysis and bisulfite conversion, NGS adapters were incorporated onto DNA fragments using a single cycle of random priming followed by Adaptase. The resulting low pass WGBS sequence demonstrated robust neuron type classification, readily separating excitatory and inhibitory cells, and further identified distinct populations of inhibitory cells. This method enabled unbiased characterization of brain epigenomic diversities without the need for the isolation of specific cell populations. Compared to other single cell bisulfite sequencing methods, the Adaptase workflow was faster and produced a higher percentage of aligned reads that increased the per cell data output. This method will meet the need of large-scale single cell methylome profiling of thousands of cells to enable discovery of single cell epigenomic variation.

2038W

Integrated analysis of miRNA and mRNA transcript profiles reveals different molecular mechanisms underlying seizure and *status epilepticus*-like in the zebrafish brain. *C. Maurer-Morelli, R. Oliveira, A. Vieira, A. Matos, W. Souza, C. Rocha, B. Carvalho, I. Lopes-Cendes, V. Zago.* Genetica Medica, UNICAMP, Campinas, Sao Paulo, Brazil.

Purpose: MicroRNAs (miRNAs) are key molecules associated with seizures. This study aimed to integrated miRNA/mRNA transcript profile data of two protocols of seizure-induced in the zebrafish model in order to identify molecular mechanisms underlying seizures. **Methods:** Seven days post fertilization larvae were separated into (i) control (CTL, n=3), (ii) acute seizure (AS, n=2) and (iii) *status epilepticus*-like (SE, n=3) groups. Larvae from AS and SE groups were exposed to pentylenetetrazole 15mM for 20 minutes and 3h, respectively. After seizure-induced, animals were anesthetized and their heads isolated, quickly frozen and stored at -80°C. Each sample was composed by pooling 20 heads to obtain material for RNA extraction. Illumina's Sample Prep Kit was used according to instructions to get miRNA and mRNA libraries. Validated and pooled libraries were sequenced in the Illumina HiSeq 2500 System. We used FastQC, DESeq2, and TopHat tools for bioinformatic analyses. MiRNA and mRNAs differentially expressed were determined (adj p<0.01) and the target genes investigated by the TargetScanFish 6.2. Animal Ethical Committee/UNICAMP (# 3045-1). **Results**: We found (i) two miRNAs (dre-miR-460-3p and dre-miR-725-3p) and 2897 mRNAs in the AS *vs* CTL groups, (ii) three miRNAs (dre-miR-31, dre-miR-132-5p and dre-miR-459-3p) and 4622 mRNAs differentially expressed in the SE *vs* CTL groups, and (iii) four miRNAs (dre-miR-725-3p, dre-miR-221-5p, dre-miR-192 and dre-miR-460-3p) and 2841genes between the AS *vs* SE groups. **Discussion:** After miRNA-mRNA integration data, we highlight the dre-miR-460-3p (down-regulated) and its up-regulated target, *inhbaa* gene in the AS group. The *inhbaa* gene is related to cellular growth, survival, and differentiation, and this up-regulation could represent a response toward reduce the cell damage. From the SE-like group, we highlight the association between the up-regulated dre-miR-459-3p and the repression of *adar, grm2a* and *pcdh19* genes. The *pcdh19* gene is a protocadherin that plays an important role in brain development, and mutations in this gene are associated with severe seizures in children. To our knowledge, this is the first study showing the miRNA profile after seizures or SE-*like* states using the zebrafish. By integrating global miRNA-mRNA transcript profile, we hope to shed some light on the molecular pathways underlying seizures at different conditions and their impact on the brain. **Supported by** FAPESP 14/15640-8 and CEPID-BRAINN.

2039T

CpG methylation at the *DMPK* locus as a prenatal indicator of congenital myotonic dystrophy. *S. Lanni[1], L. Barbe[2], A. Lopez-Castel[1], S. Tome[1], C. Spits[2], K. Keymolen[3], S. Seneca[3], Z. Musova[4], D. Chitayat[5], C. Gagnon[6], J. Matthieu[6], J. Letourneau[6], K. Sermon[2], C.E. Pearson[1].* 1) Genetics & Genome Biology, The Hospital for Sick Children, toronto, Ontario, Canada; 2) Department for Reproduction and Genetics, Vrije Universiteit Brussel, Belgium; 3) Centre for Medical Genetics, Universitair Ziekenhuis Brussel, Belgium; 4) Department of Biology and Medical Genetics, Charles University 2nd Medical School and University Hospital Motol, Prague, Czech Republic; 5) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Groupe de recherche interdisciplinaire sur les maladies neuromusculaires, Saguenay, Quebec, Canada.

Myotonic dystrophy type 1 (DM1) is caused by the expansion of CTG repeats in the 3' UTR of the *DMPK* gene. The repeat tract tends to expand through transmission, causing more severe phenotypes in the progeny, a phenomenon called genetic anticipation. A parent-of-origin effect is evident in DM1, where larger expansions and more severe pathogenesis are associated with inheritance of the mutant allele from mothers. Congenital myotonic dystrophy (CDM), the most severe form of DM1, represents the most extreme example of parent-of-origin effect with a bias for maternal transmission, where offspring often have exceedingly large repeat expansions (>1000 repeats). However, not all CDM individuals have massive expansions, and not all individuals with massive expansions have the congenital form, suggesting that some other mark is responsible for the maternal transmission and presentation of CDM. The *DMPK* gene is located in a CpG island flanked by two CCCTC-binding factor (CTCF) binding sites. Using Massive Parallel Sequencing, we determined the CpG methylation status of the CTCF binding sites flanking the DM1 CTG repeat in a unique collection of 70 DM1 patients and 19 families, 19 of which showed CDM with maternal inheritance, and 4 DM1 human embryonic stem cell lines (hESCs). All CDM patients, but one, showed complete methylation both upstream and downstream of the CTG repeat. The same methylation pattern was observed in two hESCs carrying maternally-derived expansions, and in one third of the classical DM1 patients with maternal inheritance of the expansion. Apart from blood samples, we analyzed 4 maternally-derived CVS that also showed methylation both up- and downstream of the repeat. In contrast, two hESC lines with paternally-derived DM1 expansions, along with all DM1 patients with paternal inheritance of the mutation, never showed methylation upstream of the CTG repeat. Three CVS with paternally-derived expansions also showed methylation downstream but not upstream of the CTG repeat. Importantly, methylation patterns did not strictly correlate with CTG tract length. Analysis of this large and comprehensive collection highlights CpG methylation at the CTCF binding sites flanking the CTG repeat as a prenatal indicator of CDM that is stronger than the repeat size itself.

2040F

Differential regulation of amyloid-β precursor protein (APP) and β-secretase 1 (BACE1) by transcription factor (TF) modulating drugs in human cells. *B.L. Bayon[1,3], K. Nho[2], B. Maloney[3], N. Chopra[3,4], D.K. Lahiri[1,3,4].* 1) Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 2) Radiology, Indiana University School of Medicine, Indianapolis, IN; 3) Psychiatry, Indiana University School of Medicine, Indianapolis, IN; 4) Program of Neuroscience, Indiana University School of Medicine, Indianapolis, IN.

TFs play a role in the survival of neuronal cells and regulate genes associated with Alzheimer's disease (AD). Expression levels of some TFs are perturbed in AD. BACE1 is responsible for the rate-limiting cleavage of APP to amyloid-β (Aβ). The TF Sp1 positively regulates APP and induces BACE1. To test the involvement of Sp1 in APP regulation, we treated human cells with Sp1 inhibitors mithramycin A (MTM) and tolfenamic acid (TA). We also transfected a variety of mammalian cell lines, a primary human neuron culture, and mixed brain cultures derived from human neurospheres with siRNAs. Minimal changes in confluence, cytotoxicity, neurite length, and neurite outgrowth were seen after Sp1 knockdown via siRNA or treatment with Sp1 modulating drugs. TA did not change Aβ40 levels, or affect APP and BACE1 levels. MTM led to a significant decrease in the expression of APP and BACE1, and significantly decreased Aβ40 levels. Neither treatment with Sp1-inhibiting drugs alone nor transfection with Sp1 siRNA alone affected cell viability of human cells illustrating that Sp1 levels can be safely reduced in cells. However, combination treatments of MTM plus TA or of Sp1 inhibition by siRNA in combination with TA treatment led to high cytotoxicity. To better understand the activation of multiple TFs and elucidate the status of signaling pathways after MTM or TA treatment, we used a TF luciferase transactivation reporter array. This array revealed decreased activation of several TFs with MTM treatment that correspond with a decrease or no change with TA treatment. These TFs were used as a preliminary screen for future assays. To understand the relationship between genetic variants in screened TFs and AD-related endophenotypes, we performed an association analysis of single nucleotide polymorphisms (SNPs) in select TFs with biomarkers in the AD Neuroimaging Initiative cohort. We identified a SNP (rs2551922) in *CREB1* significantly associated with Aβ levels in the cerebrospinal fluid. A SNP (rs11170553) in *SP1* is significantly associated with entorhinal cortex thickness and with global cortical Aβ load measured by Florbetapir PET. These associations demonstrate a possible correlation with these TFs and AD pathology. Compounds that can modify TF binding to sites on the BACE1 and APP promoters could provide a means to limit the production of Aβ and slow the symptoms of AD. These data show that modulation of TFs could potentially be a novel drug target for AD.

2041W

The *SMN1* mutation c.22-23 insA(;p.Ser8lysfs*23);causes spinal muscular atrophy by nonsense-mediated mRNA decay. *J. Bai, Y. Qu, Y. Cao, L. Ge, Y. Jin, H. Wang, F. Song.* Capital Institute of Pediatics, Bei Jing, China.

Background: Spinal muscular atrophy (SMA) is a common autosomal recessive neuromuscular disorder with an incidence of 1 in 6,000 ~ 10,000 among newborns. Most of the patients are with homozygous deletion of *SMN1* gene, and 5%-10% SMA patients are believed to have *SMN1* subtle mutations. Recently, the mutation c.22-23 insA (p.Ser8lysfs*23) has been identified as the most frequent mutation in the Chinese SMA population and be associated with a severe phenotype. In our previous study, *SMN1* mRNA and SMN protein in peripheral blood cells of this patient were much lower than that those of controls. However, the exact molecular mechanism of the mutation on the pathogenesis of SMA remain is unclear. **Objective:** The aim of this study is to investigate whether c.22-23 ins A mutation of *SMN1* gene causing SMA by triggering nonsense-mediated mRNA decay (NMD). **Methods**: Two-stage validation of NMD mechanism was supplied in our study. We firstly measured changes in the transcript levels and *SMN1* gene by real-time quantitative PCR after immortalized B-lymphoblasts of the SMA patient carrying c.22-23 insA mutation were treated with the inhibitors of NMD pathway, including puromycin and cyclohemide. Next, lentivirus-mediated knock-down of the regulator of nonsense transcripts 1(*UPF1*) was performed in the fibroblasts of the patient to detect whether the mutation lead to NMD. In our experiments, SC35 1.7Kb transcripts, a physiological NMD substrate was measured as a NMD positive gene. **Results:** The two inhibitors result in a dramatic escalation of full-length *SMN1* (*fl-*SMN1) transcript levels. And SC35 1.7kb mRNA level are also increased, which suggests that NMD pathway is suppressed by the two inhibitors. For puromycin and cycloheximide, the maximum increases of mRNA levels reach about 22- and 10-fold for *SMN1*, and about 9.6- and 7.9-fold for *SC35* in immortalized B-lymphoblasts derived from the patient, respectively. Meanwhile, lentivirus-mediated *UPF1* knock-down lead to reduction of UPF1 protein level to 22.5% compared to the negative control lentivirus. And knockdown of *UPF1* gene also promotes mRNA expression of *SC35* gene and *fl-SMN1* gene. The increases of *SMN1* and *SC35* mRNA levels reach about 4- and 7-fold in fibroblasts derived from the patient, respectively. **Conclusion:** our study provides the first evidence that c.22-23 insA mutation in *SMN1* gene triggers NMD. The SMA pathogenesis in the patient is associated with mRNA degradation of *SMN1*, but not the truncated SMN protein.

2042T

Investigating the role of SOX10-regulated alternative promoter use in Schwann cells. *E. Fogarty[1], M. Brewer[2], J. Rodriguez-Molina[3], W. Law[2], K. Ma[3], J. Svaren[4,5], A. Antonellis[1,2].* 1) Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI; 2) Human Genetics, University of Michigan, Ann Arbor, MI; 3) Cellular and Molecular Pathology Program, University of Wisconsin-Madison, Madison, WI; 4) Waisman Center, University of Wisconsin-Madison, Madison, WI; 5) Comparative Biosciences, University of Wisconsin-Madison, Madison, WI.

Remarkable genomic complexity is achieved in part by the use of multiple transcription start sites (TSSs) within a single locus. The use of alternative TSSs can be regulated through the activity of cell type-specific transcription factors and defining these interactions provides insights into the biology of the cell type of interest. Schwann cells are the myelinating cells of the peripheral nervous system and the transcription factor SOX10 is critical for their development and maintenance. Many SOX10 target genes have been functionally characterized, including other transcription factors critical for myelination and genes implicated in inherited demyelinating peripheral neuropathies. Thus, identifying SOX10 targets is relevant to Schwann cell biology and pathology. As Schwann cell gene expression has been relatively understudied, we performed cap analysis gene expression on cultured Schwann cells to globally identify TSSs. Our efforts to date resulted in the identification of 10,679 putative TSSs. To investigate novel transcriptional biology, we assessed this dataset for TSSs that map at least 5 kb from an annotated RefSeq exon, and found 407 potentially novel TSSs. Given the importance of SOX10 and to prioritize TSSs for functional characterization, we integrated these data with SOX10 and H3K4me3 ChIP-Seq and DNase-Seq datasets. This revealed 55 previously unannotated promoters that bind to SOX10: 33 within known RefSeq loci and 22 not associated with an annotated gene. Importantly, RNA-Seq of cultured Schwann cells validate the presence of these newly identified TSSs. Gene ontology analyses of the 33 known loci showed no enrichment of particular biological processes, but genes with known or predicted roles in myelinating cells were identified. For example, we revealed an alternative promoter at the demyelinating peripheral neuropathy locus MTMR2. Functional studies showed that SOX10 is necessary and sufficient for the activity of the alternative promoter and expression of the associated transcript. The resulting alternative protein isoform shows differential nuclear localization, suggesting that it may have a divergent function. We propose that this SOX10-regulated gene product explains the sensitivity of Schwann cells to loss of MTMR2, which is broadly expressed. Our data support further identification of SOX10-regulated TSSs genome-wide and indicate that this will be a fruitful approach for investigating isoform-specific gene function in Schwann cells.

2043F

Allele-specific silencing as therapeutic strategy to tightly modulate gene expression in disorders due to gene duplication: A proof-of-principle in Autosomal Dominant Leukodystrophy (ADLD). *E. Giorgio[1], A. Bartoletti-Stella[2], A. Brussino[1], C. Mancini[1], S. Cavalieri[1], E. Di Gregorio[3], L. Gasparini[4], P. Cortelli[2], S. Capellari[2], A. Brusco[1,3].* 1) Dep. Medical Sciences, University of Torino, Torino, Italy; 2) Dep. of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy; 3) "Città della Salute e della Scienza" University Hospital, Medical Genetics Unit, Torino, Italy; 4) Dep. Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy.

Allele-SPecific silencing by RNA Interference (ASP-iRNA) is a therapeutic strategy to downregulate a single mutant allele with minimal suppression of the corresponding wild-type (WT) allele. This approach has been effectively used to target autosomal dominant mutations or single nucleotide polymorphisms (SNPs) linked with aberrantly expanded trinucleotide repeats. We reasoned that ASP-iRNA could have been the paramount choice to target also genetic diseases due to gene duplication. Its main advantage is avoiding a harmful excessive downregulation of the target gene, a potential side effect of iRNA strategies. Here, we demonstrate the utility and efficacy of ASP-iRNA in Autosomal Dominant adult-onset demyelinating LeukoDystrophy (ADLD), a hereditary, progressive and fatal disorder affecting myelin in the central nervous system. ADLD is generally caused by lamin B1 (*LMNB1*) overexpression due to gene duplication. It stands to reason that a genetic modulation of lamin B1 mRNA represents a promising strategy to treat ADLD, and on the other way its excessive reduction may cause severe neurologic disorders, as shown by knockout mouse model. We choose to targeted the non-duplicated allele of the *LMNB1* gene by ASP-siRNAs, exploiting a frequent single nucleotide polymorphism (SNP) located in its coding sequence (MAF: 0.45), with the aim of reducing expression close to a wild-type. We designed and screened a siRNA library with a 19+2 geometry centred on the SNP and evaluated siRNAs efficacy and specificity by using a customized dual luciferase reporter system (modified from the psiCHECK™-2 system, Promega). We identified four siRNAs with a high efficacy (silencing 65%-87%; p<0.0001) and allele-specificity (p<0.001). These were further tested in four ADLD patient-derived fibroblast lines. Three out of four siRNAs selectively silenced the target allele (p< 0.0005) and restored *LMNB1* mRNA levels to controls values [*LMNB1* expression in patients / controls: before treatment = 2.42 (p<0.0005); after treatment =1.07 (p=ns)]. Most importantly, siRNA treatments almost completely rescued ADLD-specific splicing alterations, suggesting that ASP-iRNA may be useful in recovering cell function in ADLD. Our work is a proof-of-concept that ASP-silencing can be a therapeutic option for diseases with gene overexpression, opening new therapeutic possibilities for all Mendelian and syndromic disorders associated with gene(s) duplication.

2044W

Prenatal stress of *Cntnap2* heterozygous mice demonstrates a gene by environment interaction model for developmental brain disorders. *R.S. Alisch[1], Q. Zhang[2], K. Chen[3], S. Li[1], S. Keles[3], L.S. Papale[1].* 1) Psychiatry, Univ. of Wisconsin SOM - Madison, Madison, WI; 2) Statistics, University of Nebraska, Lincoln, NE; 3) Statistics, Biostatistics, and Medical Informatics, and, University of Wisconsin, Madison, WI.

The autism spectrum disorders refer to a group of complex neurodevelopmental brain disorders. Alterations in environmentally sensitive epigenetic mechanisms (*i.e.,* DNA methylation) have been reported in numerous genetic disorders associated with autism, suggesting that epigenetic mechanisms involving gene and environment interactions are a common pathway for many cases of autism. 5-hydroxymethylcytosine (5hmC) is a novel environmentally sensitive DNA modification that is highly enriched in post-mitotic neurons and perturbations of 5hmC are linked to brain-related disorders, leading us to hypothesize a role for 5hmC in the complex etiology of developmental brain disorders, including autism. As a first test of this hypothesis, we generated a genome-wide profile of striatal 5hmC in an autism mouse model (*i.e.,* a *Cntnap2[-/-]* homozygote) that exhibits core autism features, including deficits in striatal GABAergic signaling. Homozygous mice exhibit a genome-wide disruption of 5hmC in the orthologs of a significant number of genes implicated in human autism (*N*=68/233), suggesting a role for 5hmC-mediated epigenetic modulation in the pathogenesis of autism. Notably, *Cntnap2[+/-]* heterozygous mutant mice lack autistic-like features. Since 5hmC levels are environmentally sensitive, we next hypothesized that heterozygous mice subjected to an early environmental stress would have disruptions in 5hmC and may exhibit autistic-like behaviors similar to the homozygous mice. To test this hypothesis, we subjected heterozygous mice to seven days of variable prenatal stress and subsequently examined a variety of adult behaviors, including anxiety, depression, and sociability. Unlike wild-type littermates, prenatally stressed female heterozygous mice exhibited autistic-like social deficits. In addition, genomic profiling revealed disruptions in 5hmC throughout the striatal genome, which were similar to those seen in homozygous mice. Finally, integration of the above data with striatal gene expression data revealed that differential hydroxymethylation correlated with altered transcript levels of genes shown to contribute to the deficits in GABAergic signaling and social behavior of the homozygous mouse (*e.g.,* GABA receptors and *Oxytocin*). Together, these data demonstrate a gene by environment interaction that leads to deficits in social interactions and suggest that 5hmC likely contributes to this outcome and may be involved in the etiology of developmental brain disorders.

2045T

Altered DNA methylation in autism spectrum disorder: Sex differences and potential for biomarker development. *V. Hu[1], Y. Hong[1,2], M. Xu[1,3]*. 1) Biochemistry & Molecular Medicine, George Washington Univ School of Medicine & Health Sci, Washington, DC; 2) School of Life Sciences, Anhui Univ., Hefei City, Anhui 230601, P.R. China; 3) Shanghai East International Medical Center, Shanghai, P.R. China.

   Statement of purpose: The goals of this study were to: 1) investigate changes in the DNA methylation profiles of lymphoblastoid cell lines (LCL) derived from individuals with autism and their respective sex-matched siblings; 2) describe ASD-relevant pathways and functions that may be impacted by the differentially methylated genes; 3) identify sex-dependent differences in methylation patterns; and 4) examine the potential for class prediction based on epigenetic biomarkers in peripheral tissues. Methods: To reduce the effects of clinical heterogeneity on the methylation analyses, we subtyped individuals with ASD on the basis of cluster analyses of scores from the Autism Diagnostic Interview-Revised diagnostic instrument as previously described (Hu and Steinberg, Autism Research 2:67, 2009), and limited the cases for this study to those with severe language impairment. All diagnostic and phenotypic data was obtained through the Autism Genetic Resource Exchange (AGRE), from which LCL were obtained for the methylation analyses. LCL from sex-matched unaffected siblings were used as controls. Affymetrix Human Promoter 1.0R GeneChips were used to investigate differentially methylated promoter regions in the DNA of cases and controls. For each sample, methylation-enriched DNA prepared using a MeDIP protocol and input (total) DNA were hybridized on separate chips. Partek GS Software was used to identify differentially methylated regions using the workflow recommended by Affymetrix for methylation analyses on tiling arrays, and Multi-experiment Viewer (MeV) microarray analysis software was used for additional statistical and class prediction analyses, with the samples divided into a training set for identification of putative biomarkers, and a test set for validation of the biomarkers. Pathway and functional analyses of the differentially methylated genes were performed using Ingenuity Pathway Analysis (IPA) software. Summary of results and conclusion: This study revealed significant differences in DNA methylation in LCL derived from individuals with ASD and from their sex-matched siblings without ASD and showed, for the first time, sex-dependent differences in methylation profiles. Class prediction analyses further demonstrated that limited panels of differentially methylated genes may be used to classify ASD-affected and unaffected individuals with moderately high sensitivity and specificity, suggesting the potential for their use as epigenetic biomarkers.

2046F

MicroRNA biomarkers in whole blood for diagnosis of autism spectrum disorder. *M. Nakata[1], R. Kimura[1], K. Tomiwa[2], T. Awaya[3], T. Kato[3], Y. Funabiki[4], T. Heike[3], M. Hagiwara[1]*. 1) Anatomy and Developmental Biology, Kyoto University Graduate School of MedicineTodaiji Medical and Educational Center, Kyoto, Japan; 2) Todaiji Medical and Educational Center, Nara, Japan; 3) Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; 4) Human Coexistence, Kyoto University Graduate School of Human and Environmental Studies, Kyoto, Japan.

   Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and a pattern of repetitive stereotyped behaviors. Many studies have been reported that high rates of mental health problems are observed commonly in ASD, especially depression and anxiety. Therefore, early detection is important to intervene early and prevent later psychiatric symptoms. Despite all the research efforts, however, reliable diagnostic biomarkers are not established as it stands. MicroRNAs (miRNAs) are a class of short non-coding RNAs that regulate gene expression. Increasing evidence has shown that miRNAs may play essential roles in neurodevelopmental disorders. In this study, we evaluated blood-based microRNAs as novel biomarkers for diagnosis of ASD. We determined miRNA expression profiles of ASD using Agilent miRNA microarray analysis in discovery sample set (30 ASD patients and 30 Controls). An independent replication sample set was used for the biomarker validation by real-time RT-PCR. Further we constructed ROC (Receiver Operating Characteristics) curves to determine the sensitivity and specificity of individual miRNA as a diagnostic biomarker. We found that three miRNAs had significantly altered in ASD patients compared to healthy controls and also got a good ROC curve according to the accepted classification of biomarker utility. These miRNAs were expressed in not only blood but also human brain. Our findings provide novel and reliable biomarkers for diagnosis of ASD.

2047W

Functional study of lncRNAs that associate with synaptic vesicles in autism spectrum disorder. *L. Zhao[1], W. Ju[2], M. Zhong[1], Y.H. Yu[1], W.T. Brown[2], N. Zhong[1,2].* 1) Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China; 2) 1.New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA.

Studies have demonstrated that epigenetic regulation of long non-coding RNA (lncRNA) is involved in brain development and neuro-psychiatric diseases, including autism spectrum disorder (ASD). However, there are few reports on the lncRNAs that associate with synaptic vesicle (SV) cycling. We have previously reported differential expression of SV-associated lncRNAs/mRNAs among 25 pairs (ASDs *vs.* controls) of children's blood. We further validated 15 pairs of synaptic lncRNAs-mRNAs with an independent set of 100 cases, including 25 with ASD compared to age- and gender-matched controls that include 25 of attention deficit hyperactivity disorder, 50 non-syndromatic mental retardation, and 25 normal controls. In addition to ASD blood specimens, two fibroblast cell lines in four groups (ASD, normal control, fragile X with autism, and fragile X without autism) were subjected to RNA sequencing. Our data showed 26 SV-lncRNAs were differentially expressed. These eight fibroblast lines were applied to generate iPSC lines, in addition to two epithelial lines (one from a child with ASD, and the other from a normal control). From the 10 fibroblast re-programmed iPSCs, NPCs were differentiated. After comparing lncRNAs between ASD *vs.* non-ASD in neuronal progenitor cells (NPCs) and between lncRNAs identified from NPC *vs.* iPSC, bioinformatic analyses were conducted and the involved pathways were characterized. Two SV-associated lncRNAs have been subjected to functional analysis. The lncRNA-SYT15 was knocked out with CRESPR-Cas9 system that showed decreased lncRNA-SYT15 resulted deduction of mRNA-SYT15 in both neuronal and trophoblast cell cultures, suggesting lncRNA-SYT15 cis-regulates the mRNA-SYT15 transcription that was also confirmed by over expressed lncRNA-SYT15. Longitudinal assessment of lncRNBA-STX8 has demonstrated that lncRNA-STX8 is differentially expressed from the first trimester to the late stage of pregnancy, parallel to the differential expression pattern of mRNA-STX8, suggesting that lncRNA may have a regulatory function during the entire pregnant period. Our studies have provided evidence that SV-lncRNA is one of the epigenetic regulators, likely involved in SV cycling that may play a pathogenic role in ASD. .

2048T

Histone deacetylase deficiency leads to changes in hippocampal histone lysine profiles and is involved in anxiety and depression-like behavior in mice. *S. Röh[1], C.T. Wotjak[2], T. Strom[4], W. Wurst[3], E.B. Binder[1], J.M. Deussing[2], M. Jakovcevski[2].* 1) Max Planck Institute of Psychiatry, Department of Translational Research, Munich, Germany; 2) Max Planck Institute of Psychiatry, Department of Stress Neurobiology and Neurogenetics, Munich, Germany; 3) Helmholtz Center Munich, institute of Developmental Genetics, Neuherberg, Germany; 4) Helmholtz Center Munich, Institute of Human Genetics, Neuherberg, Germany.

It has been shown that broad spectrum deacetylase inhibitors (HDACi) can alter and improve depression-related symptoms in humans and that stressors affect the epigenome by changing histone acetylation. Therefore, understanding the essential molecular mechanisms of how single histone deacetylases (HDACs) contribute to ameliorate stress-related pathologies can improve the efficacy of current antidepressant drugs through increased specificity. In our study, we focus on one specific histone deacetylase (HDAC), a common target of HDACi, and its effects on the behaviour of adult mice under non-stressed and stressed conditions. Using a loss of function mouse model (Hdac knock-out mice), we tested behavior as well as epigenomic and gene expression changes by next generation sequencing. Our results suggest that HDAC is involved in basal, non-stressed anxiety-related behavior, the locomotor response to novelty, and anhedonia, indicating altered stress resilience. Hdac knockout mice from two independent batches were more susceptible to the effects of chronic social defeat stress. Altered behavior was accompanied by changes in bulk histone acetylation levels for several histone marks in knock-out mice at baseline and in stressed wild-type mice. Accordingly, ChIPseq data of histone acetylation show massive acetylation changes between all conditions. In stark contrast, transcriptional activity was only mildly affected and rather uncorrelated to acetylation changes between all conditions. To shed light on the question of whether this outcome could be explained by cell type/specific effects of acetylation, we have generated data sets to investigate this further.

2049F

Establishment of a strong link of smoking with cancer pathogenesis through DNA methylation analysis. *Y. Ma[1], M. Li[1,2].* 1) State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China; 2) Institute of NeuroImmune Pharmacology, Seton Hall University, South Orange, NJ USA.

Tobacco smoking is a well-documented risk factor for various cancers, which are regulated by complex mechanisms, including DNA methylation and other epigenetic reprogramming. Previous epigenome-wide association studies revealed a large number of methylation sites associated with smoking. However, to date, no study has established a link between smoking-associated DNA methylation and cancer. To attack this issue, we identified two sets of genes methylated by smoking from blood and buccal samples, and then determined whether these methylated genes were enriched in pathways implicated in the pathogenesis of cancers. For the methylated genes from the blood samples, we identified 57 significantly enriched pathways; many of them are reported to be associated with different types of cancer. Among these pathways, 11, including RAR activation, actin cytoskeleton signalling, and aryl hydrocarbon receptor signaling, were also significantly overrepresented by the methylated genes in the buccal samples. Furthermore, our findings indicate that the aryl hydrocarbon receptor signalling pathway plays an essential role in the initiation of smoking-attributable cancer. Based on the findings from the current study and prior biological evidence, we propose a schematic model to elucidate the biological effects of smoking on cancer pathogenesis. Finally, we constructed a subnetwork consisting of 48 non-redundant genes in the 11 oncogenic pathways detected in both the blood and buccal samples. Of them, several genes, such as *DUSP4*, *AKT3*, *SMARCA4*, and *SMAD6*, show strong evidence for involvement in smoking-related lung cancer. In conclusion, our findings provide the first robust and systematic evidence supporting the idea that smoking is indeed an important inducer of cancer at an epigenetic level.

2050W

Genes and memory: A comprehensive study of the association between quantitative trait loci and memory in independent samples. *V.V. Vukojevic[1,2,5], A.M. Milnik[1,4,5], P.D. Demougin[1,2], F.P. Peter[1,2], D.Q. de Quervain[3,4,5], A.P. Papassotiropoulos[1,2,4,5].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Department Biozentrum, Life Sciences Training Facility, Klingelbergstrasse 50/70, 4056 Basel, Switzerland; 3) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 4) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland; 5) Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel,CH-4055 Basel, Switzerland.

Background: Complex traits, e.g. cognitive performance, exhibit a striking amount of phenotypic variability. A large amount of this phenotypic variation across population is due to the underlying genetic architecture. Therefore, understanding the genetic basis of the quantitative trait variation, e.g. quantitative trait loci (QTL), is one of the major challenges in current biology of complex cognitive traits. **Analysis:** In the present study we performed a genome wide analysis of DNA methylation and expression cis-QTLs (meQTLs and eQTLs, respectively) in a European sample of healthy young adults (N = 533). We validated our findings in an independent sample of similar size and origin (N = 319), and further investigated the association of the significant loci with episodic and working memory performance across two samples. Finally, we extended our analysis to a third sample of an African decent (N = 380), and investigated the amount of conserved QTLs. **Results:** The cross-investigation of cis-meQTLs in the discovery and replication sample revealed 154'135 unique meQTLs (corrected for 1.4×10^{11} tests; 448'243 CpGs, 313'505 SNPs). These meQTLs were significantly associated to 19'929 unique CpGs, targeting a total of 8'599 genes (FDR of 5%). The same set of SNPs was additionally enriched in eQTLs: 37'721 SNPs associated to 6'578 transcripts (4'330 genes, 67% concordance with meQTLs overlapping genes; FDR of 5%). Extended analysis of QTLs in an African sample revealed a significant overlap with European samples, suggesting a high level of conservation across populations (92'399 meQTLs, 12'217 CpGs; FDR of 5%). A second level analysis revealed that meQTL associated CpGs were pronouncedly related to cognitive phenotypes across two investigated European samples (310 CpGs, 286 genes; FDR of 5%). **Discussion:** Our findings suggest that QTL associated CpGs also show an enrichment of significant hits in association analyses with complex traits. This might be of special importance for studies that rely on proxy tissues, like in the case of cognition and psychiatric disorders.

2051T

Alcohol use is associated with dose-dependent DNA methylation and alternative transcript expression of synaptic genes. *R.P. Cervera-Juanes, L.J. Wilhem, C. Huang, R. Shah, S. Gonzales, K.A. Grant, B.M. Ferguson.* Oregon National Primate Research Center, Oregon Health and Science University, Beaverton, Oregon.

Alterations in DNA methylation have been associated with alcohol exposure, and proposed to contribute to escalating alcohol abuse; however the specific molecular mechanisms involved remain obscure. Clarifying the genes regulated by DNA methylation and distinguishing their contributions to alcohol dependence are critical for fully understanding and treating alcohol use disorders (AUDs). In addition, understanding the effects of the dose of alcohol consumed will be critical for developing more specific approaches in the treatment of AUDs. We used the rhesus macaque alcohol self-administration model to investigate the relationship between the amount of alcohol consumed and the effects on DNA methylation and gene expression in the nucleus accumbens (NAc). We combined a genome-wide approach for the targeted selection of CpG rich genomic regions, bisulfite sequencing and a regional clustering to identify significant differentially methylated regions (DMRs) among subjects that differed in their alcohol consumption levels. The methylation levels of 2.7 million CpGs were compared between light-moderate and heavy drinking subjects. We identified 17 DMRs, 15 of them were located in the gene body, coinciding with predicted enhancer and promoter regions. In 14 of these DMRs, the average CpG methylation level was significantly correlated with the average daily amount of ethanol consumed. Eight of the DMRs map to genes encoding proteins with functions linked to cell adhesion, cytoskeleton binding, neurotransmitter release and receptor trafficking, underlining their relevance to alcohol-associated synaptic adaptation. Moreover, gene expression data revealed a prominent role of the DMRs in regulating alternative transcript expression of synaptic genes. For instance, *JAKMIP1*, which is known to regulate GABA-B and NMDA receptor signaling, is associated with a DMR that was significantly correlated with both the amount of ethanol consumed and the differential expression of two alternative 1st exons. Thus, our results suggest that some of the neural adaptations induced by alcohol are mediated by changes in DNA methylation and alternative transcript expression of synaptic genes, such as *JAKMIP1*. Moreover, the genes identified by this study implicate new targets for the development of alcohol dependence treatments. This study was supported by NIH grant U01AA020928, R24AA019431 and OD011092.

2052F

Hippocampal DNA methylation changes affect oxidative stress genes in fetal alcohol-exposed adolescent mice. *E.J. Chater-Diehl, B.I. Laufer, S.M. Singh.* Department of Biology, Western University, London, Ontario, Canada.

Alcohol use during pregnancy can lead to a range of neurological abnormities termed Fetal Alcohol Spectrum Disorders (FASD). The mechanisms by which alcohol (ethanol) induces FASD are poorly understood. We have previously developed a model to generate FASD-like mice showing characteristic learning and memory impairment. Our laboratory has explored whether epigenetic modifications are related to these changes. We found persistent fetal-ethanol-exposure-induced changes in histone modification, DNA methylation, and gene expression of oxidative stress genes in adult mouse hippocampus. Oxidative stress is a key component of the primary effects of ethanol in FASD etiology. It is unclear if these changes arose later in life as a compensatory long-term response, or were maintained from the initial early-life exposure. Here, we hypothesized that changes to oxidative stress genes were also present in adolescent mice. Mouse pups were injected with saline or ethanol on postnatal days 4 and 7, the period neurodevelopmentally equivalent to human trimester three. At 21 days of age, hippocampi were isolated and used for gene and miRNA expression microarray, and methylated DNA immuno-precipitation & sequencing (MeDIP-seq). The gene expression analysis on the hippocampus identified enrichment of oxidative phosphorylation genes (*Cox7b2*, *Ndufa1*, *Ndufa7*, *Uqcrq*) as well as genes implicated in alcoholism and olfactory transduction. Long non-coding RNA expression changes and reciprocal microRNA/target gene expression changes were also identified. We next performed MeDIP-seq. Similar to our previous adult data, similar oxidative stress pathways were implicated, through few genes overlapped with the expression data. Our results support a growing body of evidence that early ethanol exposure, epigenetics, and oxidative stress are related late into life. Specifically, these data support a role for changes in oxidative stress gene expression and DNA methylation in FASD at several stages. The emergence of this theme from our group and others raises the possibility of diagnostic and therapeutic interventions for FASD individuals. It is also relevant beyond alcohol research, furthering our understanding of the scale and duration of environmental effects on the (epi)genome. .

2053W

TET1-mediated epigenetic modulation in stress-induced response. *Y. Cheng[1], Z. Miao[2], L. Chen[1], Y. Li[1], L. Lin[1], B. Yao[1], Z. Wang[1], N. Xin[2], L. Huang[1], X. Xu[2], P. Jin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; 2) The Institute of Neuroscience, Soochow University, Suzhou 215006, P.R. China.

Depression is a common but serious mental disorder that characterized by a pervasive and persistent low mood. Depression can be triggered by various causes, in which chronic stress has been shown to directly link to depression. 5-hydroxymethylcytosine (5hmC) has been reported highly enriched in the central neural system, and is now recognized as an important epigenetic mark in developmental and neurological disease. In the present study, we show that chronic restraint stress (CRS) can induce the depression-like behavior in wild type (WT) mice, and results in the global 5hmC decrease in prefrontal cortex (PFC). We profiled genome-wide distribution of 5hmC and identified stress-induced dynamically hydroxymethylated loci in WT mice (WT-SI-DhML). Gene ontology (GO) analyses indicated these WT-SI-DhML were located in the genes that involved in the pathways associated with major depressive disorder. We then utilized global Ten-eleven translocation protein 1 knockout (Tet1-KO) mice to investigate the role of Tet1 in stress-response. Interestingly we found that the loss of Tet1 led to the resistance to CRS. To link the 5hmC status with stress-resistant phenotype, we further profiled 5hmC using the PFC DNA samples of Tet1 KO mice, and identified phenotype associated SI-DhML (PA-SI-DhML). Go analyses indicated that these PA-SI-DhML were significantly associated with the genes involved in the response to stimulus and neuronal function as well as those involved in axon guidance. Subsequent motif search revealed a strong enrichment of hypoxia-induced factor (HIF) binding motifs among PA-SI-DhML. Biochemically we demonstrated the physical interaction between TET1 and HIF1α under CRS condition, although CRS had no impact on the expression of TET1 and HIF1α. By using anti HIF1α antibody, our ChIP-seq data of WT PFC identified differentially changed HIF1α binding regions (HBRs) upon CRS. The HBRs with increased ChIP-seq reads (Stress/Control) were highly associated with WT-SI-DhML and PA-SI-DhML, suggesting TET1 may be recruited by HIF1α upon CRS and influence the 5hmC dynamics. These results together suggest that TET1, by interacting with HIF1α, plays important role(s) in stress-induced responses.

2054T

The efficacy of oxBS and the Infinium MethylationEPIC array to characterize variations in 5-methylcytosine and 5-hydroxymethylcytosine between different human tissue types. *J. Gross[1], F. Lefebvre[2], P.E. Lutz[1], F. Bacot[3], D. Vincent[3], G. Bourque[3,4], G. Turecki[1,5].* 1) McGill Group for Suicide Studies, Douglas Mental Health Univ Inst, Montreal, Quebec, Canada; 2) Canadian Centre for Computational Genomics, Montreal Node, McGill University, Montreal, Quebec, Canada; 3) McGill University and Genome Quebec Innovation Centre, McGill University, Montreal, Quebec, Canada; 4) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 5) Department of Psychiatry, McGill University, Montreal, Quebec, Canada.

Investigating 5-methylcytosine (5mC) has led to many hypotheses regarding molecular mechanism underlying human diseases and disorders. Many of these studies, however, utilize bisulfite sequencing, which cannot distinguish 5mC from its recently-discovered oxidative product, 5-hydroxymthylcytosine (5hmC). Furthermore, previous array-based technologies do not have the necessary probes to adequately investigate both modifications simultaneously. Here, we use DNA from post-mortem human brain, human blood, and human saliva to validate the efficacy of oxidative bisulfite conversion and Illumina's Infinium MethylationEPIC array to analyze 5mC and 5hmC. We show the presence of 5mC and 5hmC at greater than 650,000 and 450,000 relevant loci, respectively, in the human brain. While the number of loci containing 5mC remains stable across tissue-types, the number of 5hmC-containing probes drops drastically in human blood and saliva. Nevertheless, 5mC- and 5hmC-containing probes are equally distributed across chromosomes, genomic features, and, in particular, distal regulatory regions. Our results confirm the sensitivity of the EPIC array to detect 5mC and 5hmC in all three tissue-types and show this methodolgy to be an important and cost-effective advancement in profiling cytosine modifications in human phenotypes.

2055F

Beyond DNA methylation: Characterizing H3K27ac using ChIP-seq in post-mortem human brain tissue in a multi-omics analysis of Alzheimer's disease. *S.J. Marzi[1], T. Ribarska[4], A. Smith[2], K. Lunnon[2], L. Schalkwyk[3], J. Mill[1,2].* 1) King's College London, London, United Kingdom; 2) University of Exeter, Exeter, United Kingdom; 3) University of Essex, Colchester, United Kingdom; 4) DKFZ, Heidelberg, Germany.

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and progressive neuropathology in specific regions of the brain. As the most common form of dementia that affects an increasingly ageing population it is a growing public health problem. Disease progression is accompanied by an accumulation of amyloid β and neurofibrillary tangles across the cortex, however the underlying aetiology is still not fully understood. Several studies have implicated epigenetic processes in disease progression. Our lab recently identified widespread changes in DNA methylation associated with AD, and in particular reported a robust association with cortex-specific methylomic variation in the *ANK1* gene, which has now been replicated across 4 independent cohorts. Building on this work we used chromatin immunoprecipitation sequencing (ChIP-seq) to profile differences in histone acetylation (H3K27ac) – a histone modification found in active promoter and enhancer regions – associated with AD in an overlapping set of post-mortem entorhinal cortex samples from AD cases and matched controls (n=50). Modelling normalised read counts in ~155,000 robustly detected peak regions, we identify ~550 differential peaks between AD cases and controls. Several of these are associated with neuronal pathways and most are explained by differences in neuronal vs glia proportion estimates, which were obtained based on DNA methylation profiles. Analyses correcting for neuronal proportion, sex and age at death result in 21 differential peaks at Bonferroni-corrected significance level of a=0.05. This is the largest study of H3K27ac undertaken in post-mortem brain samples to date and to our knowledge the first one of Alzheimer's Disease. In addition to histone acetylation the samples have been profiled for DNA methylation, DNA hydroxymethylation and genotypes, which we are currently using to identify co-localized signals identified across multiple layers of (epi-) genetic information, and characterize functional consequences of non-coding variation.

2056W

Quantitative DNA methylation analyses using digital droplet PCR sensitively detects surreptitious smoking in adolescents. *R.A. Philibert[1,3], A. Andersen[1], S. McElroy[1], K. Vercande[1], L. Delevan[1], E. Papworth[1], N. Hollenbeck[1], W. Lugo-Morales[1], M. Gerrard[2], R. Gibbons[2].* 1) Psychiatry, University of Iowa, Iowa City, IA; 2) Dept of Psychology, University of Connecticut, Stoors, CT; 3) Behavioral Diagnostics Inc, Iowa City, IA.

The initiation of smoking is associated with the initiation of other externalizing disorders including drinking, risky sex and conduct disorder spectrum illness. Psychoeducational interventions can arrest the escalation of each of these disorders. Unfortunately, their implementation is hindered by the unreliability of self-report- which can range as high as 50% in high-risk populations. Previously, we have shown that quantitative DNA methylation analyses can accurately detect smoking in adult populations. However, whether these measures can be used to detect smoking in adolescents is not known. We present the initial results from a longitudinal examination of 450 high school students at elevated risk for cigarette smoking using repeated biological measures including exhaled carbon monoxide, ELISAs, and digital droplet PCR (ddPCR) assessments of methylation in DNA from whole blood. According to self-report, only 4 of the first 192 sophomores surveyed reported any type tobacco consumption in the past week. However, 17 of the first 213 serum samples from sophomores subjects show elevated cotinine levels with many levels clearly consistent with active smoking. Similarly, only 8 of the first 192 sophomores surveyed report any use of THC in the week prior to intake, but at least 26 of the 213 first serum samples surveyed showed elevated serum levels of tetrahydrocannabinol. Finally, summative smoking history of the first 184 subjects was assessed via ddPCR measurement of cg05575921 methylation, an established epigenetic biomarker of cigarette consumption in adults. All subjects with established histories of smoking via self-report or through clear serological detection demonstrated significant demethylation at cg05575921. Logistic regression analysis demonstrates that a model that incorporates self-report, parent smoking status and serological status significantly predicts DNA methylation status. In summary, these data demonstrate that adolescents show high levels of unreliable self-report as evidenced by cross section serological assessment. Exploratory methylation analyses suggest that the true rate of smoking in these adolescent subjects may be even higher. We conclude that biological indexes that incorporate DNA methylation measures may be useful in assessing smoking histories in adolescent subjects and that further studies to examine the sensitivity and specificity of methylation analyses for the detection of adolescent smoking are in order.

2057T

Epigenetic effects of glucocorticoids exposure during hippocampal neurogenesis. *N. Provencal[1], J. Arloth[1,2], C. Anacker[3], A. Cattaneo[4,5], T. Klengel[6], N.S. Mueller[2], C.M. Pariante[4], E.B. Binder[1].* 1) Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; 2) Institute of Computational Biology, Helmholtz Zentrum München, Germany; 3) Columbia University, and Division of Integrative Neuroscience, New York State Psychiatric Institute/Research Foundation for Mental Hygiene, Inc., New York; 4) Psychological Medicine Dept, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; 5) IRCCS Fatebenefratelli, University of Brescia, Brescia, Italy; 6) Psychiatry Dept, School of Medicine, Emory University, Atlanta and McLean Hospital, Harvard Medical School, Belmont, Massachusetts.

Exposure to early life stress (ELS) is a well-known major risk factor for developing psychiatric and behavioural disorders later in life. A growing body of evidence indicates that exposure to ELS can lead to long lasting changes in a number of systems including the endocrine system, the immune system and brain structure and function. However, our understanding of the mechanisms underlying these effects is limited. One proposed mechanism that might lead to some of these long-lasting effects is that excessive glucocorticoids (GC) release after ELS exposure induces long-lasting epigenetic alterations in important regulatory genes. Indeed, accumulating evidences suggest that epigenetic mechanisms are in part responsible for the embedding of ELS where the type and timing of stress exposure are important moderating factors. Using human hippocampal progenitor cells (HPCs) and multi-omic data analysis integrating gene expression, DNA methylation (5mC) and hydroxymethylation (5hmC) at a genome-wide level, we identified long-lasting 5hmC and 5mC alterations induced by GCs exposure during neurogenesis, where a significant portion of these marks were maintained after neuronal differentiation. Moreover, the sites showing GC-induced methylation changes are enriched in regulatory regions as well as in genes differentially methylated during fetal brain development as well as in genes previously associated with child abuse in human hippocampus and blood cells. To some extend, they also reflect epigenetic changes induced by acute GCs exposure in human blood cells. Together, these results suggest that GC-induced epigenetic alterations in HPCs might reflect GC actions during ELS and be in part responsible for the increased risk for psychopathology. A mechanistic understanding of the long-term epigenetic consequences of early life stress using both, cellular models where we can study in depth the molecular mechanisms at play, as well as patients exposed to ELS may allow novel, targeted intervention and prevention strategies for behavioral, psychiatric and other stress-associated disorders.

2058F

Epigenome-wide association analysis of DNA methylation in panic disorder. *M. Shimada-Sugimoto[1], T. Otowa[2], T. Miyagawa[1,3], M. Tochigi[4], N. Sugaya[5], Y. Kawamura[6], T. Umekage[7], H. Kaiya[8], K. Kasai[9], H. Tanii[10], Y. Okazaki[11], K. Tokunaga[1], T. Sasaki[12].* 1) Department of Human Genetics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan; 2) Graduate School of Clinical Psychology, Teikyo Heisei University Major of Professional Clinical Psychology, Tokyo, Japan; 3) Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 4) Department of Neuropsychiatry, Teikyo University School of Medicine, Tokyo, Japan; 5) Department of Epidemiology and Public Health, Graduate School of Medicine, Yokohama City University, Kanagawa, Japan; 6) Department of Psychiatry, Shonan Kamakura General Hospital, Kanagawa, Japan; 7) Division for Environment, Health and Safety, The University of Tokyo, Tokyo, Japan; 8) Panic Disorder Research Center, Warakukai Med. Corp., Tokyo, Japan; 9) Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 10) Department of Psychiatry, Institute of Medical Life Science, Graduate School of Medicine, Mie University, Mie, Japan; 11) Department of Psychiatry, Koseikai Michinoo Hospital , Nagasaki, Japan; 12) Department of Physical and Health Education, Graduate School of Education, The University of Tokyo, Tokyo, Japan.

Panic disorder (PD) is a major anxiety disorder and characterized by recurrent unexpected panic attacks and anticipatory anxiety. According to previous twin and family studies, PD is considered to be a multifactorial disorder emerging from interactions among multiple genetic and environmental factors. To date, genetic studies reported several susceptibility genes with PD. Nevertheless, the pathogenesis of PD remains to be clarified. Epigenetics is considered to play an important role in etiology of complex traits and diseases, and DNA methylation is one of the major forms of epigenetic modifications. It has been reported that DNA methylation may play a certain role in pathogenesis of psychiatric disorders. In this study, we performed an epigenome-wide association study of DNA methylation in PD. The methylation levels of DNA samples extracted from peripheral blood of 48 PD patients and 48 age- and sex-matched healthy control individuals were studied. We applied the stringent data filtering and data normalization method for an accurate quantification of DNA methylation. Totally, methylation levels of 376,602 probes were determined. As a result, 42 CpG sites showed significant differences of methylation between PD and healthy control subjects at 5% FDR correction, though the methylation differences were relatively small. Many of the significant CpG sites (30/42 CpG sites) were located upstream of genes and such 30 CpG sites were all hypomethylated in PD subjects. According to previous studies, hypomethylation of upstream of gene region is often associated with higher gene expression. Therefore, the hypomethylation of the 30 CpG sites may be related to higher expression of the annotated genes. Additionally, we predicted cell mixture distributions of leukocyte subsets using DNA methylation data in order to examine the possibility that the significant associations were derived from a bias of leukocyte subsets' proportions. A pathway analysis using genes annotated to the significant 42 CpG sites revealed several intriguing pathways. Although future studies with larger number of samples are necessary to confirm the small methylation abnormalities related to PD, there is a possibility that several CpG sites might be associated with PD.

2059W

Reducing epigenetic heterogeneity in autism spectrum disorder by sub-stratifying based on genetic variants. *M.T. Siu[1], D.T. Butcher[1], S. Choufani[1], A.L. Turinsky[1,2], C. Cytrynbaum[1,3,4], D.J. Stavropoulos[5,6], S. Walker[7], Y. Lou[1], S.W. Scherer[1,4,7], M. Brudno[1,2,8], R. Weksberg[1,3,4,9,10].* 1) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 2) Centre for Computational Medicine, Hospital for Sick Children, Toronto, ON, Canada; 3) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, ON, Canada; 6) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 7) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 9) Institute of Medical Science, School of Graduate Studies, University of Toronto, Toronto, ON, Canada; 10) Department of Pediatrics, University of Toronto, Toronto, ON, Canada.

Autism spectrum disorder (ASD), the most common pediatric neurodevelopmental disorder (NDD), demonstrates a significant amount of etiologic and phenotypic heterogeneity. Over 200 ASD-risk genes have been identified, but genetics alone can only account for ~25% of cases. Aberrant epigenetic mechanisms are known to negatively impact biological pathways important for normal brain development (e.g. Angelman syndrome). A role for epigenetic dysregulation in ASD etiology is suggested by the number of ASD-risk genes that function as epigenetic regulators. Some studies have also identified epigenetic alterations, specifically DNA methylation (DNAm), in ASD. However, the impact of these findings is limited by small sample sizes and inconsistent results across studies, likely due to variance across tissue types, different analytical methods and the heterogeneity of the ASD groups examined. In order to investigate the role of epigenetic dysregulation in ASD, genome-wide DNAm, the most commonly assessed epigenetic mark, was measured using the Illumina Infinium HumanMethylation450 BeadChip array in blood from a heterogeneous ASD group (n=52). This was contrasted with individuals substratified based on known genetic alterations, forming more homogeneous comparison groups; DNAm was also assessed in blood from patients with heterozygous loss of function mutations in *CHD8* (*CHD8*[+/-]; n=7), a chromatin modifier, and 16p11.2 deletions at the 600 kb risk locus (16p11.2del; n=9). These represent two of the most common genetic alterations associated with a significantly increased risk for ASD. No ASD-specific DNAm patterns were found to fully distinguish heterogeneous ASD cases from controls. However, unique DNAm patterns were identified for the *CHD8*[+/-] and 16p11.2del groups when compared with controls. These DNAm patterns consist of specific sets of differentially methylated sites (adjusted p<0.05, absolute difference >5%) comprising a DNAm signature. These signatures classified individuals with *CHD8*[+/-] mutations and 16p11.2del distinctly from each other, as well as from the heterogeneous ASD group and controls. Furthermore, these signatures properly classified additional *CHD8*[+/-] sequence variants, 16p11.2del cases and independent controls. These data demonstrate an improved method of molecular classification of a heterogeneous NDD such as ASD and will help to enhance our understanding of the underlying biological mechanisms of this disorder.

2060T

Differential DNA methylation at birth associated with psychiatric phenotype in individuals with 22q11.2 deletion syndrome. *S. Weinsheimer[1], A. Starnawska[2], C. Søholm-Hansen[3], T. Sparsø[1], W. Mazin[1], L. Olsen[1], M. Bertalan[1], A. Buil[1], J. Bybjerg-Grauholm[3], M. Bækvad-Hansen[3], D. Hougaard[3], M. Nyegaard[2], T. Werge[1], iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research.* 1) Institute for Biological Psychiatry, Mental Health Center Sankt Hans, Roskilde, Denmark; 2) Aarhus University, Aarhus, Denmark; 3) Statens Serum Institute, Copenhagen, Denmark.

Individuals with 22q11.2 deletion syndrome (DS) have an increased risk of comorbid psychiatric disorders including schizophrenia, attention deficit hyperactivity disorder, depression, as well as mental retardation. Most carriers of the 22q11.2 deletion have the long 3 Mbp form of the hemizygous deletion; however, there remains a large variation in the development and progression of psychiatric disorders, which suggests that alternative factors such as epigenetic dysfunction contribute to the pathogenesis. **Purpose:** We performed the first epigenome-wide association study (EWAS) in 22q11.2 DS to investigate DNA methylation associated with psychiatric phenotype. **Methods:** DNA methylation was measured from neonatal dried blood spots using the Illumina 450k methylation array in a cohort of 164 individuals with 22q11.2 DS, including 48 diagnosed with a psychiatric disorder. We used regression models to identify differentially methylated CpG sites with adjustment for sex, type of 22q11.2 deletion and age of the blood spot. **Results:** Among several CpG sites with p-value < 10^{-6}, we identified cg23546855 (p-value = 2.15×10^{-7}) mapping to *STK32C* to be associated with psychiatric diagnosis in 22q11.2 carriers. Pathway analysis of the top findings resulted in the identification of ten Gene Ontology pathways to be significantly enriched (p-value < 0.05 after Benjamini-Hochberg correction) including 'Neurogenesis', 'Neuron development', 'Neuron projection development', 'Neuron projection morphogenesis', and 'Cell morphogenesis involved in neuron differentiation'. Additionally, we identified differentially methylated CpG sites in *LRP2BP* (p-value = 5.37×10^{-8}) to be associated with mental retardation, *TOP1* (p-value = 1.86×10^{-7}) with behavioral disorders, *NOSIP* (p-value = 5.12×10^{-7}) with disorders of psychological development, and *SEMA4B* (4.02×10^{-7}) with schizophrenia. In conclusion, our study suggests an association between differential DNA methylation at birth and development of psychiatric disorder later in life in 22q11.2 DS individuals.

2061F

Epigenome-wide changes in hypothalamus following enhanced maternal care reprogram stress responses in early life. *D. Yasui[1], A. Vogel Ciernia[1], K. Dunaway[1], A. Singh[4], T. Baram[4], J. LaSalle[1,2,3].* 1) UC Davis Med Sch Dept of Med Microbiology & Immunology, Davis, CA; 2) UC Davis Genome Center, Univ California Davis, Davis, CA; 3) UC Davis MIND Institute, Sacramento, CA; 4) UC Irvine, Department of Anatomy and Neurobiology, Irvine, CA.

In mammals, the hypothalamic-pituitary-adrenal (HPA) circuit regulates stress responses. In this "augmented maternal care" (AMC) paradigm, enhanced licking and grooming from postnatal days 2-9 (P9) leads to reduced corticotrophin releasing hormone (Crh) expression in the hypothalamic paraventricular nucleus (PVN) thereby attenuating the hypothalamic-pituitary-adrenal (HPA) circuit throughout development. We hypothesize that AMC epigenetically sets future gene expression levels and thus behavior by altering specific DNA methylation patterns. In support of this hypothesis, *Crh* promoter methylation levels were found to be significantly higher in AMC than control PVN when assayed by bisulfite pyrosequencing analysis, consistent with *Crh* repression. However, *Crh* methylation levels were unaffected in thalamus where HPA responses are not engaged, suggesting that elevated *Crh* methylation levels induced by AMC are specific to the PVN. To identify genome wide DNA methylation changes produced by AMC, whole genome bisulfite sequencing (WGBS) analysis was performed on AMC and control PVN to identify additional differentially methylated regions (DMRs). Surprisingly, there was a significant (P>0.042) 2.91% reduction in genome wide CpG methylation in PVN in response to AMC. To identify potential DMRs between the two regions we focused on partially methylated (>70% CpG methylation) domains (PMDs), as they comprise up to 40% of the genome in early life and are enriched in repressed genes. A hidden Markov model (HMM) trained to detect PMDs identified 38 DMRs, covering ~5% of the uniquely aligned rat genome. Selected DMRs were validated by pyrosequencing. RNA-seq analysis on AMC and control PVN yielded a total of 759 differentially expressed transcripts including *Crh*, *Galanin*, *Ube3a* and other transcripts involved in HPA neuronal responses. This combined WGBS and RNA-seq analysis has led to the identification of key hypothalamic gene networks involved in the epigenetic regulation of the attenuated stress response underlying the resilience paradigm. These results provide insight into the early-life epigenetic programming of hypothalamic genes with implications for treatment of human stress-related disorders.

2062W

Mapping methylation and expression quantitative trait loci in human prefrontal cortex. *H. Zhang, F. Wang.* Department of Psychiatry, Yale University School of Medicine, New Haven, CT.

Genetic variants associated with neuropsychiatric disorders are potential expression and/or methylation quantitative trait loci (eQTLs and/or mQTLs) in the brain. They may alter the expression or epigenetic status of genes in specific brain regions such as the prefrontal cortex (PFC; implicated in cognition, perception, memory, and decision making), thus leading to an increased risk of the illness. The present study aimed to map eQTLs and mQTLs in human brain PFC, thus facilitating characterizing the function of disease-associated genetic variants and understanding the role of the PFC in the development of neuropsychiatric disorders. A joint analysis of the relationship among genetic variants, gene expression, and DNA methylation was performed using genotype data of 254,286 common single nucleotide polymorphisms (SNPs) (identified by the Illumina Infinium HumanExome BeadChip 12 v1 array assay), expression data of 16,553 transcripts (or exons) (profiled by the Illumina HumanHT-12 v4 Expression BeadChip assay), and methylation data of 430,407 CpGs (generated by the Illumina Infinium HumanMethylation450 BeadChip assay) from postmortem PFC tissues of 46 Caucasian Australians (32 males and 14 females; half them were affected with alcohol use disorders or AUDs). The correlation of SNP genotypes (in an additive model) and expression levels of exons (or methylation levels of CpGs) was analyzed using the R package Matrix eQTL, with sex, age, AUDs, and postmortem intervals (PMI) as covariates. A total of 1,464 eQTLs (1,245 cis-eQTLs and 219 trans-eQTLs mapped to 549 genes) and 48,041 mQTLs (26,846 cis-mQTLs and 21,195 trans-mQTLs mapped to 6,525 genes) identified in the PFC survived multiple testing correction ($q<0.05$). Moreover, 356 SNPs (mapped to 45 genes) were found to be both eQTLs ($q<0.05$) and mQTLs ($q<0.05$). Among them, 120 SNPs showed an influence on gene expression and DNA methylation in the same direction, whereas the remaining 236 SNPs showed an influence on gene expression and DNA methylation in the opposite direction. DAVID functional annotation analysis indicated that the 45 genes mapped by the 356 SNPs were enriched in pathways such as glutathione metabolism, drug metabolism, and post-translation modification. Additional bioinformatics analysis is need to understand whether these eQTLs and mQTLs identified in the PFC are enriched in neuropsychiatric disorder-associated genetic variants identified by genome-wide association studies (GWAS).

2063T

Methylome-wide association analysis in blood implicates *NLRC5 and TRIM69* **genes for intravenous drug use in HIV-infected patients.** *X. Zhang[1], A. Justice[4], Y. Hu[2], H. Zhao[3], Z. Wang[3], J. Krystal[1], K. Xu[1].* 1) Department of Psychiatry, Yale Medical School, New Haven, CT; 2) Center for Biomedical Bioinformatics, National Institute of Cancer, MD; 3) Department of Biostatistics, Yale School of Public Health, New Haven, CT; 4) Yale University School of Medicine, New Haven Veterans Affairs Connecticut Healthcare System, West Haven, CT.

IMPORTANCE: Intravenous drug (IVD) use is the second most common risk factor for human immunodeficiency virus (HIV-1) infection and IVD use worsens HIV progression. However, the mechanisms of how IVD impacting HIV risk and outcomes is still unknown. **OBJECTIVE:** To identify differential DNA methylation patterns between IVD users and non-IVD users in HIV-infected patients and to identify IVD-associated biological pathways. **MATERIALS AND METHODS:** A total of 553 HIV-infected male veterans were selected from the Veteran Aging Cohort Study. DNA was extracted from peripheral blood. We profiled DNA methylation for 485,521 CpG sites using Illumina HumanMethylation 450K Beadchip. Data processing and normalization was followed by the recent developed pipeline (Lehne et al, 2015). We removed 11,648 probes on X and Y chromosomes and an additional 36,535 probes within 10bp SNPs. Probe intensity was normalized using quantile method. Cell type proportion was estimated with Houseman method using minfi R package. Linear regression model was applied as: *β value (Quantile normalized) ~ IVD + age + race + WBC+ cell types*. Significant level was set at false discovery rate <0.05. Genes with FDR < 0.1 were used for a gene set enrichment analysis to identify IVD-associated pathways by using Ingenuity Pathway Analysis. **RESULTS:** We identified 37 CpG sites reached methylome-wide significance for HIV-infected IVD use. Two CpG sites in the promoter region of the *NLR family, CARD domain containing gene 5 (NLRC5)*, which plays an important role in regulating major histocompatibility complex class I gene expression, were significantly less methylated in IVD users compared to non-IVD users (cg07839457: t=-6.02, p_{adj}=0.0014; cg08159663: t=-4.7, p_{adj}=0.045). In addition, 3 CpG sites in the promoter region of *tripartite motif containing 69 (TRIM69)* were also less methylated in IVD users (cg10439456: t=-5.53, p_{adj}=0.004; cg22107533: t=-5.19, p_{adj}=0.013; cg10439456: t=-5.53, p_{adj}=0.004). *TRIM69* is related with immune system and antigen processing through ubiquitination and proteasome degradation. More interestingly, we found the significant enrichment of genes in the antigen presentation pathway ($p_{nominal}$=4.3x10^{-8}) and the interferon signaling pathway ($p_{nominal}$=1.2x10^{-4}) for HIV-infected IVD use. **CONCLUSIONS:** Our findings suggest IVD use changes DNA methylation in the genes involving immune regulation and antigen processing that are critical for HIV-1 infection and progression.

2064F

Enhancer and super-enhancer gene targets in genetic disease. *C. Barr[1,2], Y. Feng[1,2], A. Dineen[1,2], A. Sarkar[1,2].* 1) Genetics & Development, Krembil Research Institute, UHN, Toronto, ON, Canada; 2) Program in Neurosciences and Mental Health, Hospital for Sick Children, Toronto, ON, Canada.

A major challenge confronting genetic studies of complex traits is moving from GWAS findings to identifying the genes in these regions creating risk and the mechanisms by which genetic variants alter function. For most of the association findings, the associated markers reside in gene regulatory regions, particularly enhancers and super-enhancers. Enhancers can reside anywhere within, or megabases (MB) from the gene that they regulate (target gene). Further, they can enhance transcription of multiple genes (~2.5) and their targets are often not the nearest gene (27-40% target the nearest gene). Thus, the assumption that the gene nearest a GWAS-significant marker will be the risk gene will in many cases be incorrect. To identify the target genes of the enhancers with associated markers for schizophrenia and for immune mediated disorders, we analyzed Capture-HiC data sets. The results identify the target genes of the enhancers, with targets located both inside and outside of the associated region. For example, for the 36 brain super-enhancers within the 108 published schizophrenia GWAS regions, we identified 229 interacting genes, with some of these key suspects located outside the associated region. Specifically, the brain and neural precursor cell (NPC) super-enhancer located within the schizophrenia associated *RERE* gene region interacts with multiple genes over a 2 MB region, including those implicated in neuronal survival (e.g. *LZIC*, *NMNAT1*), genes implicated in Parkinsons (*PARK7*) and bipolar disorder (*PER3*). Using CRISPR/Cas9, we deleted part of the *RERE* super-enhancer in human neural precursor cells (hNPCs) and HEK293 cells. Quantitative reverse transcriptase PCR (RT-PCR) of the targets are in progress and preliminary results indicate that in HEK293 cells that *RERE* and *PER3* were reduced in expression while *PARK7* was unaffected. RT-PCR studies for these genes are in progress for the CRISPR/Cas9 edited hNPCs. In parallel studies, we are using RT-PCR and chromatin conformation capture (3C) in a genotyped panel of human brain tissues to correlate associated alleles with transcription and altered promoter interactions of the predicted target genes. In summary, Capture-HiC analysis provides important new leads in pinpointing the target genes of enhancer-mediated regulation emanating from the GWAS findings. Identification of the multiple gene targets allows for functional studies and has important implications for modeling disease.

2065W

Methylome-wide association study of major depressive disorder in post-mortem brain tissues from 212 subjects. *K. Aberg[1], G. Turecki[2], A. Shabalin[1], L.Y. Xie[1], G. Kumar[1], R.F. Chan[1], M. Zhao[1], M.W. Hattab[1], H. Herman[1], S.L. Clark[1], B. Dean[3], E.J.C.G. van den Oord[1].* 1) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, USA; 2) Department of Psychiatry, McGill University, Canada; 3) The Florey Institute of Neuroscience and Mental Health, The Victorian Brain Bank Network, Australia.

We performed a methylome-wide association study (MWAS) using 212 post-mortem brain tissue samples from Broadmann area (BA) 10, to study the disease etiology of major depressive disorder (MDD) [MIM#608516]. To further study the potential consistency/variation in disease-associated methylation marks across brain regions we investigated a second brain tissue (BA25), from the same individuals, for a subset of 63 subjects. The brain tissues were obtained from the Victorian Brain Bank Network, Douglas–Bell Canada Brain Bank, Harvard Brain Tissue Resource Center, Stanley Medical Research Institute and The Netherlands Brain Bank. The methylomes were assayed with a sequencing-based methylation enrichment approach that applies a high affinity methyl-binding domain protein (MBD-seq). In contrast to standard bisulfite conversion methods, MBD-seq uniquely detects CpG methylation, but not hydroxymethylated CpGs, and, as the approach does not require conversion of nucleic acids, the DNA sequence retains its full integrity thus alleviating the challenges of aligning bisulfite-converted genomes. Furthermore, in contrast to commonly used methylation arrays, typically targeting 2-3% of CpGs in the human genome, MBD-seq allows for the investigation of the majority of all >26 million CpGs. Specifically, in direct comparison with whole-genome bisulfite sequencing (WGBS) MBD-seq can correctly identify the methylation status of 94% of the CpGs detected by WGBS. After rigorous quality control, our analysis involved methylation measures from >17 million CpGs that were evaluated for association with MDD. Our main analysis, including all 212 samples from BA10, suggested many markers with small effects (lambda=1.068; 253 CpGs with p<1e-05). These markers, which all showed the same direction of effect in the main subsets of the samples, are further investigated for their combined poly-epigenetic risk involved in MDD etiology. Furthermore, the comparison of the results from BA10 and BA25 identified a significant overlap (p=0.0082) of MDD-associated sites, suggesting considerable consistency in methylation marks of potential importance in the brain regions. Top findings included CpGs in *ANKS1B* [MIM*607815], a gene with a suggested role in normal brain development, that are currently followed-up for functional relevance in human cells using epigenetic editing. This study presents a rigor investigation of human brain that identified methylation marks of potential importance for MDD etiology.

2066T

DNA methylation profiles of monozygotic twins discordant for ADHD. *T.V.M.M. Costa[1], F.A. Marchi[2], C. Milani[1,3], J.G. Damasceno[1], E.A. Zanardo[1], G.M. Novo-Filho[1], M.M. Montenegro[1], F.A.R. Madia[1], A.T. Dias[1], A.M. Nascimento[1,3], M. Rocha[4], M.B. Veronesi[1], C.A. Kim[3], V. Schuch[5], C.B. Mello[5], M. Muszkat[5], L.D. Kulikowski[1,3,4].* 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo, Brasil; 2) Laboratório NeoGene, AC Camargo Cancer Center, São Paulo, Brasil; 3) Laboratório de Citogenômica, Unidade de Genética, Departamento de Pediatria, Instituto da Criança, HC/FMUSP, São Paulo, Brasil; 4) Centro de Reprodução Humana e Genética, Departamento de Saúde Coletiva, Faculdade de Medicina do ABC, Santo André, Brasil; 5) Núcleo de Atendimento Neuropsicológico Infantil Interdisciplinar (NANI), Departamento de Psicobiologia da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil.

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurobehavioral disorder diagnosed in childhood. It has high heritability (around 80%) and is etiologically heterogeneous. However, the genetic architecture of ADHD is still largely unknown. Behavioral genetic studies have demonstrated that genetic influences play a role in the etiology of ADHD. It is also known that environmental factors are also important in development of the disorder. Therefore, it is necessary to study whether epigenetic factors are related to the etiology of ADHD. Thus, DNA methylation profiles of monozygotic twins are an important tool to study the role of epigenetic factors in ADHD. Diagnosis was performed according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). DNA was extracted from blood lymphocytes of one ADHD patient and his twin brother – without ADHD (two boys, ages 12). After, DNA methylation profiles were performed using Illumina Infinium HumanMethylation450 BeadChip. Arrays data were processed within the environment R. We calculate the difference between the β values (Δβ) of all probes of both twins (with and without ADHD). Limma package was used to obtain significant probes considering Bonferroni correction and p < 0.05. Finally, we selected the probes with Δβ > 0.1. The probes were annotated according to the data provided by Illumina using the genome hg19 reference. In total, 1,023 probes were selected. 631 of these probes are hypomethylated in the ADHD patient relative to his twin without ADHD. Thus, our results suggest that DNA methylation profile of ADHD patient is hypomethylated when compared with his twin brother without ADHD. However, it is important to investigate methylation profiles of more cases of twins discordant for ADHD to verify if indeed there is a relationship between hypomethylated methylation profiles and the development of ADHD. Grants: FAPESP: 2015/05350-5, 2014/02565-8 and FINEP-CT INFRA 0160/12 SP8.

2067F

WGBS reveals autism-associated hypomethylation and differentially-methylated regions in umbilical cord blood samples from the prospective MARBLES study. *C.E. Mordaunt[1,2,3,4], K.W. Dunaway[1,2,3,4], R.J. Schmidt[3,4,5], C.K. Walker[3,4,6], S. Ozonoff[3,7], I. Hertz-Picciotto[3,4,5], J.M. LaSalle[1,2,3,4].* 1) Medical Microbiology and Immunology; 2) Genome Center; 3) MIND Institute; 4) Center for Children's Environmental Health; 5) Public Health Sciences; 6) Obstetrics and Gynecology; 7) Psychiatry and Behavioral Sciences; University of California, Davis; Davis, CA.

   Autism spectrum disorders (ASD) have complex etiologies, likely involving multiple genetic and environmental insults in perinatal life. Genetic susceptibility can interact with environmental risk factors such as pesticides, air pollution, and persistent organic pollutants. The perinatal period is critical for both nutritional protective factors, such as the methyl donor folate, and interactions with genetic regulators of one-carbon metabolism. The epigenetic layer of DNA methylation, at the interface of genetic and environmental risk and protective factors, holds promise for earlier diagnosis and improved understanding of complex ASD etiologies. In this study, we investigated human umbilical cord blood samples from the MARBLES (Markers of Autism Risk in Babies - Learning Early Signs) prospective study by whole-genome bisulfite sequencing (WGBS) to identify DNA methylation differences predictive of ASD diagnosis by age three. ASD cord blood samples showed significantly lower global percent CpG methylation compared to typically-developing (TD) controls (ASD 76.6% vs TD 77.3%, p = 0.01; n = 26 ASD, 26 TD). ASD-associated hypomethylation was observed across most chromosomes, within gene bodies, and within early-life partially-methylated domains (PMDs), defined from human placenta. 20 kb windowing of the genome demonstrated a global shift of hypomethylation, with 74% of windows hypomethylated in ASD samples. This suggests a methylation deficiency in ASD during perinatal life, which could be a cumulative effect of genetic variants, environmental exposures, and/or shortage in methyl donors. Smaller differentially-methylated regions (DMRs) enriched for CpG islands were also identified in ASD cord blood (5,863 total DMRs). 56% of DMRs were hypomethylated, and all were less than 2.5 kb in size. DMRs hypomethylated in ASD cord blood overlapped genes significantly enriched for functions in synaptic transmission (62 genes, Bonferroni-adjusted p = 3.3x10^{-5}) and for the oxytocin signaling pathway (35 genes, Bonferroni-adjusted p = 2.1x10^{-7}). In addition to examining WGBS in association with ASD diagnosis and neurodevelopmental outcome measures, WGBS will be examined in relation to demographic, genetic, environmental, and nutritional information collected in the MARBLES study. These results are expected to improve understanding of perinatal factors in ASD etiology and aid in future preventative and therapeutic treatments.

2068W

Genomewide DNA methylation analysis in an antipsychotic naive first episode of psychosis cohort. *P.N. Moretti[1,2,3], M.L. Santoro[1,2,3], V.K. Ota[1,2,3], S. Jong[4], E.S. Gouvea[5], C.S. Noto[1,3], L.M.N. Spíndola[1,2,3], G. Xavier[2,3], C.M.F. Carvalho[2,3], Q. Cordeiro[5], R.A. Bressan[1,3], A. Gadelha[1,3], G. Breen[4], S.I. Belangero[1,2,3].* 1) Morphology and Genetics, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Psychiatry Department – Universidade Federal de Sao Paulo (UNIFESP), Sao Paulo, Brazil; 3) Interdisciplinary Laboratory of Clinical Neurosciences (LiNC), Psychiatry Department – Universidade Federal de Sao Paulo (UNIFESP), Sao Paulo, Brazil; 4) MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; 5) Psychiatry Department, Irmandade da Santa Casa de Misericórdia de Sao Paulo (ISCMSP), Sao Paulo, Brazil.

   Schizophrenia is a severe psychiatric illness with considerable morbidity and motality as well as a very high cost of care. The first episode of psychosis (FEP) usually occurs in adolescence or early adulthood, and is a critical period for future schizophrenia prognosis. The evaluation of FEP patients can be helpful for understanding the onset and progression of the disorder, especially before treatment with antipsychotics, without the influence of the progression of symptoms and pharmacological treatment. Beyond successful GWAS analyses, the identity of other biomarkers for the disease remains uncertain. Our objective was to analyze genomewide DNA methylation in a longitudinal cohort of antipsychoticnaïve Brazilian FEP patients, followed up after two months (FEP2M) of antipsychotic (Risperidone) treatment. We selected 60 healthy controls, 60 FEP samples before and after treatment, aged 1835 yo and fulfilling DSMIV criteria for psychotic diagnoses . Blood was collected from all participants and DNA was extracted. For DNA methylation analysis, we performed sodium bisulfite treatment, and used the Infinium HumanMethylation450 BeadChip Kit according to the manufacturer's protocol. Ten differently methylation regions (DMR) were observed between FEP and FEP2Min in a paired analysis. Notably, three of the ten regions were identified before in PGC2 Schizophrenia GWAS (regions on chromosome 6, 2 and 17). We did not observe any DMRs between healthy controls and antipsychotic-naïve FEP subjects, suggesting the observed changes were a consequence of treatment.

2069T

Differential DNA methylation in the dorsal and ventral striatum of individuals with chronic cocaine dependence. *K. Vaillancourt[1,2], C. Ernst[1,3], G.G. Chen[1], A. Bramoulle[1], J-F. Théroux[1], D.C. Mash[4], G. Turecki[1,2,3].* 1) McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, QC, Canada; 2) Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA.

   Background: Cocaine dependence is a chronic relapsing disorder whose development and trajectory is impacted by multiple genetics and environmental factors. Transcriptional changes accompany the transition from recreational cocaine use to cocaine dependence and epigenetic mechanisms may mediate these effects. Recently, several studies have identified epigenetic marks that are associated with the acquisition of compulsive drug seeking in animal models, but little is known about the role of epigenetics in human cocaine dependence. Of particular interest is DNA methylation as it represents a mitotically stable mark that has been shown to be altered by environmental experience. **Methods**: We used Reduced Representation Bisulfite Sequencing (RRBS) on post mortem nucleus accumbens and caudate tissue from 25 dependent cocaine users and 25 drug-free and age-matched controls. This approach has allowed us to detect genome-wide cytosine methylation at base pair resolution. Differentially methylated regions (DMRs) were validated in this cohort, and replicated in an independent cohort of cases and controls (n=18 per group), using targeted deep bisulfite sequencing. We used fluorescence activated cell sorting (FACS) to separate neuronal (NeuN+) from non-neuronal (NeuN-) nuclei in order to identify cell type specificity of our results. **Results**: Our study has identified multiple clusters of hyper- and hypomethylation associated with chronic cocaine dependence, in both brain regions. Differential methylation spans gene bodies, including introns and exons, as well as promoter and intergenic regions. High throughput discovery of cocaine-associated networks and pathways allow us to investigate cell-type specific epigenetic changes that accompany the development and persistence of chronic cocaine dependence. Funded by NIDA (DA033684).

2070F

miR-1202 potentially important in regulating dopaminergic neuron differentiation. *D. Cao, L. Li, W. Law, J. Tu, K. Miu, H. Cheung, W. Chan.* School of Biomedical Sciences, Faculty of Medicine, the Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.

   Micro-RNAs (miRNAs) comprise a distinct class of small non-coding RNAs identified to be potentially important post-transcriptional gene regulators. Their activities in cells have important consequences for normal development and physiology, diseases, and evolution. In the past decades, many miRNAs have been shown to be dynamically expressed in brain or during neural differentiation from stem cells. However, it remains largely unknown about how specific miRNAs regulate neurogenesis. Through analyzing miRNA expression profile at different time points during neural induction from hiPSCs, hsa-miR-1202 was identified to be significantly increased during the process. hsa-miR-1202 is a primate-specific microRNA. Interestingly, hsa-miR-1202 was recently shown to be dysregulated in brain tissues of patients suffering from major depression syndromes suggesting that hsa-miR-1202 may have important functions in human brain. Downstream target prediction and in silico functional analysis showed that hsa-miR-1202 is probably involved in regulating development and activities of dopamine neurons. To study how hsa-miR-1202 regulates dopaminergic (DA) neuronal development, we differentiated human iPSCs into DA neurons. We showed that expression of has-miR-1202 was elevated in the DA neuron progenitors and DA neurons. We are in the process of manipulating the expression of hsa-miR-1202 during DA neuronal patterning stage in order to define its role in DA neuronal differentiation. This study will extend our understanding of the specific function of hsa-miR-1202 in DA neuronal development and reveal specific role of hsa-miR-1202 in DA neuron-related diseases. .

2071W

Small nucleolar RNA and antidepressant response. *R. Lin[1], J.P. Lopez[1], L. Fiori[1], C. Cruceanu[1], R. Belzeaux[1], CANBIND Working Group[3], J. Foster[2], S. Kennedy[2], G. Turecki[1].* 1) Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, University Health Network, University of Toronto, Toronto, Ontario, Canada; 3) CANBIND.

Statement of Purpose: Major depressive disorder (MDD) is a prevalent disorder treated primarily by antidepressants. Although effective, on average 30%-40% of patients experience an inadequate response to treatment after several attempts. Thus, there is a great need to identify biomarkers associated with MDD that predict and/or mediate response to antidepressant treatment. Recent discoveries have pointed towards small non-coding RNAs (sncRNAs) as feasible biomarkers. While the majority of studies have focused on microRNAs, evidence suggests that small nucleolar RNAs (snoRNAs), which are involved in alternative splicing and chemical modifications of RNAs, may also act as novel biomarkers for antidepressant response. **Methods:** A sample of 258 depressed patients were treated either with the antidepressant duloxetine (N=124) or placebo (N=134) for a period of 8 weeks. Blood samples were collected at baseline (T0) and 8 weeks into treatment (T8), and RNA sequencing was performed to detect expression changes of sncRNAs. At T8, patients were grouped into responders or non-responders to duloxetine, where responders were characterized as having a 50% reduction in the MADRS score from T0 to T8. The expression of the top snoRNAs, from responders of duloxetine, showing significant differences in their expression profiles from T0 to T8 was assessed in neural progenitor cells (NPCs) treated with duloxetine using qRT-PCR. **Results:** Small RNA sequencing revealed SNORD43, SNORD11B, SNORD17, SNORD77, SNORD99, and SNORD52 to be significantly up-regulated in responders 8 weeks after duloxetine treatment. NPCs treated with duloxetine for 2 weeks also showed a significant up-regulation of all 6 candidate snoRNAs when compared to untreated controls. This is the first study profiling snoRNAs in MDD and antidepressant response, and these preliminary results suggest that the candidate snoRNAs may be good biomarkers of antidepressant response.

2072T

Differential DNA methylation of neuro-immune regulatory genes associated with symptom severity and amygdala-prefrontal dysfunction in pediatric post-traumatic stress disorder. *T.J. Keding[1,2], L.A. Papale[2], R.S. Alisch[1,2], R.J. Herringa[1,2].* 1) Neuroscience Training Program, University of Wisconsin-Madison, Madison, WI; 2) Department of Psychiatry, University of Wisconsin-Madison School of Medicine & Public Health, Madison, WI.

Studies of adult post-traumatic stress disorder (PTSD) suggest a role for DNA methylation on genes related to hypothalamic-pituitary-adrenal (HPA) axis regulation, including the glucocorticoid receptor gene *NR3C1* and its regulator *FKBP5*, as well as genes involved in the regulation of the immune system. Furthermore, DNA methylation levels on such genes are dynamically related to symptom severity and improvement. However, studies have yet to investigate whether these relationships exist in pediatric PTSD, especially with regards to symptom severity, trauma-exposure, and abnormalities in brain circuitry. Saliva from youth with trauma exposure and PTSD (n = 22, 8-18 years) and age- and sex-matched healthy comparison youth (n = 22) was collected. DNA was extracted and genome-wide 5-methylcytosine (5mC) was profiled using the HumanMethylation450 BeadChip. Differentially methylated regions (DMRs) between groups were examined using a mixed effect model, controlling for age, sex, and cell type composition. False discovery rate (FDR)-corrected multivariate and partial correlation analyses were conducted to relate methylation at DMRs to symptom severity, trauma-related measures, gray matter volume, and task-based brain activation and amygdala functional connectivity within the PTSD group. Eighty-one group-associated DMRs were identified, 79 of which showed hypermethylation within the PTSD group relative to healthy controls. Enrichment analyses revealed that 14 of these genes were associated with signal transduction regulation and 11 with immune response regulation. Additionally, a subset of these genes (N = 28) showed significant associations with symptom severity, trauma-exposure, and amygdala-prefrontal structure and function within the PTSD group (FDR-corrected p < 0.05). Eleven of these genes were related to cellular and organismal development processes, including *GABRA3*, *GABRA5*, and *SYNGAP1*, which showed specificity to the nervous system. Systems pathways analysis further revealed that these three genes are heavily involved in the development and migration of neurons, neuronal apoptosis, as well as learning, memory, and cognition. Together, these findings link DNA methylation to symptom severity and trauma-related brain measures in pediatric PTSD. The DMRs found here represent potentially modifiable molecular substrates in the brain that ultimately could be targeted to aid children exposed to early life trauma that otherwise may result in PTSD.

2073F

Exploring the etiology of FASD: Hippocampal gene expression after continuous prenatal alcohol exposure plus postnatal maternal separation in mice. *B. Alberry, S.M. Singh.* Biology, Western University, London, Ontario, Canada.

Prenatal alcohol exposure (PAE) can result in Fetal Alcohol Spectrum Disorder (FASD), characterized by behavior abnormalities. In earlier experiments using a mouse model, we established the effect of alcohol on long-term neural gene expression, which may come about through additional epigenetic modifications resulting in altered gene expression. This research extends the model to include early life stress (ELS) via postnatal maternal separation. To generate PAE mice, pregnant C57BL/6 females had access to water and 10% ethanol in water, while control mice had access to water only. ELS treatment followed with half the pups from each litter isolated from the dam and littermates postnatal days 2-14 for 3 h/day. All pups underwent various behavioral tests during development and into adulthood, including the Barnes Maze Test for learning and memory. Mice were sacrificed on postnatal day 70. Hippocampal RNA was isolated for assessment of gene expression including microRNAs using two independent methods: Affymetrix Mouse Gene 2.0 ST gene expression arrays and RNA-Seq. Genes implicated by both measures were compared for pathway and network analysis. Results show many genes have altered expression in the adult hippocampus following PAE, with and without ELS. The results are compatible with observed behavioral alterations resulting from PAE and ELS. A number of pathways, including alcoholism, common in FASD are implicated. Results argue that the effect of PAE and ELS in FASD is realized by alterations in gene expression in a large number of genes affecting numerous pathways. This experimental model represents a more realistic model of FASD and critical genes and pathways identified may provide a foundation for prospective interventions following early FASD diagnosis.

2074W

Long noncoding RNAs regulation of glucocorticoid signaling in brain and blood tissue across psychiatric disorders. *G. Guffanti[1], A. Wingo[2], T. Jovanovic[2], C. Nemeroff[3], A. Myers[3], F. Macciardi[4], K. Ressler[1].* 1) Harvard Medical School, Belmont, MA; 2) Emory University, Atlanta, GA; 3) 3Univ. of Miami, Miami, FL; 4) University of California Irvine, Irvine, CA.

BACKGROUND: Evidence is accumulating that non-coding RNAs might play arole in GR-mediated regulation of glucocorticoid signaling. lncRNA GAS5 has recently been characterized as a suppressor of transcriptional activity of glucocorticoid responsive genes in mice. Human genetic studies examining stress pathways have mostly focused on the glucocorticoid receptor (GR) and its target genes. Transcriptional and post-transcriptional regulatory function of lncRNAs might explain the missing "link" between the genetic make-up predisposing to PTSD and stress- and trauma-related disorders and GR-related neurobiology in a tissue specific way. We set up to test the hypothesis that lncRNA GAS5 regulation might be perturbed in PTSD. METHODS: The sample derives from the larger cohort of the Emergency Department (ED) collaboration for the analysis of the transcriptome of peripheral blood of PTSD patients. PTSD was diagnosed immediately after the trauma and confirmed up to three months later, yielding 23 cases and 23 trauma-exposed controls. Raw paired-end RNA-sequencing was aligned to the hg19 genome build using the Tuxedo pipeline. Further processing to delineate the transcriptional profiles of lncRNA was performed using the ad hoc lncRNA pipeline called PLAR. We tested for statistical significance of observed changes in gene expression between PTSD patients and trauma-exposed controls using the statistical negative binomial model implemented in R-Bioconductor EdgeR. RESULTS: We have characterized GAS5 expression profiles in a subset of 46 subjects among cases and controls from the GTP peripheral RNA-seq cohort. Out of the 29 non-redundant isoforms of GAS5, we were able to detect the expression of 12 isoforms. At least two GAS5 isoforms revealed significant down-regulation in PTSD patients compared to trauma-exposed controls (p < 10-4). Based on the information reported in the gene expression database GTEx, these isoforms seem to be expressed in peripheral blood tissue but do not seem to be expressed in AM nor HC in favor of other isoforms. CONCLUSIONS: The results highlight the importance of tissue specificity in the assessment of lncRNA expression profiles, and support the original hypothesis that lncRNA GAS5 plays a role in PTSD, and suggest that lncRNA GAS5 may be differentially expressed in stress-related syndromes compared to traumatized controls. The mechanism by which GAS5 performs its regulatory function remains to be elucidated. .

2075T

Disruption of activity-dependent signaling pathways in Potocki Shaffer Syndrome. *R.S. Porter[1,2], Y. Nakamura[1], S. Agarwal[1], H.G. Kim[3], S. Iwase[1].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 3) Section of Reproductive Endocrinology, Infertility & Genetics, Department of Obstetrics & Gynecology, Augusta University, Augusta, GA.

Histone modifications and histone variants are important epigenetic factors that can lead to differential higher-order structures of the genome. A variety of protein factors are responsible for placing, reading, and removing the various histone marks. Many of these factors when mutated lead to neurodevelopmental disorders such as intellectual disability and autism, suggesting that proper epigenetic regulation is crucial for normal brain development. Potocki Shaffer Syndrome (PSS) is a contiguous gene deletion syndrome involving multiple genes on chromosome 11p11.2. PSS is characterized by intellectual disability, craniofacial abnormalities, and bone malformation. Intellectual disability in PSS has been attributed to loss of *PHF21A*, a protein that recognizes unmethylated histone 3 lysine 4 (H3K4me0) and is part of the Lysine Specific Demethylase 1 (*LSD1*) complex. *LSD1* acts to demethylate mono- or di-methylated histone 3 lysine 4(H3K4me1/2) to repress nearby genes. The *LSD1-CoREST* complex is important for mediating repression of neuron-specific genes in non-neuronal cells. However, the mechanisms by which *PHF21A* contributes to brain development remain unknown. Our lab recently performed RNA-Sequencing on two patient samples with PSS and two unaffected control individuals. Preliminary pathway analysis using LR Path and GSEA MSigDB of differentially regulated genes suggested disruption of learning and memory pathways, in line with the clinical phenotype. Specifically, *CREB* target genes were differentially regulated more than expected by chance. Transcription factor target pathway analysis showed misregulation of neuronal activity dependent signaling pathways, specifically *CREB*, *AP1*, and *MEF2*. With the hypothesis that *PHF21A* is required for proper induction of activity-dependent signaling in the nervous system, these disrupted pathways will be functionally tested in binding assays with *PHF21A* as well as luciferase assays to examine whether the expression of these transcription factor targets changes upon *PHF21A* loss. This work will provide insight into the molecular mechanisms of intellectual disability in PSS and can be the basis for the design of rational therapeutics.

2076F

A unique epigenetic fusion protein on a ring X chromosome found in a patient with Kabuki syndrome. *G.O. Pilarowski[1,2], J.D. Robertson[1,2], C. Applegate[1], E.S. Wohler[3], D.A. Batista[1,3], L.A. Goff[1,4], H.T. Bjornsson[1,5].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States; 2) Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States; 3) Cytogenetics and Microarray Laboratory, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States; 4) Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States; 5) Department of Pediatrics at the Johns Hopkins University School of Medicine, Baltimore, Maryland, United States.

Kabuki syndrome (MIM: #147920; #300867) is a genetic disorder characterized by unique facial features, growth retardation, and intellectual disability. The majority of patients (70%) have a mutation in either of two genes encoding histone modifying enzymes (KMT2D and KDM6A), resulting in an inability to open chromatin. A subgroup of Kabuki syndrome patients with unknown genetic etiology have a Turner syndrome karyotype (45,X) and most of these are mosaic for a ring X chromosome (45,X/46,X,r(X)). This ring X does not include the *XIST* locus, indicating that the ring is constitutively active. We studied the breakpoints in one such patient with a clinical diagnosis of Kabuki syndrome and a cytogenetic diagnosis of Turner syndrome (46,X, t(1:9)(q21;q32) ,+r.ish r(X)(DXZ1+,XIST-) [16]/45,X, t(1:9)(q21;q32)[4]), no mutations in *KMT2D/KDM6A* and a non-consequential translocation inherited from a healthy father. One possibility is that overexpression of one of the 107 genes on the ring would cause the Kabuki phenotype; yet none of these genes involve histone machinery. Surprisingly, we found that the fused ends of the ring X chromosome form a novel fusion transcript of two histone modifiers, created by the 5' end of *PHF8* and the out of frame 3' end of *HDAC8*. PHF8 normally functions to remove H3K9 mono- and di-methylation, showing enhanced activity when bound to H3K4 tri-methylation. Here we demonstrate that the fusion protein is produced in cells, it localizes to the nucleus, and contains an intact PHD finger domain. Therefore, we hypothesize that the fusion protein acts in a dominant negative manner by binding to H3K4me3 with its PHD finger domain. Such binding could mask the H3K4me3 mark from the full length PHF8, thus rendering it unable to remove the closed chromatin mark H3K9me1/2. We are employing single cell RNA-sequencing on both populations of cells found in the patient (45,X cells and 46,X,r(X) cells). Through single cell analysis we will explore whether the presence of this ring X leads to transcriptional dysregulation and subsequently we will interrogate the effects of the fusion protein itself. Although the paradigm of novel fusion proteins leading to disease has been established in cancer, to our knowledge this may be the first example of a Mendelian disease caused by the creation of a novel fusion protein. Our findings also suggest that disruption of sequential stages of chromatin opening forms a unifying mechanism of disease in Kabuki syndrome.

2077W

Genome-wide DNA methylation testing in patients with developmental delay and intellectual disabilities. *B. Sadikovic[1,2], K. Boycott[3], C. Schwartz[4], G. Pare[5], C. Howlett[1,2], P. Ainsworth[1,2], L. Schenkel[1]. 1) Department of Pathology and Laboratory Medicine; Western University; London, ON, Canada; 2) Molecular Genetics Laboratory; Molecular Diagnostics Division, London Health Sciences Centre, London, ON, Canada; 3) Children's Hospital of Eastern Ontario Research Institute; University of Ottawa, Ottawa, ON, Canada; 4) Center for Molecular Studies, J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, South CA, USA; 5) Department of Pathology and Molecular Medicine and Department of Clinical Epidemiology and Biostatistics; McMaster University; Hamilton, ON, Canada.*

Developmental Delay and Intellectual Disabilities (DD/ID) are caused by the interaction of genetic and environmental factors that is facilitated by epigenetic mechanisms. The most widely studied epigenetic mechanism is the DNA methylation, which is a covalent modification occurring primarily at cytosines within CpG dinucleotides. DNA methylation defects have been shown to play a key role in many pediatric and adult onset disorders, including imprinting diseases and carcinogenesis. However, there is little known about genome-wide DNA methylation changes in patients with DD/ID. Our laboratory has performed a genome-wide DNA methylation testing in 1000 patients with a wide range of disorders associated with DD/ID to clinically validate the use of a genome-wide DNA methylation assay and define epi-signatures of various DD/ID conditions. Briefly, the DNA methylation array was performed in peripheral blood using the Infinium HumanMethylation450 Beadship. And results were analyzed using a custom algorithm with Partek Genomic Suite software. Methylation patterns of individual patients were compared to normal reference cohort of >300 individuals and were prioritized based on the statistical parameters including methylation difference, *p* value, and F value; and functional parameters including distance to the CpG islands and gene promoter, and the proximity to the known disease-causing genes. We have validated this approach for sensitive detection of imprinting disorders including Angelman, Prader-Willi, Beckwith Wiedemann, and Russell-Silver syndrome, as well as Fragile X syndrome. We also discovered novel highly specific diagnostic epigenetic signatures in the peripheral blood of patients with Alpha Thalassemia/Mental Retardation syndrome X-Linked, Floating-Harbor syndrome, Autosomal Dominant Cerebellar Ataxia with Deafness and Narcolepsy (ADCA-DN) and Claes-Jensen X-linked Mental Retardation syndrome. These findings demonstrate clinical utility of genome-wide DNA methylation testing in patients with a wide-ranging spectrum of epi/genetic conditions, and provide better understanding for the pathology of DD/ID syndromes, in which specific DNA methylation changes could lead directly to an aberrant expression of genes. .

2078T

Genome-wide DNA methylation analysis of atypical 7q11.23 rearrangements implicate *GTF2IRD1* **and** *GTF2I* **in DNA methylation.** *E. Strong[1], C.B Mervis[2], C.A Morris[3], L.R Osborne[1,4]. 1) Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY; 3) Department of Pediatrics, University of Nevada School of Medicine, Las Vegas, NV; 4) Department of Medicine, University of Toronto, Toronto, Ontario, Canada.*

Williams syndrome (WS) is a developmental disorder caused by the hemizygous deletion of approximately 25 protein-coding genes on chromosome 7q11.23. The reciprocal duplication leads to a distinct disorder, 7q11.23 duplication syndrome (Dup7). Both syndromes present with variable cognitive and behavioural phenotypes, including intellectual ability that is more limited than other family members', anxiety disorders and attention deficit hyperactivity disorder, however little is known about their underlying molecular mechanisms. Five genes within 7q11.23 have been associated with epigenetic mechanisms and we have recently shown that deletion or duplication of this chromosome segment causes dose-dependent changes in DNA methylation. To better understand the molecular mechanism of aberrant DNA methylation in these disorders and identify the gene(s) responsible we assessed genome-wide DNA methylation in participants with atypical deletions or duplications of 7q11.23. We measured genome-wide DNA methylation in whole blood samples derived from 3 participants with atypical deletions and 3 with atypical duplications of 7q11.23 using the Illumina HumanMethylation450k array. Atypical rearrangements had differing contributions of the genes previously linked with epigenetic regulation (*BAZ1B, BCL7B, WBSCR22, GTF2IRD1, GTF2I*). By comparing DNA methylation profiles across the top 1000 most variable positions in 18 typically developing controls (TD), 6 atypical, 20 WS and 10 Dup7 participants, we found that only the atypical rearrangements that affect the *GTF2IRD1* and *GTF2I* genes clustered with their respective syndromic cohorts. Likewise, hierarchical clustering of the top 500 most differentially methylated (DM) sites within the WS cohort and top 500 most DM sites within the Dup7 cohort revealed that only atypical rearrangements spanning *GTF2IRD1* and *GTF2I* cluster with their respective syndromic cohorts. Pyrosequencing validation of several sites confirmed these methylation changes and showed close correlation with array findings. These data suggest that loss or gain of the GTF2I genes at the telomeric end of the 7q11.23 interval are key contributors to the large changes in DNA methylation we have reported in our WS and Dup7 cohorts. Current work focuses on validating and investigating the individual contribution of *Gtf2i* and *Gtf2ird1* to aberrant methylation using mouse models to better understand the molecular mechanism of aberrant DNA methylation in WS and Dup7.

2079F

SMCHD1 regulates a limited set of gene clusters on autosomal chromosomes. *A.G. Mason[1], R.C. Slieker[2], J. Balog[1], R.J.L.F Lemmers[1], C.J. Wong[5], J.W. Lim[5], G.N. Filippova[5], B.T. Heijmans[2,3], R. Tawil[4], S.J. Tapscott[5], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 3) Netherlands Consortium for Healthy Aging, PO Box 9600, Leiden 2300, RC, The Netherlands; 4) Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA; 5) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Facioscapulohumeral muscular dystrophy (FSHD) affects approximately 1 in 20,000 to 1 in 8,500 individuals. It is clinically characterized by the progressive and often asymmetrical weakness and wasting of the facial and upper extremity muscles. The most common form of FSHD, FHSD1, is caused by the deletion of a subset of D4Z4 macrosatellite repeats on chromosome 4, whereas the less common form, FSHD2, is most often caused by mutations in the *SMCHD1* gene. Both mutations result in the incomplete epigenetic repression of the D4Z4 encoded retrogene *DUX4 in somatic cells*, leading to the variegated expression of DUX4 in skeletal muscle cells. *SMCHD1* encodes for a protein that regulates chromatin repression at different loci, having a role in methylation establishment and/or maintenance. To investigate the global effects of SMCHD1 mutations on methylation we carried out 450k methylation analysis on mononuclear monocytes obtained from 23 female *SMCHD1* mutation carriers and 24 female unaffected controls. In addition, we performed reduced representation bisulfite sequencing (RRBS) on two FSHD2 and unaffected control myoblast cell lines. From the assays we identified several clustered genomic loci with decreased methylation: the clustered protocadherin (*PCDH*) locus on chromosome 5, the tRNA and 5S rRNA clusters on chromosome 1, the *HOXB* and *HOXD* clusters on chromosomes 17 and 2 respectively, and the D4Z4 macrosattelite repeats on chromosomes 4 and 10. In addition an increase in RNA expression was seen in FSHD2 myoblast cells for some of the *PCDHβ* cluster isoforms, tRNA isoforms, *HOXB* isoforms, and 5S rRNA in comparison to controls, as well as the increased DUX4 expression from the D4Z4 repeats previously described. We identified SMCHD1 binding at the DNAseI hypersensitivity sites HS17-17' and HS19-20, known to regulate the expression of the *PCDHβ* cluster, and within the chr1 tRNA cluster, with SMCHD1 binding being significantly decreased at HS17-17' and HS19-20 in *SMCHD1* mutation carriers compared to controls. These studies suggest that in a myogenic context SMCHD1 acts as a repressor on the *PCDH* cluster and some additional autosomal gene clusters, with *SMCHD1* mutations resulting in hypomethylation of the cluster, increased expression of some gene cluster isoforms, and in the case of the *PCDH* cluster decreased binding of SMCHD1 to cluster regulatory sites.

2080W

Expression of selected circular RNAs in myotonic dystrophy type 1. *K. Czubak, M. Wojciechowska, P. Kozlowski.* European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Recent studies have discovered a new class of RNA molecules called circular RNAs (circRNAs), whose generation depends on a non-canonical form of alternative splicing. Unlike the other types of RNA, circRNAs are covalently closed circular molecules. Most of them contain exons of protein-coding genes and are processed cotranscriptionally, which implies that their formation competes with the formation of linear transcripts. To date, several mechanisms of circRNAs biogenesis have been proposed. One of them involves RNA-binding protein, muscleblind (MBNL). The MBNL protein is a splicing factor acting as an activator or repressor of splicing. The downregulation of this protein, caused by its sequestration by expanded CUG tract in *DMPK* transcript, plays an important role in the pathogenesis of myotonic dystrophy type 1 (DM1). Taking into account the sequestration of MBNL in DM1 cells and the crucial role of this protein in circRNA biogenesis, one may expect decreased level of circRNAs in DM1. To investigate this issue, from circRNA databases and from the literature we selected several circRNAs, taking into account their expression, ratio to linear transcript, and the presence of MBNL binding sites in surrounding introns. For 7 of the selected circRNAs, including circHIPK3, circN-FATC3 and circMBNL, we designed and optimized splicing-specific droplet digital PCR (ddPCR) assays. The specificity of all designed circRNA-specific amplicons was confirmed by the Sanger sequencing method. Each designed assay allows both absolute quantification of circRNA copies and estimation of its ratio to the linear counterpart. With the use of the designed assays, we analyzed the expression of the circRNAs in samples derived from muscle biopsies of DM1 patients and from DM1 myoblast cell lines, and compared it with the expression in the corresponding normal samples. As expected, our preliminary results showed decreased expression of some of the analyzed circRNAs in DM1. However, since these results are limited to highly-expressed circRNAs, in the next step we will extend the analysis to a higher range of different types of circRNAs (high/low expression, single-exonic/multiexonic, exonic/intronic). In conclusion, our analysis showed the downregulation of the selected circRNAs in DM1 samples and confirmed the suitability of ddPCR technique for the parallel analysis of the expression of circular and linear RNA. Acknowledgements: NCN 2015/17/B/NZ2/01182.

2081T

Variation in mitochondrial DNA copy number is associated with epigenetic modification of the nuclear genome. *C.A. Castellani[1], S.E. Ellis[1], F.N. Ashar[1], R.J. Longchamps[1], A. Tin[2], J.A. Lane[3], J. Bressler[4], J. Coresh[2], J.S. Pankow[5], M. Fornage[4], N. Pankratz[3], E. Boerwinkle[4], D.E. Arking[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Epidemiology and the Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, MN; 4) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 5) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN.

Reduced mitochondrial DNA copy number (mtDNA-CN) has been associated with human disease and overall mortality. Changes in mtDNA-CN have been proposed to have a role in regulating epigenetic changes as evidenced by the observation of both hyper- and hypo- methylation of nuclear genes in cell lines depleted for mitochondria. Thus, increased understanding of the cross-talk between nuclear and mitochondrial DNA is critical to elucidating the impact of mtDNA-CN on human health and disease. To measure nuclear DNA methylation, DNA was extracted from buffy coat for 1,567 African American participants from the Atherosclerosis Risk in Communities (ARIC) study and hybridized to the Illumina Infinium 450K microarray. We used the Genvisis software to determine mtDNA-CN from genotyping microarray raw probe intensities (Affymetrix Human SNP 6.0). Linear mixed effect regression analyses were performed to determine the association between nuclear DNA methylation and mtDNA-CN, adjusted for age, sex, collection site, visit, chip position, plate, smoking status, imputed cell composition and surrogate variables. Residual bootstrapping was performed to determine the probe-specific genome-wide significance cutoff (p<6.22e-08). Twenty-five independent loci were genome-wide significant. The top associations include two CpG probes found in the 5'UTR of *KIAA0825* on chromosome 5 (p=1.12e-43; p=9.25e-30). Interestingly, we also observe that increased mtDNA-CN is associated with global hypermethylation (p<2.2e-16, ß=0.149). The results establish that variation in mtDNA-CN is associated with nuclear epigenetic modifications.

2082F

Effects of World Trade Center (9-11) particulate matter on mitochondrial damage observed in human lung cell in vitro. *A. Dilorenzo, V. Gatson, M. Hassan, C. Gonzales, R. Madrid, L. Seder, E. Tobias.* Dept Biol, Sci Hall - 109, Montclair State Univ, Upper Montclair, NJ.

An array of chronic health conditions have been associated exposure to the dust generated from the New York City World Trade Center (WTC) attack and collapse on September 11, 2001. This study seeks to isolate cellular mechanisms thought to play a role in the decreased lung function of World Trade Center (WTC) workers and victims. Human Lung Fibroblast Cells were exposed to World Trade Center (WTC) Dust at various ppm concentrations and simulated stress environments (via decreased serum levels). Previous studies have shown that the exposed cells showed decreased levels of cell proliferation as WTC Dust ppm in test media increased. The particle concentrations of WTC Dust assessed for each serum concentration ranged between 25 and 250 ppm. The serum concentrations used included 10% Fetal Bovine Serum (FBS) as the non-stressed level, followed by levels simulating stress environments. Exposure levels as low as 25 ppm WTC have shown increased levels of apoptosis. Experiments are in progress to determine if the components of the WTC dust specifically damage mitochondria which can be shown by reduced ATP output using the Promega Mitochondrial Tox-Glo Assay. Another marker of mitochondrial damage is the formation Reactive Oxidative Species (ROS). Promega ROS- Glo Assay will be employed to determine if the toxic, apoptotic and mutagenic effects of WTC dust can be shown as the result of oxidative stress. If WTC Dust is influencing free radical formation, then increases in ROS can be measured by hydrogen peroxide as a marker of cellular damage.

2083W

microRNA-328 suppresses scleral stem cells leading to an increase risk for myopia. *S.H. Juo[1], E. Hsi[1], C.L. Liang[2].* 1) Kaohsiung Medical University, Kaohsiung, Taiwan; 2) Asian University, TaiChung, Taiwan.

Myopia has become a pandemic disease, however, the underlying mechanism for myopia development is poorly understood. During the development of myopia, the sclera has dynamic changes and remodeling leading to an elongation of eyeball. We previously reported that over-expression of microRNA-328 can increase a risk for myopia. Recent studies identified the existence of rodent scleral stem cells (SSCs) that highly express CD44. Since microRNA-328 can directly target CD44, we explored the relationship among microRNA-328, SSCs and myopia development in teh mouse model. Myopia is induced by the following method: the right eye was covered to induce form deprivation myopia (FDM) in a mouse of 23 days old, while the fellow eye (i.e. left eye) was not covered. The normal control mice were not treated in either eyes. Since the myopic eyes have the longer axial length (AXL), the difference of AXL of the same animal is a surrogate of myopia. SSCs were isolated from the right eyes from either experimental or control mice for the measurement of SSC number and ability of differentiation to various cell types. The correlation between stem cell abundance and myopia severity was tested. In addition, microRNA-328's effects on CD44 expression and stem cell functions were assessed. Using fluorescent staining, we showed that most SSCs reside in the ciliary body. The induction of myopia was failed in some mice, and therefore we had 3 types of eyes (myopic, normal and "fail" to be induced to myopic eyes). The SSC percentage in normal eyes, myopic eyes and the "fail" eyes was 3.02%, 1.95% and 2.39%, respectively. The difference of SSCs between myopic and normal eyes was statistically significant (p= 0.015). In addition, the correlation between the difference of AXL and stem cell abundance is -0.7 (p=0.0024) in the experimental mice, which means more SSCs less myopic. These results indicate sufficient SSCs are important to prevent myopia development. Since microRNA-328 can target CD44 mRNA, a reduction of microRNA-328 level by its anti-sense could increase the SSC proliferation rate (p<0.0001), elevate ALP expression (p=0.03) and ALP activities (p=0.0003). In addition, a decrease of microRNA-328 could substantially enhance the stem cell differentiation to osteocytes, chondrocytes and adipocytes. In conclusion, we demonstrated the relationship among microRNA-328, SSCs and myopia development, which implies a potential application of stem cell treatment in myopia.

2084T

Functional role of an intronic SNP rs2279590 in *CLU* gene and its regulators in the progression of pseudoexfoliation syndrome. *B. Padhy[1], B. Hayat[1], G.G. Nanda[1], R. Mendke[2], P. Mohanty[2], D.P. Alone[1].* 1) School of Biological Sciences, National Institute of Science Education and Research, Khurda, Odisha, India; 2) JPM Eye hospital and Research institute, Cuttack, Odisha, India.

Purpose: An intronic SNP in clusterin gene rs2279590 is a risk factor for age related disorders like Alzheimer's, Diabetes and pseudoexfoliation. Following work depicts the functional significance of rs2279590 and role of its regulators in progression of pseudoexfoliation. Methods: Anterior eye tissues of study subjects (Control, pseudoexfoliation syndrome-PEXS and pseudoexfoliation glaucoma-PEXG) were collected in accordance with declaration of Helsinki and target gene expression was checked by quantitative real-time PCR (qRT-PCR). Electromobility shift assays (EMSA) were carried out using Light Shift chemiluminescent kit. Chromatin immunoprecipitation (ChIP) assays were done by Pierce-Agarose ChIP kit and Dual-Luciferase (R) reporter assay (DLR Promega) system was used for reporter assays. Results: Clusterin mRNA was two-fold upregulated in individuals with "GG" homozygosity at rs2279590 compared to "AA". TFSEARCH revealed a putative binding element for Heat shock factors (HSFs) in the flanking region of rs2279590. EMSA confirmed a defined binding of HSFs to the flanking region of rs2279590 with wild allele "A" but not with allele "G". Binding was further confirmed by *in vivo* ChIP assays with specific antibodies for HSFs for immunoprecipitation. Luciferase reporter vector containing an insert (18bp) identical to the surrounding region of rs2279590 with allele "A" showed reduced expression compared to that of allele "G" whereas insertion of a larger fragment (200bp) surrounding rs2279590 drastically reduced reporter activities suggesting a nearby region cumulatively acting as a negative regulator. EMSA also validated the presence of another putative binding site for an unknown regulator located at a nearby 3' region of rs2279590. qRT-PCR revealed a significant upregulation of heat shock factor-1 in anterior eye tissues (lens capsule and conjunctiva) of PEXS but not in PEXG individuals whereas heat shock protein-70 a major cytosolic chaperone is found to be significantly downregulated in PEXS individuals. Conclusions: HSFs negatively regulate clusterin expression by binding to allele "A" at rs2279590. They do so together with another unknown regulator located 3' to rs2279590. "G" being a risk allele for age related neurodegenerative disorders, Alzheimer's and PEXG, support a neurotoxic role of increased Clusterin despite its chaperonic functions. Furthermore, status of HSF1 and HSP70 suggests an impaired cytosolic unfolded protein response in PEXS individuals.

2085F

Epigenetic changes in monozygotic twins discordant for nonsyndromic cleft lip and palate. *A. Vedenko[1], J.I. Young[1], S. Lang[1], J.T. Hecht[2], S.H. Blanton[1].* 1) Dr.John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 2) Department of Pediatrics, McGovern Medical School, University of Texas Health, Houston, TX.

Nonsyndromic cleft lip with or without cleft palate (NSCLP) is the most common craniofacial birth defect, with an estimated birth prevalence of 1/700 newborns. Etiologic studies have traditionally focused on genetic and environmental risk factors, but recent work has highlighted a role for epigenetic processes in mediating susceptibility to complex disorders. To investigate the contribution of epigenetic factors to NSCLP, we examined epigenetic changes in monozygotic twins discordant for NSCLP. The study of discordant monozygotic twins allows the assessment of the epigenome virtually independent of the influence of underlying genomic sequence variation. Thus, we performed whole genome bisulfite sequencing on saliva DNA from 3 monozygotic twin pairs discordant for NSCLP. Analysis of the full methylome identified significantly distinct methylation patterns within the twin pairs. We searched for intra-twin methylation differences using a candidate gene approach as well as a genome-wide unbiased manner. Following the hypothesis that genes previously associated with NSCLP are candidates for carrying metastable epialleles, we first examined our set candidate genes for evidence of differential methylation. We identified and confirmed differentially methylated sites for *GSK3B*, *TP63*, *VEGFC*, *FGF9*, *FGF14*, *PDGFC*, *PDGFRA* and *CAV3*. In addition, high stringency genome-wide analysis of the discordant methylomes revealed 436 genes containing significant methylation differences. Notably, pathway enrichment analysis of the differentially methylated genes identified cell-cell adhesion as the top biological process (GO:0098742, adjusted p-value: 0.001). Altogether, our data suggest that DNA methylation differences originating from stochastic or environmentally induced factors play a critical role in mediating phenotypic differences within MZ twins and, therefore, in the etiology of NSCLP.

2086W

Identification of regulatory element variants in craniosynostosis. *Y. Zhou[1], A.L. Fenwick[1], S.J. McGowan[2], J.R. Hughes[3], A.O.M Wilkie[1], S.R.F Twigg[1].* 1) Clinical Genetics Group; 2) Computational Biology Research Group; 3) MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

Craniosynostosis, the premature fusion of the cranial sutures, is a serious disorder with a prevalence of 1 in 2,100. Monogenic causes are associated with an increased frequency of complications, often leading to long-term health problems. Despite recent successes in identifying new genetic causes, a significant proportion of patients (~60%) analysed by exome or whole genome sequencing remain undiagnosed. One potential explanation is that some causative variants lie in non-coding DNA, in regions not routinely analysed in NGS data. As an initial step towards identifying such mutations, we carried out genome-wide chromatin analysis of human fetal (12-14 weeks post conception) cranial sutures and suture-derived cells (human & mouse), using histone H3 K27ac and K4me1 marks to map active regulatory elements (REs). In addition, ATAC-seq and Capture C were performed on suture-derived cells, probing open chromatin regions and revealing looping interactions between selected promoters and putative enhancers. As the search space is large, we have initially focused on genes for which haploinsufficiency leads to craniosynostosis, as RE mutations in such genes could lead to pathological under-expression. Saethre-Chotzen syndrome, for example, is caused by heterozygous loss-of-function mutations of *TWIST1*, which encodes a bHLH transcription factor. The intersection with targeted resequencing data from the *TWIST1* locus of 152 patients with craniosynostosis has pinpointed several candidate RE variants. These include 1) a cluster of single nucleotide variants from 6 patients within a 500 bp region that intersects with a putative RE (human & mouse) ~5 kb from the *TWIST1* transcription start site (TSS) and 2) a ~2.9 kb deletion in a patient with craniosynostosis, mediated by non-allelic homologous recombination between *AluY* elements, enriched for H3K27ac/K4me1 (human & mouse) ~14 kb from the *TWIST1* TSS. In summary these chromatin data illustrate the feasibility of identifying putative REs from early-stage embryonic tissues and primary cells, and suggest that several key REs are conserved between humans and mice, enabling further functional testing of their developmental role. Intersection with resequencing data from craniosynostosis patients highlights possible causative non-coding variants around *TWIST1*. This combined approach will lead to a better understanding of craniosynostosis, improve molecular diagnosis and increase understanding of cranial suture development.

2087T

The emerging landscape of dynamic DNA methylation in early childhood. *C.J. Xu[1,2], M.J. Bonder[2], C. Söderhäll[3,4], M. Bustamante[5,6,7,8], N. Baïz[9], U. Gehring[10], S.A. Jankipersadsing[1,2], P. van der Vlies[2], C. van Diemen[2], B. van Rijkom[2], J. Just[9], I. Kull[11], J. Kere[3,12], J.M. Antó[5,6,7,8], J. Bousquet[13,14,15], C. Wijmenga[2], I. Annesi-Maesano[9], J. Sunyer[5,6,7,8], E. Melén[16], Y. Li[2], D.S. Postma[1], G.H. Koppelman[17].* 1) University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC Research Institute, Groningen, the Netherlands; 2) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 3) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; 4) Department of Women´s and Children´s Health, Karolinska Institutet, 171 77 Stockholm, Sweden; 5) ISGlobal, Centre for Research in Environmental Epidemiology, Barcelona, Spain; 6) Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain; 7) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 8) Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Spain; 9) Epidemiology of Allergic and Respiratory Diseases Department (EPAR), Sorbonne Université, UPMC Univ Paris 06, INSERM, Pierre Louis Institute of Epidemiology and Public Health, Saint-Antoine Medical School, Paris, France; 10) Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; 11) Department of Clinical Science and Education, Stockholm South General Hospital, Karolinska Institutet, and Sachs' Children´s Hospital, SE-118 83 Stockholm, Sweden; 12) Folkhälsan Institute of Genetics and Research Programs Unit, University of Helsinki, Helsinki, Finland; 13) University Hospital, Montpellier, France; 14) MACVIA-LR, Contre les Maladies Chroniques pour un VIeillissement Actif en Languedoc-Roussillon; 15) INSERM, VIMA: Ageing and chronic diseases. Epidemiological and public health approaches, U1168, Paris and UVSQ, UMR-S 1168, Université Versailles St-Quentin-en-Yvelines, France; 16) Institute of Environmental Medicine, Karolinska Institutet, Centre for Occupational and Environmental Medicine, Stockholm County Council and Sachs' Children´s Hospital, Stockholm, Sweden; 17) University of Groningen, University Medical Center Groningen, Department of Paediatric Pulmonology and Paediatric Allergy, Beatrix Children's Hospital, GRIAC Research Institute, Groningen, the Netherlands.

   DNA methylation is known to influence many aspects of life, but it is unknown how genome, environment and disease influence DNA methylation dynamics in childhood. By analysing 538 paired DNA blood samples from children at birth and at 4-5 years old and 726 paired samples from children at 4 and 8 years old from four European birth cohorts using the Illumina Infinium Human Methylation 450k chip, we have identified 14,150 consistent age-differential methylation sites (a-DMSs) at epigenome-wide significance. Genes with an increase in age-differential methylation were enriched in pathways related to 'development', while genes with a decrease were enriched in 'regulation'. Changes in 101 a-DMSs (0.71%) were partly regulated by single nucleotide polymorphisms with more prominent genetic effects seen in later childhood. Maternal smoking was associated with more strongly decreased methylation levels in early life. a-DMS-associated genes were significantly more likely to be linked with disease. Our study provides new insights into the dynamic epigenetic landscape of the first 8 years of life.

2088F

Dissecting the epigenetic and transcriptional landscapes in spermatogonial stem cells ageing. *T. Lee, J. Liao, S. Ng, W. Chan.* School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.

   Spermatogenesis maintains male fertility through a highly complicated process, consisting of spermatogonial proliferation, meiosis of spermatocytes and spermiogenesis of haploid spermatids. This process is sustained by a pool of rare male germline stem cells known as spermatogonial stem cells (SSCs) that either self-renew or differentiate to progenitor spermatogonia to initiate spermatogenesis. Similar to other developmental programs, this process is tightly coordinated through epigenetic and transcriptional programs and any failure in this process could have catastrophic effects on male reproductive biology and cause significant health impacts in the offspring. One of the ageing-associated characteristics in the male reproductive system is the dramatic decline in stem cell content. However, the molecular regulations underlying the function deterioration of aged SSCs remain largely elusive. In this study, we used highly purified SSCs from young and aged mice along with an integrated genomic approach to evaluate ageing-associated alterations in the transcriptome and epigenome. We first examined the transcriptional landscapes of SSCs and delineated the molecular mechanisms underlying SSC ageing, which are characterized by the dynamic interplay between cell-intrinsic and environmental signals. We also constructed an in-depth map of epigenetic regulation of SSCs including DNA methylation (5mC), DNA hydroxymethylation (5hmC) and histone modifications (H3K4me3/H3K27me3). Our analysis demonstrated that alterations in both DNA methylation and histone modifications influence changes in gene expression and represents one of the driving forces behind ageing-related declined in the functionality of SSCs. Taken together, the findings of this study have uncovered comprehensive information of the transcriptional and epigenetic regulation in SSC ageing, which will provide the foundation for novel therapeutic options against spermatogenic failure and reproductive ageing.

2089W

Difference in plasma circulating cell-free microRNAs between young and aged mice: Effects of miRNAs that are predominantly present in young mouse serum on myogenic differentiation in C2C12 cells and primary myoblasts. *M. Fukuoka[1], N. Ito[2], S. Takeda[2], H. Hohjoh[1].* 1) Department of Molecular Pharmacology, National Institute of Nuroscience, National Center of Neurology and Phychiatry, Kodaira, Tokyo, Japan; 2) Department of Molecular Therapy, National Institute of Nuroscience, National Center of Neurology and Phychiatry, Kodaira, Tokyo, Japan.

MicroRNAs (miRNAs) that are 21~23-nucleotide-long small non-coding RNAs function as mediators in gene silencing, and play essential roles in gene regulation by inhibiting translation of messenger RNAs (mRNAs) and by digestion of mRNAs. Recently miRNAs have been found in various body fluids including plasma/serum as well as inside cells, and a part of such miRNAs appear to be enclosed in extracellular vesicles such as exosomes and be associated with RNA-binding proteins, e.g., Argonaute2. Such cell-free (extracellular) miRNAs might be implicated in cell-cell communication and be capable of becoming useful biomarkers. Aging is characterized by a progressive loss of physiological integrity, in which functional declines in physiology such as a decrease in hearing, visual acuity, muscle strength and memory skills occur. To see such aging at a molecular level, we investigated plasma-circulating (cell-free) miRNAs prepared from young (6-week-old) and aged (2-year-old) C57BL6/J mice by expression profile analyses in this study. The results indicated that the level of myogenic miRNAs, which are involved in myogenic differentiation, was higher in the plasma of young mice than in that of aged mice. In addition, young mouse serum, but not aged serum, appeared to have the potential for inducing myogenic differentiation of mouse C2C12 cells (a myoblast cell line). We further investigated the effects of the miRNAs increased in young mouse serum on myogenic differentiation using C2C12 cells: miRNAs of interest were introduced into C2C12 cells and the expression of the *Myogenin* and *Myh1* genes that were used as indicators for myogenic differentiation was examined. As a result, a certain miRNA appeared to have a strong induction for myogenic differentiation relative to other examined miRNAs, and might be associated with the increase in the level of miR-1, which plays a role in gene regulation during myogenic differentiation. In addition to the above experiments, we introduced synthetic mimics of that miRNA into mouse primary myoblasts and confirmed their induction to myogenic differentiation. Our findings indicated that miRNAs played an important role in gene regulation during myogenic differentiation, and suggested the possibility that decrease in certain miRNAs in aged mouse plasma might be associated with inefficient muscle regeneration in aged mice. Our further studies are in progress, and we would like to discuss this topic together with additional data in the meeting.

2090T

Early-life socio-economic status is associated with accelerated epigenetic ageing in monocytes. *E. Gatev[1], N. Gladish[1], J. MacIsaac[1], L. McEwen[1], M. Jones[1], G. Miller[2], S. Mostafavi[1], M. Kobor[1].* 1) UBC, Vancouver, British Columbia, Canada; 2) Northwestern University, USA.

Socio-Economic Status (SES) during Early Life (EL) has been associated with susceptibility to chronic diseases in adulthood. With their first years in difficult socio-economic conditions, children have increased vulnerability to infectious, cardiovascular, respiratory and cancer diseases in adulthood. Low SES increases the likelihood of chronic psychological stress from living in unfavourable conditions. There is evidence that chronic stress contributes to the development and progression of pathologies. The plasticity of dynamic DNAm suggests a mechanism where persistent environmental factors, like chronic stress, cause chromatin modifications that affect developmental programming, and remain detectable in adult life. In this way, variability of environmental exposures across individuals is reflected in age-related DNAm trajectories. "Epigenetic age" constructs measure deviations from the average age-related DNA methylation trajectory in a population, and quantify inter-individual variability of age-related DNAm. We develop a novel bioinformatic approach to construct a monocyte-specific (MS) epigenetic clock that incorporates prior knowledge about DNAm sites associated with age. We use this MS epigenetic clock to asses epigenetic age acceleration. Our MS epigenetic clock has greatly improved accuracy over the pan-tissue clock of Horvath, while using much fewer MS DNAm sites. It includes novel MS DNAm sites that are associated with multiple genes involved in cell differentiation, growth and transcription factor binding.

2091F

The effect of age on sex-specific DNA methylation patterns in a Bangladeshi population. *R.J. Jansen[1], L. Tong[2], M. Argos[3], F. Jasmine[2], M.G. Kibriya[2], J.A. Baron[4], H. Ashan[2], B.L. Pierce[2].* 1) North Dakota State University, Fargo, ND; 2) University of Chicago, Chicago, IL; 3) University of Illinois, Chicago, IL; 4) University of North Carolina, Chapel Hill, NC.

It is well-known that epigenetic changes occur as human age, and there is a growing interest in characterizing age-related changes in DNA methylation. In this study, we set out to characterize the effect of age on DNA methylation in a sex-specific manner and determine if these effects vary by genomic context. We used the Illumina HumanMethylation 450K Beadchip and DNA derived from whole blood for 400 adult participants (189 males and 211 females) from Bangladesh to identify age-associated CpG sites and regions, and characterize to what extent these age-associated sites were located in CpG islands (vs. shore, shelf, or open sea), and gene regions (vs. intergenic). We conducted a genome-wide search for age-associated CpG sites (among 423,604 sites) in two ways: 1) using a reference-based adjustment for cell type proportion (the R package: MethylSpectrum) and 2) using a reference-free approach for cell type proportion adjustment (the R package RefFreeEWAS). Using method 1 there was a noticeable shift of the model estimates toward hypermethylation which was not present using method 2; potentially representing the effects of unmeasured confounding variables. As is consistent with previous reports, CpG sites that were differentially methylated with respect to age were more likely to be hypermethylated compared to hypomethylated, and to be found in CpG Islands and promotor gene regions compared with other locations. For men and women respectively, significant associations with age (p-values < 10^{-8}) for individual CpG sites were 568 and 1819, of these 272 (47.9%) and 707 (38.9%) CpGs were found in island regions while 185 (32.6%) and 631 (34.7%) were found in promotor gene regions. Although we observed good correlation with previously developed methylation age models (r≈0.8), among the top 100 CpG sites only 2 for males and 5 for females are included in these prediction models. Our results indicate that age-associated methylation tends to be consistent for both sexes (correlation among top 100: 0.98, and 0.99 males and females, respectively). Our population may have unique age-related methylation changes that are not included in the established methylation age prediction models.

2092W

DNA methylation in the apolipoprotein E locus in late life. *I.K. Karlsson[1], A. Ploner[1], P. Garagnani[3,4], C. Franceschi[3,4], T. Morris[5], S. Hägg[1], N.L. Pedersen[1,2].* 1) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 2) Department of Psychology, University of Southern California, Los Angeles, California, USA; 3) Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40126 Bologna, Italy; 4) Interdepartmental Center "L. Galvani", University of Bologna, 40126 Bologna, Italy; 5) Medical Genomics Group, UCL Cancer Institute, University College London, London, UK.

The Apolipoprotein E (*APOE*) gene has been shown to be of importance in both dementia and cardiovascular disease (CVD), as well as longevity. Recent evidence suggests a role for epigenetics in several diseases, including dementia and CVD, and epigenetic mechanisms may mediate gene-environment interactions that promote disease. This project aims to investigate if variation in DNA methylation of the *APOE* gene differs in relation to dementia, CVD, and healthy aging. For this purpose, DNA methylation levels were measured using the Illumina 450K array in 426 Swedish twins (mean age 78.4), as well as in 82 healthy Italian centenarians (mean age 105.5), 63 offspring to healthy centenarians (mean age 71.6), and 47 controls who were offspring to non-long lived parents (mean age 69.8). DNA methylation levels in the *APOE* gene have previously been categorized in three groups: a hypermethylated region in the promotor, a hypomethylated region in the first two introns and exons, and a hypermethylated region in the 3´exon. The mean M-values across these three groups were used in linear regression models to compare differences in methylation between dementia cases and controls, CVD cases and controls, centenarians, their offspring, and offspring controls. We found a higher level of DNA methylation in the *APOE* promotor in dementia cases compared to controls after adjusting for sex and age at blood sample (β 0.07, 95% CI 0.02-0.12). Healthy centenarians were found to have a significantly lower level of DNA methylation in the same region, and also in the 3´exon when compared to offspring controls after adjusting for sex (β -0.04, 95% CI -0.05 to -0.02 and β -0.02, 95% CI -0.03 to -0.01, respectively). We did not detect any difference in methylation levels between CVD cases and controls, or between offspring to centenarians and controls. In conclusion, we found that DNA methylation levels in the promotor region of the *APOE* gene were higher in dementia compared to controls, and lower in centenarians compared to controls. No difference was found in methylation of the *APOE* gene in relation to CVD or in offspring to centenarians. This may shed more light on the strong but largely unexplained association between *APOE* and dementia.

2093T

A longitudinal study of stochastic epigenetic mutations and aging. *Y. Wang[1], R. Karlsson[1], J. Jylhävä[1], T. Morris[2], P. Garagnani[3,4], A.K. Hedman[5], N.L. Pedersen[1], S. Hägg[1].* 1) Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 2) Medical Genomics Group, UCL Cancer Institute, University College London, London, UK; 3) Department of Experimental, Diagnostic and Specialty Medicine Experimental Pathology, University of Bologna, Bologna, Italy; 4) Center for Applied Biomedical Research, St. Orsola-Malpighi University Hospital, Bologna, Italy; 5) Department of Medical Sciences & SciLife Lab, Uppsala University, Uppsala, Sweden.

   DNA methylation plays an important role in gene regulation together with other epigenetic markers. Aberrant DNA methylation patterns can result in genomic instability, be involved in pathogenesis of age-related diseases and be associated with human aging. In a cross-sectional study, Gentilini et al. 2015 found that the number of stochastic epigenetic mutations (SEMs) increased exponentially with chronological age, defining SEMs as extreme outliers in a population with methylation levels three times interquartile below the first quartile ($<Q1-3 \times IQR$) or above the third quartile ($>Q3+3 \times IQR$). Here we replicated those results and furthermore investigated if SEMs could persist and accumulate over the aging process in a longitudinal perspective. We characterized DNA methylation patterns in whole blood using HumanMethylation450 BeadChip array. Samples (n=1011) were collected from 385 old Swedish twins up to five times over 20 years (age from 48 to 94 at the baseline). A longitudinal analysis using a mixed-effects model showed that the number of SEMs grows exponentially with age ($p=1.26 \times 10^{-11}$). However, technical variance (detection p-value), batch effect and cellular composition were found to be strong confounders even though they had been adjusted for in prior preprocessing steps. In single site analysis, we identified some SEMs arising from a normal methylation level at one time point and maintained as mutations in later time points. Moreover, we observed that 3443 CpG sites have a much higher chance of mutation than other sites ($p=2.2 \times 10^{-16}$). These CpG sites are more prone to epigenetic mutations and they are different from previously identified age-associated CpG sites. In summary, we verified that the number of SEMs increase exponentially with age from longitudinal observations, indicating that acquired SEMs can be persistent and accumulate in the aging process.

2094F

Longitudinal study of gene expression in a Swedish population sample. *B. Balliu[1], K.S. Smith[1], C. Lindgren[2], A. Morris[3], L. Lind[4], E. Ingelsson[4,5], S. Montgomery[1,6].* 1) Department of Pathology, Stanford University School of Medicine, Stanford, CA; 2) Wellcome Trust Centre for Human Genetics; 3) Institute of Translational Medicine, University of Liverpool; 4) Department of Medical Sciences, Uppsala University; 5) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 6) Department of Genetics, Stanford University School of Medicine, Stanford, CA.

   While DNA sequence is more or less stable throughout life, gene regulation and expression is dynamic, fluctuating in response to different environmental exposures. Increased knowledge about how genes are regulated and expressed, and how this changes over time in relation to environmental triggers and disease outcomes can give important insights to fundamental questions about gene function. Using a unique study sample with repeated sampling, we performed whole-genome RNA sequencing analysis in white blood cells at two time points with ten years apart. The PIVUS study was initiated in 2001-2004, where 1,016 70-year-old individuals from the general population participated in an extensive examination of their health, including biological samples and subsequent genome-wide SNP genotyping. Participants were re-examined at age 75 and 80. We sequenced the blood transcriptomes of 62 PIVUS participants at both age 70 and 80 (104 samples) and observed that gene expression is strongly correlated (mean=.97) between the two time points, and that for the majority of samples (~66%), measurements in the same individual at the two different ages cluster together. The remaining of the samples clustered with samples from the same age group. This illustrates that although most individuals demonstrate an expression pattern that is relatively similar to themselves over 10 years of aging, there is also a large proportion who are more similar to peers from their age group than themselves 10 years earlier or later. The latter is likely to reflect the aging process, changes in environment, subclinical traits or disease outcomes. Moreover, we identified 2,699 genes showing differential expression between the two time points (1% FDR), and a strong enrichment (.01%FDR) for age related GO terms (immune and stress response), KEGG pathways (MAPK signaling and metabolic pathways), and disease-associated genes (inflammation, vascular disease). We further discovered 403 and 303 eGenes, *i.e.* genes with a cis-eQTL (1% FDR), at the first and second time point, respectively. cis-eQTLs detected at both time points were largely shared ($\pi 1= 73$ %) indicating that genetic regulation of gene expression is mostly stable over time. In summary, we have performed one of the first long-term longitudinal studies of genetics of gene expression that highlights transcriptome dynamics late in life.

2095W

DNA methylation and aging in mice and humans. *K. Mozhui[1,2], A.K. Pandey[2].* 1) Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN.

Aging has a profound influence on the epigenetic structure of the DNA. In humans, methylation patterns at specific CpG sites have been used to define the 'epigenetic clock', a molecular biomarker of age that could provide a means to discern the health and fitness of an individual. The mechanisms that establish these intriguing patterns in the methylome remain largely elusive. It is also unclear whether CpG markers for age identified in humans (that rely mostly on microarrays) can be generalized to experimental models such as the mouse. Here we use an affinity-based capture of DNA using methyl-binding domain (MBD) protein and high-throughput DNA sequencing (MBD-seq) for a genome-wide assay of DNA methylation in aging mouse liver. Consistent with previous reports, we find that the most pronounced effect of age is an increase in methylation at CpG sites encompassing the transcription start site (TSS) of genes (e.g., *C1ql3*, *Srd5a2*, and *Rfx4*). In contrast to the hypermethylation at specific CpG islands, there is age-dependent hypomethylation at most other regions. The differentially methylated regions (DMRs) identified in mice have high sequence conservation in humans and the age-dependent patterns are consistent between the two species, and detected in more than one human tissue (blood and liver). In humans, these conserved age-hypermethylated regions are binding sites for members of the polycomb repressor complex (PRC), including EZH2. A targeted chromatin immunoprecipitation assay confirms that these DMRs are associated with EZH2 in mice, and there may also be age-dependent reduction in EZH2 occupancy. This adds to growing body of evidence that EZH2 and other members of PRC are part of the protein machinery that establish and shape the aging epigenome. Given the level of conservation and consistency, the mouse model can serve as a powerful experimental basis to study the nature and molecular modifiers of the epigenetic clock.

2096T

DNA methylation in aging twins across 10 years: Heritable and environmental contributions. *C.A. Reynolds[1], E. Munoz Diaz[1], Y. Wang[2], T. Morris[3], N.L. Pedersen[2,4], S. Hägg[2].* 1) Department of Psychology, University of California, Riverside, Riverside, CA, USA; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 3) Medical Genomics Group, UCL Cancer Institute, University College London, London, UK; 4) Department of Psychology, University of Southern California, Los Angeles, CA, USA.

DNA methylation may result from environmental exposures as well as be influenced by genetic factors, with dynamics in methylation associated with increasing age in adulthood. Age-related change in DNA methylation in adulthood has been examined largely using cross-sectional data, with few investigations of genetic and environmental influences that may contribute to stability and change in late life. Methylation profiles (Illumina HumanMethylation450 array) were evaluated in 53 pairs of twins from the Swedish Adoption Twin Study of Aging (SATSA). The pairs selected for analysis included 21 MZ and 32 DZ pairs with an average of 10 years between two in-person assessments. At the first assessment the average age was 62.9 years (range 50.7 - 77.9 years) and 72.5 years at the second (range 59.9 – 86.8). Pre-processing of data included QC adjustments at the sample and probe level, and normalizations and adjustments for confounders (cell counts, batch effects), resulting in an analysis set of 329,341 probes. Biometrical model fitting was conducted on M-values adjusted for age and sex. ACE and ADE models were evaluated for each CpG site, where A represents additive genetic contributions, D non-additive genetic contributions, C common environmental contributions, and E nonshared or stochastic environmental contributions to variation in M-values. E was the sole contributor to variation in methylation across 77.1% of sites at age 62.9, and to 80.5% of sites 10 years later. C contributed to variation on 7.1% and 7.0% of sites, respectively at each age, where common environmentality was .28 on average. At age 62.9 years, 15.8% of sites evidenced some genetic variance (A and/or D), and broad heritabilities were .46 on average; 10 years later, 12.4% evidenced genetic variance with broad heritabilities of .45 on average. Genetic influences were statistically signficant (p < 1.00E-07) for 1.8% sites at age 62.9 and 1.4% sites 10 years later. Persistence of significant genetic influences across the two ages was observed for 0.50% of CpG sites, whereas 0.56% were significant at one age but not another. Overall, results suggest small genetic contributions to variation in DNA methylation in late adulthood. New influences emerge over time, primarily in terms of stochastic environmental contributions and for a subset of CpG sites, novel genetic influences may contribute to variation.

2097F

Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome. *R. Jeremian[1], E. Oh[1], G. Oh[1], D. Groot[1], A. Petronis[1], V. Labrie[2].* 1) Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, MI, USA.

Transcriptional variation in histologically- and genetically-identical cells is a widespread phenomenon in tissues, yet the processes conferring this heterogeneity are not well understood. To identify contributing factors, we analyzed epigenetic profiles associated with the *in vivo* transcriptional gradient of the mouse lactase gene (*Lct*), which occurs in specialized epithelial cells (enterocytes) along the proximal-to-distal axis of the small intestine. We found that epigenetic signatures at enhancer and promoter elements align with transcriptional variation of *Lct*. Age- and phenotype-specific environmental cues (lactose exposure after weaning) induced changes to epigenetic modification density and CTCF binding at select regulatory elements, which corresponded to alterations in the intestinal *Lct* mRNA gradient. Thus, epigenetic modifications in combination with CTCF binding at regulatory elements account for the transcriptional gradient in *Lct* in cells of the same type, and this epigenetic divergence within enterocytes may contribute to the functional specialization of intestinal subregions. Our findings emphasize the importance for future studies examining epigenetic contributions to the transcriptional variances in multiple genes within a cell type, as these could help explain why tissue subregions can perform diverse biological functions and vary widely in disease susceptibility and treatment. In addition, our findings provide an important lesson for epigenetic studies of phenotypes, as failure to consider within cell-type transcriptional variation and epigenetic divergence limits the detection of biologically significant effects. Future epigenetic and gene regulation studies in health and disease will be greatly refined by not only isolating the cell type of interest, but by sampling a single cell type across tissue subregions, and across aging and environmental parameters. .

2098W

3' UTR fragments accumulate in aging neuronal subtypes of mice and humans. *P.H. Sudmant[1], M. Heiman[2], C.B. Burge[1].* 1) Biology, Massachusetts Institute of Technology, Cambridge, MA; 2) Picower Institute for Learning and Memory Massachusetts Institute of Technology, Cambridge, MA.

Molecular differences among cell types and tissues can result in varied responses to genetic or environmental insults such as those incurred during aging. To assess how the transcriptome of closely related neuronal subtypes diverges during development and aging we performed translating ribosome affinity purification (TRAP) and sequencing of mRNA from D1 and D2 medium spiny neurons (MSNs) of the mouse striatum across various time-points. Surprisingly, aged (2yrs) D1 MSNs diverged substantially in their expression patterns from aged D2 MSNs specifically in translation initiation factors and components of oxidative phosphorylation. These gene expression patterns were accompanied by the pronounced accumulation of ~2000 RNA fragment species consisting of the 3' UTRs of genes without their cognate CDS and 5' UTR sequences. We find that these 3'UTR fragment species also accumulate in aging human brains to varying degrees among different neuronal tissues and provide evidence that 3' UTR fragments can be induced by oxidative stress. These results shed light on a putative novel RNA species that may impact the aging nervous system.

2099T

The role of the ENCODE Data Coordination Center. *J.M. Davidson[1], E.T Chan[1], I. Gabdank[1], J.A. Hilton[1], A.K. Narayana[1], K.C. Onate[1], M.O. Ho[1], B.T. Lee[2], L.D. Rowe[1], N.R. Podduturi[1], D.I. Glick[1], U.K. Baymuradov[1], V.S. Malladi[2], S.R. Miyasato[1], T.R. Dreszer[1], J.S. Strattan[1], F.Y. Tanaka[1], B.C. Hitz[1], C.A. Sloan[1], E.L. Hong[1], J.M. Cherry[1], ENCODE CONSORTIUM.* 1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305; 2) Center for Biomolecular Science and Engineering, School of Engineering, University of California Santa Cruz, Santa Cruz, CA 95064.

In recent years there has been an influx in the quantity of publicly available, large genomic datasets from individual labs and large consortia. Here we describe the efforts of the Data Coordination Center (DCC) in improving the accessibility of data available for the Encyclopedia of DNA Elements (EN-CODE) project. ENCODE is a collaborative effort to generate a comprehensive catalog of functional elements in human and mouse genomes. The ENCODE database currently includes more than 40 experimental techniques in over 400 tissue types and cell lines to analyze DNA and RNA-binding proteins, transcription and chromatin structure. All experimental data and computational analyses of these data are submitted to the DCC for validation, tracking, storage, and distribution to the scientific community. To ensure that the data generated by the production labs and the analysis performed on these data are accurately represented, the ENCODE DCC works closely with members of the Consortium groups to capture structured metadata related to experimental conditions, data quality metrics, and analysis methods. These experiments can be accessed via the ENCODE portal (http://www.encodeproject.org). Portal users query the database by searching for specific metadata terms, such as "p53", or "K562" or by utilizing faceted searching of the structured metadata. The portal also supports the visualization of certain data files by launching a Genome Browser track hub. Data files can be downloaded either directly from the experiment pages at the portal or via bulk download by programmatically accessing the ENCODE REST API. By providing direct data downloads based on flexible and powerful search capabilities that rely on highly organized metadata, the DCC strives to expand the access of ENCODE data to the scientific community.

2100F

The contribution of epigenetics to the phenotypic variation in a set of monozygotic twins with Gaucher disease. *N. Tayebi[1], S. Hassan[1], F. Donovan[2], S. Chandrasekharappa[2], M. Biegstraaten[3], C. Hollak[3], E. Sidransky[1].* 1) Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI/NIH, Bethesda, MD; 2) Cancer Genetics and Comparative Genomics Branch/Genomics Core, NHGRI/NIH, Bethesda, MD; 3) Academic Medical Center, Amsterdam, The Netherlands.

Background: Gaucher disease (GD) is an inherited autosomal recessive disorder characterized by mutations in *GBA1*, encoding the lysosomal enzyme glucocerebrosidase. The defective enzyme leads to lysosomal accumulation of glucocerebroside, characteristically in macrophages. Phenotypic heterogeneity among patients with the same genotype, even between sib pairs and identical twins, has prompted the search for genetic modifiers or epigenetic modification. Studying the contribution of epigenetic marks, such as DNA methylation, MicroRNAs and histone modifications, has been challenging due to the large variance between genomes. One approach to investigating epigenetic variation has been to evaluate samples from twins. In 2011 we described twin sisters, confirmed to be monozygotic by DNA fingerprinting, who both had GD, but with divergent phenotypes. Mutational analysis revealed that they were homozygous for the *GBA1* mutation N188S, c.670C>T. On examination, one sister exhibited severe visceral involvement, epilepsy, and cerebellar atrophy, while the other had no signs or symptoms related to GD.   **Method**: To detemine the cause of this discrepancy, we performed epigenetic evaluation of gene inactivation by looking at methylation of CpG islands from the genomic DNA of each twin. The bisulfite-based DNA modification was used to discriminate between cytosine and methylated cytosine in genomic DNA extracted from peripheral blood. Treated DNAs were applied to an Illumina HumanMethylation450 BeadChip. Data were analyzed using GenomeStudio and GeneMANIA for network analysis. **Results:** In the more affected twin, five methylated, and three demethylated genes were identified at CpG islands (P-value= 0.01-0.001) as well as two methylated genes from enhancer regions (P-value 0.05-0.01). The significant methylated genes, which included *CACNA1A*, were validated using bisulfite-treated DNA and Sanger sequencing. RNA expression of the selected genes is being evaluated by real-time PCR, using RNA extracted from at least two different tissues from each twin. **Conclusions:** This study enhances our understanding of the complex genotype-phenotype relationship in GD, and identifies a possible cause of phenotypic variation. Additionally, the data suggest an effect of epigenetic events on disease manifestations, which may contribute to phenotypic variation in monogenic and multifactorial diseases.

2101W

Gestational age acceleration in neonates predicts NICU course. *A. Knight[1], A.K. Smith[1], K.N. Conneely[1], P. Dalach[2], Y.L. Loke[2], J. Cheong[2,3,4], J.M. Craig[2,4], L.W. Doyle[2,3,4], C. Theda[2,3,4].* 1) Emory University, Atlanta, GA; 2) Murdoch Childrens Research Institute, Parkville, Victoria, Australia; 3) Royal Women's Hospital, Parkville, Victoria, Australia; 4) University of Melbourne, Parkville, Victoria, Australia.

DNA methylation changes with age, and can be used to accurately predict chronological age in adults. We have recently developed a methylation-based predictor of gestational age (DNAm GA) for neonates using cord blood or blood spot samples at birth that is highly correlated to clinically-estimated gestational age (GA) based on ultrasound (r=.91; p<2.2E-16) and is accurate within 1.24 weeks in >1,200 samples. DNAm GA increased steadily in longitudinal samples taken from birth to term equivalency in two preterm neonates admitted to the neonatal intensive care unit (NICU). We hypothesized that GA acceleration, which is the residual between clinically estimated GA and DNAm GA, would predict neonates requiring the most intensive interventions in the NICU. The Victorian Infant Collaborative Study (VICS, Victoria, Australia) provides a unique resource to address this hypothesis as preterm infants born before 37 weeks gestation in 1991/1992 were prospectively enrolled in a longitudinal follow-up study. DNA was extracted from dried blood spots used for newborn screening of 183 participants (26.7±2.1 weeks) and was processed on the Human Methylation450 BeadChip. DNA methylation data was used to estimate DNAm GA. Data regarding common NICU interventions (surfactant administration, postnatal steroids, oxygen supplementation) and complications (bronchopulmonary dysplasia) were retrieved from the VICS database. Linear regression models that controlled for cellular heterogeneity and GA were used to evaluate the association of GA acceleration and each NICU outcome. Decreased GA acceleration (lower DNAm GA relative to GA) associates with increased administration of postnatal steroids (31.1% given steroids; p=.002) and surfactant (43.7% given surfactant; p=.02). Similarly, GA acceleration was lower in neonates that required more days on oxygen supplementation (p=.002). Finally, lower GA acceleration associated with the development of bronchopulmonary dysplasia, a chronic lung disorder affecting 36.0% of these neonates (p=.01). This suggests that GA acceleration may be indicative of developmental maturity at birth. Further studies should evaluate DNAm GA and GA acceleration to determine how these measures could contribute to more accurate and individualized predictions of NICU course and short and long-term outcomes.

2102T

Integrated metadata-driven access of ENCODE, modENCODE, REMC, GGR and modERN data through a common portal. *E.T. Chan, J.M. Davidson, I. Gabdank, J.A. Hilton, J.S. Strattan, A.K. Narayanan, M.C. Ho, K.C. Onate, U.K. Baymuradov, T.R. Dreszer, D. Glick, F.Y. Tanaka, B.C. Hitz, C.A. Sloan, J.M. Cherry.* Genetics, Stanford University, Palo Alto, CA.

The efforts of large-scale NIH-funded projects such as the Encyclopedia of DNA Elements (ENCODE), its model organism corollary (modENCODE), and the Roadmap Epigenomics Mapping Consortium (REMC) have resulted in the accumulation of genome-wide maps of candidate functional sequence elements and epigenetic marks in a large variety of cells and tissues in human and other model organisms. Consideration of the provenance of experimental reagents and transparency of computational analyses is crucial to the interpretation and comparison of these data. Tracking this information consistently across different biochemical assays employed in thousands of experiments across hundreds of cell and tissue types is particularly challenging at the scale of these projects.The ENCODE Data Coordination Center (DCC) has developed a flexible and rich data model to capture information such as key experimental variables, details of experimental and analysis methods, what software and pipelines were used to produce which files for the ENCODE project, and calculated quality metrics, known collectively as metadata. This data model has since been extended to additionally integrate curated data and metadata from related projects such as REMC, modENCODE, Model organism Encyclopedia of Regulatory Networks (modERN) and the Genomics of Gene Regulation (GGR). The resulting publicly accessible data corpus encompasses over 10,000 experiments and is accessible through the ENCODE portal (https://www.encodeproject.org), which features a powerful faceted browsing interface, full-text search, and a REST API so users can easily search, filter, download and visualize the collection. Data integration and release is ongoing and continual. Learn more about how to get started on the ENCODE portal here: https://www.encodeproject.org/help/getting-started/.

2103F

The validation of array determined, differentially methylated CpG sites using targeted next-generation bisulfite sequencing. *L. Yan, R. Drennan, A. Meyer, A. Miller, E.A. Mead, S. Miller, J. Xu, M. Poulin.* EpigenDx, Hopkinton, MA.

The identification of differentially methylated regions of the genome is the focus of many epigenetic studies. The Infinium HumanMethylation450 Bead-Chip (or EPIC Array) allows researchers to interrogate more than 485,000 methylation sites (or 850,000 CpG sites) per sample. This technology is useful as an initial screening process to isolate potentially differentially methylated regions in many disease states or following specific drug treatments in epigenome-wide association studies (EWAS). The need to validate hundreds of potential CpG sites has become a challenge for many researchers. We have developed a validation system based on targeted next-generation sequencing (tNGS) that allows the validation of large numbers of array determined CpG sites at single-nucleotide resolution within a single validation panel. Panels are developed using a stepwise strategy that begins with bioinformatically validating the CpG loci and obtaining the target CpG sequence and 500 nucleotides up and downstream of the CpG. Utilizing the RepeatMask program of the UCSC genome browser, those CpGs within repeat sequence of the genome are discarded. The sequences are then *in silico* bisulfite converted and bisulfite sequencing assays are designed for each CpG site. Assays are multiplexed based on the assay design score and primer annealing temperature. Multiplex PCR reactions are performed and analyzed for amplification and PCR bias following next-generation sequencing. In this study we began with 280 possible CpG sites in which 164 (greater than 50% of assays) successful in silico designs were generated. 18 multiplex PCR reactions were carried out on a methylation mixing series of DNA. 96 of the 164 assays resulted in next-generation sequencing reads which will be analyzed for PCR Bias. Although it may seem that validation of only approximately 60% of the initial assay is low return, it allows researchers to begin formulating further experiments and hypotheses while different combinations of the other 40% of the assays are recombined in multiplex PCRs or in simplex PCR reactions to increase the percentage of successfully validated assays. .

2104W

DNA methylation analysis of Sox2 regulatory regions during osteogenic differentiation of mouse embryonic stem cell. *S. Batool.* Biosciences, COMSATS Institute of Information Technology, Islamabad, Punjab, Pakistan.

Embryonic stem cells (ES) have been shown to be able to differentiate into all lineages derived from embryos *in vitro* and hence can be used as a model system not only to study developmental programming and reprogramming events but also, once these events are known can be used to create differentiated cell types for transplantation therapies in clinical applications. Sox2 has been reported to be indispensable for mouse embryonic stem cells self-renewal and also, is one of the transcription factors used to reprogram somatic cells. This study was undertaken to study the role of DNA methylation in regulation of Sox2 during differentiation of mouse embryonic stem cells. Mouse embryonic stem cells were differentiated to osteogenic lineage through the formation of embryo bodies (EBs) – cellular aggregates partially recapitulating the early embryonic development. ES were found to be differentiating as evidenced by changes in cellular morphologies, alizarin red staining and upregulation of lineage-specific markers. Gene expression and immunocytochemistry analysis showed downregulation of Sox2 during osteogenic differentiation. SRR1 and SRR2, two conserved regulatory region of Sox2, were found to be methylated by methylation-sensitive PCR at all time-points chosen for analysis in differentiating cells. Three individual CpGs in SRR2 region were then further analysed by bisulphite sequencing which appeared unmethylated in both undifferentiated and differentiated embryonic stem cells. This data shows the involvement of DNA methylation in regulating the expression of Sox2 in differentiating embryonic stem cells and so need further exploration. The preliminary results described here show that ES cell differentiation is a good experimental system to profile expression of Sox2 and investigate the modulation of epigenetic changes at its regulatory regions as cells differentiate.

2105T

Methylome analysis in non-syndromic cleft lip/palate shows methylation differences both in Brazilian and British cohorts. *L. Alvizi[1], S.A.M. Ezquina[1], P. Stanier[2], M.R. Passos-Bueno[1].* 1) Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Sao Paulo, Brazil; 2) Genetics and Genomic Medicine, UCL Institute of Child Health, London, UK.

Non-syndromic cleft lip/palate (NSCL/P) is the most common congenital craniofacial malformation, with a prevalence of 1 affected individual to 700 live-births worldwide. NSCL/P is a complex disease and heritability studies show a genetic contribution ranging from 40% to 85%.. Even though GWASs have been extensively applied for the elucidation of NSCL/P genetic factors, the identified *loci* explain only a small fraction of the observed heritability. On the other hand, epidemiologic studies show an association between low socioeconomic status and NSCL/P, pointing out an important environmental contribution. Those data suggest that alternative mechanisms might be involved in the NSCL/P etiology, as dysregulation of epigenetic markers triggered by environmental factors. Aiming to better identify epigenetic factors for NSCL/P, we propose the investigation of DNA methylation differences between NSCL/P and control samples using an EWAS (Epigenome-Wide Association Study) approach. We hypothesized that DNA methylation sites act as epialleles in relevant genes for the craniofacial development and differences in their methylation levels predispose to NSCL/P. To test those hypotheses we performed a methylome analysis comparing 68 NSCL/P with 59 control samples (whole blood DNA), using the Infinium Human Methylation 450K platform (Illumina) and RnBeads pipeline. We found 578 methylation variable positions (MVPs) at $p<10^{-7}$. EpiExplorer analysis revealed an enrichment of those MVPs in active regions of the genome, as gene promoters, in comparison to the 450K distribution. We also found a significant intersection of MVPs with 18 previously reported NSCL/P candidate genes (Fisher´s Exact Test, p=0.00044). Using Ingenuity Pathway Analysis, we identified "Regulation of the Epithelial-Mesenchymal Transition Pathway" and "WNT Beta-Catenin Signaling" amongst the top 5 significant canonical pathways in the MVPs. We selected 11 MVPs for replication in an independent cohort of 171 NSCL/P and 177 controls samples from the United Kingdom using a Bisulfite Amplicon Sequencing approach. We found significant methylation differences for 4/11 tested MVPs in the replication cohort. Our data suggests a different methylome profile between NSCL/P and controls and point out common methylation epialleles associated to NSCL/P in two distinct populations. Our result also suggests that epigenetic factors, possibly directed by the environment, may play a role in NSCL/P etiology. Support:FAPESP/CNPq.

2106F

DNA methylation alterations in cord blood of children exposed to gestational diabetes mellitus *in utero. D.T. Butcher[1], R. Retnakaran[3], A. Hanley[3], R. Weksberg[1,2], J. Hamilton[2].* 1) Genetics & Genome Biology, Sickkids Research Inst, Toronto, Ontario, Canada; 2) The Hospital For Sick Children, Toronto, Ontario, Canada; 3) Mount Sinai Hospital, Toronto, Ontario, Canada.

Obesity, diabetes and heart disease are now serious global health issues. Substantial evidence from epidemiologic studies suggests fetal and early postnatal environment, including *in utero* exposure to maternal gestational diabetes mellitus (GDM), significantly impacts the development of these conditions. GDM, defined as glucose intolerance first recognized during pregnancy, is increasingly common, occurring in up to 25% of pregnancies. The biological processes underlying the effect of *in utero* exposure to GDM, although not well understood, are likely to include alterations to the epigenome in fetal life. Such alterations may shift set points for insulin response and lipid metabolism resulting in postnatal growth trajectories that favor an obesogenic phenotype and increase the risk of cardiometabolic problems in adult life. Genome wide DNA methylation (DNAm) was assessed using the Illumina Infinium Methylation450 Beadchip array from cord blood samples of children exposed to GDM *in utero* compared to controls. We identified 453 CpG sites, representing 273 genes, which have an absolute DNAm difference of 5% with p<0.05. Multiple CpG sites in paraoxonase-1 (PON1) demonstrated a gain of DNAm in the cord blood of children exposed to DNAm compared to controls. Sodium bisulfite pyrosequencing validated the gain of DNAm of PON1 across these CpG sites. PON1 protein associates with high density lipoprotein (HDL) which hydrolyzes a number of lactones as well as additional substrates. Fetal programming via DNAm may shift the expression profile of PON1. Interestingly, in obese children and adolescents decreased activity of PON1 has been shown to correlate with central obesity, oxidative stress and inflammation. Also studies of PON1 levels and activity have been shown to be lower in serum of adults with cardiovascular, diabetes and obesity then in controls. This study allowed us to begin to evaluate the impact of maternal metabolic events during pregnancy on DNAm status in the child. Long-term follow-up of the children in this cohort will enable us to determine whether the identified DNAm alterations in the cord blood can be used as biomarkers for future metabolic disease providing opportunities for screening and intervention studies.

2107W

Conjoint SNP-SNP effects influence the human methylome - Results from an epi-genome-wide interaction analysis. *A. Milnik[1,2,3], V. Vukojevic[1,2,4], T. Egli[1,2], T. Sengstag[6,7], M. Jacquot[6], C. Vogler[1,2,3], V. Freytag[1,2], A. Heck[1,2,3], D. Coynel[2,5], D.J.F. de Quervain[2,3,5], A. Papassotiropoulos[1,2,3,4].* 1) Division of Molecular Neuroscience, University of Basel, Basel, Basel City, Switzerland; 2) Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, CH-4055 Basel, Switzerland; 3) Psychiatric University Clinics, University of Basel, CH-4055 Basel, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, CH-4056 Basel, Switzerland; 5) Division of Cognitive Neuroscience, Department of Psychology, University of Basel, CH-4055 Basel, Switzerland; 6) sciCORE, Scientific Computing Center, University of Basel, CH-4056; 7) SIB - Swiss Institute of Bioinformatics.

   Background: The epigenome is an important regulatory factor in the context of cell development, aging and disease pathogenesis. DNA methylation is an epigenetic hallmark and can be measured at both high resolution and high throughput. While it is known that single nucleotide polymorphisms (SNPs; mQTLs) influence DNA methylation levels (CpG), pairs of SNPs may also conjointly influence methylation levels. **Analysis:** To further address this question, we analyzed 2nd-order dependencies of SNPs (SNP-SNP interactions) on CpG DNA methylation (DNA derived from blood) on an epi-genome-wide scale (1.5×10^{16} tests; 395'431 CpGs; 192'955 SNPs) in a sample of N = 533 healthy young adults. Replication was done in an independent sample of N = 319 healthy young adults. **Results:** We could identify 7'532'307 Bonferroni corrected significant CpG-SNP-SNP triplets. When applying a stringent QC in the replication sample (permuted P-values; same direction of effect), 2'119 CpG-SNP-SNP triplets survived replication (based on 764 unique CpGs). 94% of all replicated SNP-SNP pairs were located on the same chromosome as the CpG. Hence, we further filtered out SNP-SNP-pairs with an R^2 > 0.1 to reduce potential confounding effects of haplotypes, resulting in 1'531 CpG-SNP-SNP triplets (575 unique CpGs). For these CpGs, adding the SNP-SNP interaction term in the model increased the amount of variance explained from 30% to 44%. **Discussion:** The results of this epi-genome-wide interaction analyses suggest that especially mQTLs in *cis* conjointly influence DNA methylation levels.

2108T

Principal component informed dimensionality reduction approach for 450k methylation data. *T.G. Richardson, N.J. Timpson, C.L. Relton, T.R. Gaunt.* MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

   The high dimensional Illumina HumanMethylation450 (450K) BeadChip array has become a widely used platform in the field of epigenetics. However, measuring DNA methylation at over 480,000 sites poses both computational and statistical challenges. We have developed a framework to reduce the dimensionality of 450K data by taking the principal components of methylation across individual CpG islands. These regions of the genome are involved in transcription initiation and chromatin structure and therefore allowed us to retain important features of the data when reducing dimensionality. We applied our approach to 450k data across 5 time points from the Accessible Resource for Integrative Epigenomic Studies (ARIES) project using 1018 mother-offspring pairs. We observed association signals between prenatal smoking and methylation at previously reported key genes (*AHRR*, *MYO1G*, *CYP1A1* & *GFI1*) when analysing the dimensionally reduced dataset. This analysis required only 35,416 tests in comparison to the >480,000 undertaken in a conventional epigenome wide association study. We also observed an attenuation in average p-value for hits when uncovering methylation quantitative trait loci (mQTL) in our dataset in comparison to analysing the raw 450k data. A possible explanation for this could be a reduced impact of measurement error at single probes when quantifying methylation across CpG islands. Furthermore, the mQTL analysis using the dimensionally reduced dataset required over 10 times fewer tests compared to the raw 450k data. Dimensionality reduction for large scale omics data is becoming increasingly important as different data types continue to grow in scale across multiple tissue types. The results of this study demonstrate the value of our approach in achieving a greater than 10 fold dimensionality reduction whilst still uncovering real biological effects.

2109F

Epigenetic signature of preterm-birth in adult twins. *Q. Tan, K. Christensen, L. Christiansen.* University of Southern Denmark, Odense , Denmark.

Preterm birth (PTB) represents a major problem for modern obstetrics due to its increasing frequency and the accompanying socioeconomic impact. As stress and depression are major causes of premature births, such conditions could alter the epigenome and thus affect long-term health in the victims. We performed a relatively large epigenome-wide association study (EWAS) on PTB using genome-wide DNA methylation data. The study samples consist of 73 pairs of indetical twins aged from 30 to 36 who have their gestational ages (33-42 weeks) recorded by The Danish Medical Birth Registry. Among them, there are 25 preterm birth pairs with gestational age<37 weeks and 48 term control pairs. Genome-wide DNA methylation data have been collected using the Illumina's Infinium HumanMethylation450 Beadchip assay. Our EWAS aimed at identifying differentially methylated genomic sites between preterm birth cases and term controls. We found 2 CpGs methylated for PTB with p value 4.7e-6 for cg09668047 on chromosome 16 and p value 4.84e-6 for cg00718694 on chromosome 2. Both CpGs are hypomethylated in preterm subjects. In addition, we also found 2 CpGs displaying sex-specific methylation patterns for PTB with p value 7.75e-7 for cg20459037 on chromosome 17 and p value 4.51e-6 for cg21146474 on chromosome 20. Interestingly, both are hypomethylated in males. The CpG site cg20459037 is linked to WDR16 gene on chromosome 17, a gene that plays crucial roles in a wide range of physiologic functions. Our results provide evidence for altered DNA methylation patterns observed in adult subjects in association with PTB.

2110W

Puberty-associated DNA methylation changes in females are near estrogen responsive genes and implicated in immune processes. *E.E. Thompson[1], J. Nicodemus-Johnson[1], K.W. Kim[1], J.E. Gern[2,3], D.J. Jackson[2], R.F. Lemanske[2,3], C. Ober[1].* 1) Department of Human Genetics, The University of Chicago, Chicago, IL; 2) Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI; 3) Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI.

The biological changes that occur following puberty in females are associated with increased risks for immune-mediated diseases, such as asthma and autoimmune diseases, both of which become more prevalent in females post-puberty. Many genetic and environmental factors are known to contribute to variation in the onset or delay of puberty, and recent evidence suggests a role for epigenetics in the timing of puberty as well. However, relatively little is known about specific epigenetic changes that arise during this window, or the genes that they potentially affect. To address this gap in knowledge, we studied global DNA methylation changes that occur during puberty in girls by focusing on peripheral blood mononuclear cells (PBMCs) collected from 38 girls pre- and post-puberty (8yrs and 14yrs, respectively), in whom puberty status was confirmed by Tanner staging. DNA from PBMCs collected from 43 boys at the same ages was used to identify and then exclude 715 age-related differentially methylated CpGs (DMCs). Of the remaining 326,556 probes on the Illumina 450K array that passed quality control, we identified 1,171 puberty related DMCs (5% FDR) in females that are in or near 1,036 unique genes. We first determined that genes near puberty-associated DMCs are enriched for estrogen responsive genes (P=3.2x10^{-6}; Fisher's Exact Test), suggesting that at least some of the effects of estrogen signaling in puberty are modified through epigenetic mechanisms. We then asked whether the genes near puberty-associated DMCs showed functional interactions by constructing protein-protein interaction analysis networks using Ingenuity Pathway Analysis (IPA). The closest genes to each of the 1,171 puberty-associated DMCs in females were connected in three significant networks (network score P<10^{-30}) that were enriched for genes involved in inflammatory response, inflammatory disease, and respiratory disease processes. Taken together, our results provide evidence for female puberty effects on global DNA methylation patterns at CpGs whose nearby genes are enriched for estrogen responsiveness and form networks centered on respiratory and inflammatory processes. These data further suggest that epigenetic changes that occur during puberty in females may contribute to sex-specific differences in susceptibility to immune-mediated diseases later in life. Supported by HL070831.

2111T

Effects of gestational alcohol alteration on the *Igf2/H19* **imprinted locus in a mouse model.** *H. Marjonen, N. Kaminen-Ahola.* Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki, Finland.

Alcohol is a known teratogen, and usage during pregnancy affects the developing fetus and causes life-long health effects. The severest end of the spectrum, fetal alcohol syndrome, includes clinical features such as morphological malformations, central nervous system impairment and growth restriction. The most widely studied epigenetic mechanism, DNA methylation, is known to mediate gene-environment interactions and have a crucial role during the embryonic development. To search the interactions between maternal alcohol consumption and epigenome we have focused to investigate the *Insulin-like growth factor 2 /H19* locus which is a well characterized imprinted region and acts as important regulator of both fetal and placental growth during pregnancy. As a material, we are using an inbred mouse model (C57BL/Rcc). The model is based on maternal consumption of 10% (v/v) ethanol for the first 8 days of gestation, from preimplantation period to the beginning of neurulation. This kind of moderate and chronic exposure period is developmentally equivalent to the first three-four weeks of human pregnancy. To investigate the alcohol-induced alterations in DNA methylation and gene expression of the *Igf2/H19* locus we have collected ethanol-exposed and control E9.5 and E16.5 embryos and placentas. We have explored possible alterations in DNA methylation of the locus by Sequenom EpiTyper method and in gene expression by quantitative PCR method. Our preliminary qPCR results show that alcohol causes downregulation of the *Igf2* expression in both E9.5 and E16.5 placentas. However we did not observe significant alcohol-induced alterations in DNA methylation in *Igf2/H19* imprinting control region or in *Igf2* differentially methylated region.

2112F

Genetic variation in MHC proteins is associated with T-cell receptor expression biases. *E. Sharon[1,2], L.V. Sibener[3,4,5], A. Battle[6], H.B. Fraser[2], K.C. Garcia[3,4,7], J.K. Pritchard[1,2,7].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Biology, Stanford University, Stanford, CA; 3) Department of Molecular Physiology, Stanford University School of Medicine, Stanford, CA; 4) Department of Structural Biology, Stanford University School of Medicine, Stanford, CA; 5) Program in Immunology, Stanford University, Stanford, CA; 6) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 7) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

In each individual, a highly diverse T-cell receptor (TCR) repertoire interacts with peptides presented by major histocompatibility complex (MHC) molecules. Despite extensive research, it remains controversial whether the germline-encoded TCR-MHC contacts promote TCR-MHC specificity and if so, whether there are differences in the compatibilities between different TCR V-genes and different MHC alleles. Here, we applied eQTL mapping to test for association between genetic variation and TCR V-gene usage in a large human cohort. We observed strong trans associations between genetic variation in the MHC locus and usage biases in TCR V-genes. Fine mapping of the association signals reveals particular amino acid residues in MHC genes that influence TCR V-gene usage. Remarkably, many of these residues are in direct contact or spatial proximity to either the TCR or the presented peptide in co-crystal structures. Our results show that MHC variants, several of which are linked to autoimmune diseases, can directly affect TCR-MHC interaction, and provide the first examples of *trans*-QTLs mediated by protein-protein interactions.

2113W

Toll-like receptor signaling pathway genes in association with trauma induced sepsis. *D.O. McDaniel[1], S. Martin-Robertson[1], L.S. McDaniel[2], G. Timberlake[1].* 1) Dept Surgery, Univ Mississippi Med Ctr, Jackson, MS; 2) Dept Microbiology and Immunology Univ Mississippi Med Ctr, Jackson, MS.

   Immunogenetic markers have been proposed as contributing factors in the development of trauma-induced clinical complications such as sepsis. Some patients with a similar degree of variables such as severity of injury, age, gender and preexisting clinical conditions develop post-trauma complications and die, but others recover. Such variation in the outcome might have association with the individual's genetic difference in immunoregulatory components of their innate immunity. The innate immune responses mediated by toll-like receptors (TLRs), induce early inflammatory response to the pathogen or damage associated molecular patterns (DAMPs) associated with trauma injury. A panel of molecular markers corresponding to TLR signalling-pathway genes were tested to evaluate trauma patients within the first week of admission. Blood was drawn at 24 hr time intervals up to one week from patients that met the inclusion criteria. Leukocytes were utilized in mRNA transcript analysis for TLR signalling-pathway genes using a gene-array. TLR-1, -4, -8 and -10 were almost equally expressed at day 1 in all patients. But, TLR-2 was 2.7 fold increased comparing day 1 vs. day 3 in monocytes of patients who later developed sepsis. TOLLIP, a toll inhibitory protein, also known as Cox-2 gene was 4-fold increase in patient with no sepsis. The IRAK-1 (IL-1 associated kinase-1), a major mediator of the TLR signalling-pathway was 2.5-fold increased in patients with no sepsis and reduced in patients with sepsis. The HMGB-1 a DAMP associated molecule was increased 5.7-fold, and CD86 an HLA-Class II receptor molecule was 15.2-fold increased in patients who later developed sepsis. In conclusion, profiling the innate immune response signalling-genes was informative and correlated with clinical outcomes which may help to predict the development of sepsis.

2114T

Chromatin state variability: A guide to uncover non-coding functional genomic regions and potential regulated genes. *L. Pinello[1], A. Gusev[2], H. Finucane[2], J. Huang[1], A. Price[2], G.C. Yuan[1].* 1) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; 2) Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA.

   With the increasing amount of epigenomic data, a pressing challenge is to understand the mechanisms underlying chromatin states changes and their role in cell-type specific establishment and maintenance. Here we propose a computational method based on information theoretic approaches to systematically quantify the variability of chromatin states and study its association with gene expression variation and non-coding variants from GWAS and eQTL studies.   Using histone modification data from 9 human cell lines from the ENCODE project we identified Highly Plastic Regions (HPRs) for chromatin state variation. HPRs cover about 2% of the genome and correspond to regions with the highest Shannon's entropy calculated using the frequency of chromatin states across cell types. We found that HPRs are enriched for many important regulatory elements such as super-enhancers, promoters and Polycomb repressed regions. Moreover we find that the HPRs are highly enriched for GWAS-associated (1.5x, p<1E-4 permutation test) and eQTL non-coding variants (1.4x, p<1E-4 permutation test), and for some traits or diseases, provide significant additional power in explaining heritability than well-characterized elements (14.7x, p<1E-20, non-overlapping partition heritability analysis excluding coding, promoters, DHS and H3k27ac).   We also identify regions of co-variability based on histone modification data, and show that such regions can predict high confidence long-range interaction obtained by Hi-C and ChIA-PET assays (AUC = 0.87) and also correlate well with observed interactions counts (Pearson's correlation = 0.54). In addition, the average co-variability profile around topological associated domains (TAD) boundaries, qualitatively shows a precise depletion in correspondence with the boundary and more enrichment in the flanking regions. This pattern is in agreement with the chromosome neighborhoods defined by the TADs, and suggesting that this measure could be helpful to highlight or predict functional subdomains within the TAD.   Our analysis provides new insights into the organization of cell-type specific chromatin structure during development, and a valuable tool for investigating the mechanisms of chromatin state establishment and usage in regions subjected to dynamic changes important in gene regulation. The HPRs obtained from our analysis provide also an important guide to search for functionally important genetic variants and linked genes. .

2115F

Analysis of monoallelic expression in 1084 human single-cell fibroblasts revealed novel putative imprinted genes including *ADTRP, ATP5EP2,* *SMOC1, ERAP2* **and** *PRIM2.* C. Borel[1], F. Santoni[1], E. Falconnet[1], P. Ribaux[1], S.E. Antonarakis[1,2,3]. 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) University Hospitals of Geneva, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland.

Genomic imprinting is defined as the mutually exclusive expression of either the paternally or maternally inherited allele. Imprinted genes are implicated in the etiology of rare syndromes and have been associated with common diseases such as diabetes and cancer. Less than 100 human imprinted genes have been identified so far, and the predicted number is in the range of a few hundreds. The aim of the project is to identify novel imprinting genes with single-sell (SC) RNA sequencing. This approach has the potential to improve the detection of average to low expressed genes and to provide a comprehensive profiling of the allelic imbalance of each gene. Using samples from the probands of 2 family trio and 3 unrelated individuals, 1084 individual primary fibroblasts were RNA sequenced and more than 770'000 informative heterozygous SNVs were genotyped (WGS, 25 x depth). For each gene, we analyzed the allelic specific expression (ASE) through an in-house pipeline (read mapping with Tophat, mapping quality=50). For each individual we modeled the likelihood of a gene to be monoallelically expressed with a beta-binomial distribution and evaluated the significance of the aggregate monoallelic ratio (reads sum of the most frequent allele per site / total reads) with the log-likelihood test. Fifty two genes were retained (adj. p-value < 0.01, ASE: 0.8-1 in each individual where the gene is detectable) for the identification of the allelic parent of origin in the two trios. We were able to identify 10 novel putative imprinted genes and 8 known imprinted genes. We carefully analyzed the imprinting status of 93 known imprinted genes in our SC dataset. Among the 50 detected genes in fibroblasts, we validated the imprinting status of 24 genes. Remarkably, 26 known imprinting genes show evidence of individual variability or biallelic expression in fibroblasts questioning their imprinting status in all the tissues. To elucidate the advantages of SC RNAseq for imprinted gene finding, we investigated the ASE of the 52 genes in 201 primary fibroblast cell lines from the GenCord collection. When detected in bulk samples, novel putative imprinting genes are validated but their statistical significance is mostly insufficient for imprinting discovery. Overall, the SC approach revealed novel putative imprinting genes that were undetectable in the respective bulk samples. C.B. and F.S. contributed equally.

2116W

Altered imprinting in miscarried products of conception. C. Demetriou[1], A. Thomas[1], M. Ishida[1], S. Abu-Amero[1], S. Stanier[1], L. Regan[2], P. Stanier[1], G. Moore[1]. 1) Genetics and Genomic Medicine, Institute of Child Health, UCL, London, United Kingdom; 2) Department of Obstetrics and Gynaecology, St Mary's Campus, Imperial College London, London, UK.

Fetal growth involves a complex interaction between genes and the environment. Genomic imprinting is an epigenetic form of gene regulation that gives rise to differential expression of a given allele, depending upon its parental inherited origin. A study was undertaken to investigate imprinting anomalies in a 100 products of conception (POC), collected from miscarried pregnancies around the 10th week of gestation. With a hypothesis that a proportion of miscarried POC that are cytogenetically normal will have abnormal expression profiles at imprinted loci, possibly contributing to embryonic lethality, we investigated the imprinting status of the placental paternally expressed genes *IGF2, PEG3* and *DLK1* and the maternally expressed genes *H19, GRB10* and *PHLDA2*. This work first involved a detailed analysis of the POC DNA to establish levels of maternal cell contamination (MCC). Surprisingly, only 7 POC were free of MCC and 16 were partially contaminated, while 74% of the cohort was exclusively maternal decidua. The 23 POC with fetal material were tested for heterozygosity for each gene to identify informative samples. *IGF2* is a well-known paternally expressed gene. All 11 informative POC samples for *IGF2* were monoallelically expressed; however, 4 of these were maternally expressed. Bisulfite sequencing to test methylation status revealed normal methylation profiles. Sequencing of the *H19* imprinted control region and the CTCF binding factor revealed no mutations. Although paternal imprinting of *IGF2* is seen in brain tissue, to our knowledge, this is the first time maternal *IGF2* expression has been observed in human placental tissue. The other imprinted genes tested were expressed as would normally be expected. *HLA-C* was also tested and was found to be biallelically expressed in fetal, placental and maternal decidua tissue. Interestingly, although *HLA-C* was biallelically expressed in chorionic villi collected at 12 weeks of gestation from successful pregnancies, it was monoallelically expressed in all POC, showing either maternal or paternal expression, thus suggesting random allele switching off of *HLA-C* in the miscarried tissue. It is therefore possible that a proportion of miscarriages can be explained by gene deregulation at imprinted loci which lead to embryonic lethality. Genome-wide epigenetic remodeling occurs during the early stages of development and if disrupted may lead to aberrant gene expression with consequences to maintenance of a healthy pregnancy. .

2117T

Investigating DNA methylation as a marker for historical smoke exposure and a mediator of disease risk. *R.C. Richmond, M. Suderman, P. Haycock, G. Hemani, C. Relton, G. Davey Smith.* MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

Cigarette smoke is a powerful modifier of DNA methylation and epigenome-wide changes in response to smoke exposure in-utero have been established. We previously conducted work showing the persistence of some of these methylation marks through childhood and into adolescence. Of interest is whether signals persist into adulthood. We performed an analysis to investigate associations between maternal smoking in pregnancy and DNA methylation in peripheral blood among women in the Avon Longitudinal Study of Parents and Children. We found that, among women with a mean age of 29 years (N=866), there was an inflated signal above that expected by chance at CpG sites previously associated with in-utero smoke exposure in newborns (lambda=2.80 vs. 1.05 for all CpG sites on the Illumina Infinium HM450 array). In-utero smoke exposure was associated with hypermethylation at cg22132788, cg12803068 and cg04180046, located in the *MYO1G* gene region, at Bonferroni significance. These signals remained in sensitivity analyses adjusted for own smoking and when restricted to non-smokers. The same signals were observed in peripheral blood when the women were a mean age of 47 years (N=885). We replicated findings of a persistent methylation signal related to in-utero smoke exposure in whole blood and LCLs from participants in the 1958 Birth Cohort (mean age=45 years; N=40) and in saliva from participants in the Aberdeen Folic Acid Supplementation Trial (mean age=47 years; N=83). Persistent changes in DNA methylation might mediate some of the associations between in-utero smoke exposure and later life health. However, disentangling mediation from other mechanisms that lead to association -such as confounding and reverse causation- warrants further evaluation. We investigated whether persistent methylation at *MYO1G* plays a causal role in the aetiology of disease using a hypothesis-free Mendelian randomization approach. We first identified cis-SNPs associated with methylation at *MYO1G* and then performed an agnostic look-up of these SNPs in GWAS summary data using the "MRBase" platform (http://www.mrbase.org/). We found enrichment for a causal effect of *MYO1G* methylation on a range of metabolites. These findings could have useful applications in epidemiological studies, by using DNA methylation as a refined indicator and biosocial archive for historical exposure. The downstream health implications of methylation change in response to smoke exposure in-utero requires further evaluation.

2118F

DNA methylation gives insight into different etiologies in early-onset and late-onset preeclampsia and intrauterine growth restriction. *S.L. Wilson[1,2], K. Leavey[3], B. Cox[3,4], W.P. Robinson[1,2].* 1) Healthy Starts, Child and Family Research Institute, Vancouver, BC, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 3) Department of Physiology, University of Toronto, Toronto, ON, Canada; 4) Department of Obstetrics and Gynecology, University of Toronto, Toronto,ON, Canada.

Preeclampsia (PE), a maternal hypertensive disorder, and intrauterine growth restriction (IUGR), a pathologically small fetus, often co-occur and are both associated with an inadequately functioning placenta. Placental dysfunction is more common in early-onset PE (EOPE), diagnosis <34 weeks gestation, than late-onset PE (LOPE). However, the relationship between these disorders remains unclear. Placental DNA methylation (DNAm) profiles of EOPE, LOPE, and normotensive IUGR (nIUGR) placentas were compared to healthy controls to better understand their etiology and relationships to one another. Placental DNA from both a discovery (43 controls, 22 EOPE, 18 LOPE, and 11 nIUGR) and a validation cohort (18 controls, 19 EOPE, 11 LOPE) were run on the Illumina HumanMethylation450 array. Each cohort was normalized (functional normalization) and batch-corrected (ComBat). To minimize gestational age (GA) effects, EOPE samples were compared to pre-term control samples (GA<37 weeks), while LOPE and nIUGR were compared to term controls (GA>37 weeks). Linear regression, correcting for GA and sex was used to identify differentially methylated (DM) sites (FDR<0.05, Δβ>0.1). Within the discovery cohort, 2115 sites were DM in EOPE placentas, however no sites were DM for the LOPE or nIUGR groups. At a nominal p-value<0.01, EOPE DNAm changes were enriched for GO terms: regulation of inflammatory response, hormone transport/secretion, mesoderm development, and cell aging using ErmineJ. As we identified no DM sites in LOPE or nIUGR, we focused on validating the EOPE hits. In the validation cohort, 240 DM sites were identified for EOPE. Changes in DNAm at EOPE DM sites were correlated between the discovery cohort and the validation cohort (R=0.62, p<2.2e-16) with 69 sites meeting FDR<0.05 and Δβ>0.1 in both cohorts. These persistent hits include CpGs associated with previously reported PE-associated genes such as: *PAPPA2*, *INHBA*, and *JUNB*. Clustering on these persistent hits largely separates EOPE from controls, while some LOPE samples cluster with both EOPE and controls. Many DNAm changes were observed in EOPE and a significant portion were validated in a separate cohort. The persistent hits provide the best candidates for biomarker development for improved diagnosis. In contrast, we were unable to identify DNAm changes specific to LOPE or nIUGR. This may be due to a less severe placental phenotype, limited sample size, and/or increased disease heterogeneity as compared to EOPE.

2119W

Circulating microRNAs and prediction of asthma exacerbation in the Childhood Asthma Management Program (CAMP) Cohort. *J.S. Davis[1,2], A.T Kho[2,3], K.G. Moore[1], J.M. Sylvia[1], S.T. Weiss[1,2], K.G. Tantisira[1,2].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA.

   Identification of children at the highest risk for asthma exacerbations could result in decreased costs, tailored care and improved outcomes. Prior studies have yielded variable results in predicting exacerbations. Circulating microR-NAs (miRNA) are attractive biomarkers that may have predictive power for asthma exacerbations. This study examines circulating miRNA and asthma exacerbations in the 12 months following randomization to the inhaled corticosteroid arm in a subset of 153 subjects in the Childhood Asthma Management Program (CAMP) cohort. The study hypothesizes that a miRNA model or combined miRNA-clinical model may have superior predictive capability compared to a clinical prediction score for childhood asthma exacerbations. Serum samples obtained at randomization from the subjects were profiled using a TaqMan miRNA array set containing 754 miRNAs. Logistic regression was performed with dichotomized steroid burst count as the binary response variable and miRNA cycle threshold and a clinical asthma exacerbation prediction score as the explanatory variables. Stepwise feature selection was performed, and predictive capability was assessed. There were 125 (16.6 %) detected in at least 70 % of the samples with 12 significant miRNA in the univariate models with the strongest predictor being miR-206 with an OR 0.60. All miRNA demonstrated odd ratios < 1 indicating decrease risk of asthma exacerbation with increase in miRNA cycle threshold. Stepwise selection yielded a three miRNA model including miR-206, -146b, and -720. The clinical score model had a positive coefficient with resultant OR greater than 1 indicating an increased risk of asthma exacerbation with increase in clinical score. For testing/training sets, the combined miRNA-clinical score demonstrated good predictive power with an AUC of 0.807, which was significantly better than either clinical score model (0.671) or miRNA model (0.714). In conclusion, this study examined circulating serum miRNA in childhood asthmatics prior to treatment with inhaled corticosteroids to predict exacerbations in the subsequent year and found 12 miRNAs significantly associated with future exacerbations with each doubling of these miRNA associated with a 25-67 % increase in risk of exacerbation. A three miRNA model provided comparable predictive power to an established clinical model of exacerbations. Furthermore, a combined miRNA-clinical model had significant increase in the ability to predict future exacerbations.

2120T

MicroRNAs and their regulatory interactions in the human hair follicle. *S. Heilmann-Heimbach[1,2], L.M. Hochfeld[1,2], T. Anhalt[1,2], N. Fricker[1,2], A. Hofmann[1,2], M.M. Nöthen[1,2].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics; Life&Brain Center, University of Bonn, Bonn, Germany.

   Human hair follicle (HF) cycling involves numerous signaling cascades and cellular processes that have been reported to be controlled by micro(mi)RNAs. However, our current knowledge on the expression and function of these miRNAs in hair biology is largely based on murine data or cell culture experiments and little is known about their exact role in the human HF. We therefore sought to gain a deeper understanding of the role of 9 previously reported miRNAs in human hair biology by systematically investigating their expression and regulatory interactions in HF samples from 25 healthy male donors. After nucleic acid extraction, we performed a genome-wide miRNA and mRNA expression profiling and subsequently correlated miRNA expression levels with mRNA-expression levels from the same samples to identify potential target genes. Moreover, we tested for an enrichment of target genes in biological pathways and searched for protein-protein interactions (PPIs). Our analysis confirmed the expression of 7 of the 9 candidate miRNAs. For 3 of them (miR-24, miR-31, miR-106a), significant miRNA/mRNA correlations were detected. The identified target genes included genes with known functions in hair biology (e.g. *ITGB1*, *SOX9*), genes where the biological function of the encoded protein reflected the reported miRNA-function (e.g. *RXRA*) and numerous genes that had not previously been implicated in hair biology. A subsequent pathway- and PPI-based analysis revealed an enrichment of target genes in relevant pathways such as integrin and GnRH signaling and confirmed the interaction of target gene products in PPI-networks. In summary, our data provide novel insights into the role of miRNAs and their regulatory interactions in the human HF and will eventually contribute to a deeper understanding of healthy hair growth and the pathobiology of hair loss disorders.

2121F

Role of miR-133b in the regulation of *Col1A1***,** *TGF-β1* **and** *ACTA2* **in scleral fibroblasts.** *R. Metlapally[1], A.V. Tang[1], K.K. Wang[1], H.M. Do[1], C.F. Wildsoet[1,2].* 1) School of Optometry University of California at Berkeley; 2) Vision Science Graduate Group University of California at Berkeley.

   Purpose: The sclera, outermost white layer of the eye, comprises a collagen dominated extracellular matrix that undergoes active remodeling to maintain eye size and shape during early development. In myopia (near-sightedness), the most common eye disorder, increased remodeling leads to scleral thinning and excessive eye enlargement, increasing the risk for irreversible ocular complications. micro(mi)-RNAs have been implicated in scleral matrix remodeling during both normal ocular growth and myopia development. In this study, we investigated the role of miR-133b in the regulation of *Col1A1* (the predominant scleral collagen subtype), *TGF-β1* and *ACTA2* (alpha-smooth muscle actin), scleral genes linked to myopia as well as implicated in miR-133b driven matrix remodeling in wound healing. The long-term goal of this research is to target scleral miRNAs for developing novel therapies to retard myopia progression. **Methods:** Human scleral fibroblasts from donor eyes were cultured using standard methods. At 80% confluence, they were transfected with miRVANA™ miR-133b mimic or negative scramble control (30 nM and 100 nM, n=4), using electroporation. After 48 h, cells were lysed and reverse transcriptions performed using Ambion® Cells to CT kit, and the expression levels of *Col1A1*, *TGF-β1*, and *ACTA2* were quantified using Taqman™ QPCR assays. Analyses were performed using the deltaCt method; all data were normalized to the expression of snoRNAs RNU44 and RNU6B and negative controls, and fold changes calculated. Calcein-AM ethidium homodimer (EthD-1) cell viability assay was performed to rule out cytotoxicity due to the treatment. **Results:** As expected, transfection of scleral fibroblasts with miR-133b mimic resulted in increased expression of miR-133b for both concentrations tested (> 300-fold, p < 0.05), and decreased expression of *Col1A1* gene expression (~2-fold, p < 0.01). There were no significant changes in the expression of *TGF-β1* and *ACTA2* genes following miR-133b mimic treatment. Tests for cytotoxic effects of the applied miRNA mimic yielded negative results. **Conclusion:** miRNA expression can be predictably manipulated in scleral fibroblast cultures. Our findings implicate miR-133b in the regulation of *Col1A1*, but not *TGF-β1* and *ACTA2*, in scleral fibroblasts. These observations could prove useful in devising myopia-control therapies aimed at preventing or reversing the scleral changes seen in myopia.

2122W

Sex-specific effects of testosterone on the sexually dimorphic transcriptome and epigenome of embryonic neural stem/progenitor cells. *M. Bramble, L. Roach, A. Lipson, N. Vashist, A. Eskin, E. Vilain.* Human Genetics, UCLA, Los Angeles, CA.

   The direct effects of gonadal hormones and chromosomal sex on gene regulation in undifferentiated neural stem cells, a cellular model for brain sexual differentiation, are poorly understood. We isolated embryonic neural stem cells (NSCs) from XX and XY 13.5-14 C57/BL6/J mice to investigate gene expression differences and epigenetic changes in multipotent neural stem cell populations, prior to, and post testosterone exposure. Transcriptionally, Using RNA sequencing we identified novel sex differences in gene expression in NSCs that have not had prior exposure to endogenous gonadal hormones. Additionally, we identified unique sex-specific effects in gene expression post exogenous T exposure. We found 103 genes that are differentially expressed between XY and XX murine NSCs (FDR=0.05). After treatment with T, we found that in XX cells, 2854 genes were differentially expressed, compared to XY cells, which showed 792 transcripts differentially expressed (FDR 0.05). These findings indicate that testosterone exposure on XX cells may have a more robust effect with regards to altering gene expression than on XY cells. Most interestingly, it was found that by exposing XX NSCs to T, 42% of the original 103 basal sex differences that existed were masculinized and shifted towards a XY typical gene expression pattern. Unexpectedly, we also determined that 25% of basal sex differences were actually feminized in an XY background post T. This suggests modular effects of T exposure on sexual development. In addition to determining how T alters gene expression in NSCs we also investigated how testosterone perturbed the epigenome of our cell population. We found that testosterone is capable of causing global hypo-methylation events in neural stem cells, but most interestingly these effects are maintained over cellular linages even in the absence of testosterone. Our novel approach is the first to identify sex differences in gene expression in NSCs, and demonstrate the effects of testosterone exposure on gene expression. Additionally, we have identified that testosterone has the ability to alter the epigenome within a stem cell population and elicit changes that are maintained in downstream cellular lineages. Collectively these data have identified a new role for testosterone's influence on the developing CNS in addition to possibly unlocking key mechanistic changes that are responsible for permanent hormonal organization of the sexually dimorphic mammalian brain.

2123T

The landscape of regulatory post-transcriptionally derived small non-coding RNAs in the human transcriptome. *Y.Y. Leung[1,2], P.P. Kuksa[1,2], A. Amlie-Wolf[1,2,3], L.-S. Wang[1,2,3].* 1) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Mature small RNA products resulting from post-transcriptional processing of precursor transcripts play major roles in gene regulation by binding to protein and mRNA targets. While these small noncoding RNAs (e.g., miRNAs and tRNA fragments) are emerging as biomarkers and regulators in human diseases including cancer and neurodegenerative disorders, the genome-wide landscape of functional mature small noncoding RNAs (sncRNAs) remains poorly characterized. To delineate the genome-wide post-transcriptional processing and expression of sncRNAs, we performed the largest study to date by curating and integrating all short total and small RNA-seq data from 36 human cell lines with 72 samples from ENCODE and 42 human cell types and tissues with 198 samples from DASHR (Database of small human noncoding RNA), respectively. We applied *ab initio* annotation-free methodology, SPAR, to these datasets to systematically characterize the extent of sncRNA (15-44nt) expression across tissues and cell lines. We found 150,997 and 121,466 specifically processed sncRNAs expressed in at least 3 samples in ENCODE and DASHR. By expression, these sncRNAs are enriched for RNA length classes 20-24nt (66%), 15-17nt (10%), 32-36nt (4%). Of these expressed sncRNAs, 70% sncRNA genes were previously uncharacterized. Both annotated and uncharacterized sncRNA loci are conserved (mean 100-way conservation score = 0.3-0.4 with top quartile score >0.7-0.9), indicating that these previously unannotated loci may also be functional. 3,146 out of all loci are ubiquitously expressed suggesting that they may perform fundamental regulatory roles, whereas 90% are tissue-regulated. We integrated all results into our DASHR database (http://lisanwanglab.org/DASHR). Of the annotated mature product loci in the top 10% of expression, 39% are miRNAs. 70% of miRNAs display average expression or more. Surprisingly, the expression of 61% of piRNA loci are above average, but this pattern was not observed in snoRNAs and tRNAs. SPAR has been implemented as a web server where users can quickly view their own smRNA-seq analysis results and compare them with ENCODE and DASHR and as a standalone tool for offline batch analysis. These tools serve as a valuable resource for biomarker development and characterization of small noncoding RNAs. Together, SPAR and DASHR provide novel insights into the landscape of the small RNA processing and expression landscape across human tissues and cell types.

2124F

A role for YY1 in sex-biased transcription revealed through X-linked promoter activity and allelic binding analyses. *C. Chen[1,2], W. Shi[1,2], A. Matthews[1], Y. Li[1], D. Arenillas[1], A. Mathelier[1], C. Brown[3], W. Wasserman[3], FANTOM Consortium.* 1) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada; 2) Graduate Program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

Sex differences in susceptibility and progression have been reported in numerous diseases. Female cells have two copies of the X chromosome with X-chromosome inactivation imparting mono-allelic gene silencing for dosage compensation. Such differences in transcriptional status between the copies in female and copy numbers between sexes pose a challenge to genomic data analyses of the X. A subset of genes, named escapees, escape silencing and are transcribed bi-allelically resulting in sexual dimorphism. Here we conducted analyses of the sexes using human datasets to gain perspectives in such regulation. We first identified transcription start sites of escapees (escTSSs) based on higher transcription levels in female cells using FANTOM5 Cap Analysis of Gene Expression data. Greater similarity of DNA methylation levels between the sexes was found to be consistent with bi-allelic activity at these escTSSs. The significant over-representations of YY1 transcription factor binding motif and ChIP-seq peaks around escTSSs highlighted its positive association with escapees. Furthermore, YY1 occupancy is significantly biased towards the inactive X (Xi) at long non-coding RNA loci that are frequent contacts of previously reported Xi-specific superloops in female GM12878 cells. While revealing the unique properties of the X as reflected by genomic datasets, our study elucidated the importance of YY1 on transcriptional activity on Xi in general through sequence-specific binding, and its involvement at anchor regions of Xi-specific chromatin superloops.

2125W

Identification of lncRNAs involved in smooth muscle contraction pathway in preterm birth. *J. Wang[1,2], Z.H. Tu[1], X.L. Zhao[2,3], W. Ju[4], N. Zhong[2,3,4].* 1) Laboratory Medicine, Hainan Provincial Maternal and Children's Hospital, Haikou, Hainan, China; 2) China Alliance of Translational Medicine for Maternal and Children's Health, Beijing, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA.

Preterm birth (PTB) is the most common etiological factor causing neonatal mortality and the second most common factor for the death of children under five. It is also one of the common causes for the lifetime morbidity. Each year, there are 15 millions of PTB neonates delivered globally. Clinically, spontaneous PTB (sPTB) accounts for about 80% of PTB and may present as preterm labor (PTL) or preterm premature rupture of membrane (pPROM). The PTL shows uterine contraction without rupture of fetal membrane. Many studies have determined that infection and inflammation during, or prior to, the pregnant period may trigger the labor. The underlying mechanism, however, is yet unknown. We have conducted exome sequencing on 160 sPTL and obtained 147 clean data. SNPs, SNVs, InDels, can be identified at exons, introns, UTR3 and UTR5, and ncRNA regions. On the basis of exome data, we selectively focused on smooth muscle contraction (SMC) pathways, with targeting lncRNA that associated with the gene loci that encode proteins involved in SMC pathway. Microarray of genome wide lncRNAs, along with overlapped mRNAs, were analyzed and results were validated with quantitative RT-PCR. The lncRNA-AL590431.2 overlapped with tropomysin (TPM3), an actin-binding protein, was found down-regulated in placentas of sPTL, however, the lncRNA-AL590431.2 overlapped mRNA-TPM3 was up-regulated. Longitudinally, expression of lncRNA-AL590431.2 in maternal bloods were significantly in the second and third trimesters, compared to the first trimester among normal pregnancies. However, there was no significant alteration of the lncRNA-AL590431.2 between the second and third trimesters, compared to the first trimester among the placentas of sPTL and pPROM. The pattern of longitudinal expression of lncRNA/mRNA of Desmin, a muscle-specific class III intermediate filament, showed a similarity to the lncRNA-AL590431.2/mRNA-TPM3. Our studies demonstrated that epigenetic the regulation of lncRNA not only are involved in placentas but also present in maternal circulations.

2126T

Inter-individual variation in microRNA regulatory network. *T. Annilo, M. Kals, R. Mägi, T. Esko, A. Metspalu.* Estonian Genome Center, University of Tartu, Riia 23b, 51010 Tartu, Estonia.

The dynamic nature of evolution of miRNA gene families and their targets has been demonstrated by comparative genomics studies in a wide range of species. Genetic variation can affect all aspects of miRNA functionality, including transcription, biogenesis and binding to the target. We present a study to assess the miRNA function using naturally occurring genetic variation in the context of individual whole-genome sequences (WGS). We have combined WGS (N=2244) and ample phenotypic data collected at the Estonian Genome Center for identification of the target modulation effect caused by genetic variation in miRNA networks. Such approach allows assessment of miRNA function in the defined individual genetic and phenotypic background. Genomic coordinates for miRNA genes and flanking regions were obtained from UCSC Genome Browser and miRBase v21. Sequence variants in miRNA genes were grouped into categories: (i) seed region, (ii) mature miRNA sequence, (iii) "passenger" strand, (iv) loop region, and (v) flanking regions of 100 nucleotides. The functional impact of genetic variants on miRNA binding was estimated using PolymiRTS v3.0, miRSNP, and mrSNP databases. Experimentally validated miRNA target sites were obtained from miRTarBase v4.5, TarBase v6.0, and miRecords v4. Using this approach, we have identified more than 1000 different genetic variants in miRNA genes, approximately 60% of them novel with median allele frequency 0.14%. More than 50% of miRNA genes contained one or more genetic variant. We are currently testing the effect of these variants on mRNA expression using matched RNA-seq data (N=600).

2127F

HipSci data resources: High quality genomic and cell biological data for a reference panel of human iPSC lines. *Y. Memari[1], I. Streeter[2], R. Halai[1], R. Nelson[1], R. Durbin[1] on behalf of the HipSci Consortium (http://www.hipsci.org/).* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK; 2) EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

The Human Induced Pluripotent Stem Cells Initiative (HipSci) has developed a pipeline to derive, expand, assay and bank iPS cell lines on a production scale. This collection of well-characterised iPSCs is a valuable resource for studies of cell biology and clinical genetics, and it is accompanied by systematic data sets that can be accessed through public archives and web sites. To date HipSci has derived iPSC lines from 282 healthy individuals and 26 individuals with monogenic disease. By 2017 we aim to increase these numbers to over 450 healthy donor and 200 patient donor iPS lines. To generate the iPSCs, primary tissue is reprogrammed using a Sendai viral vector system. From each donor, 2 candidate iPS cell lines are derived, each of which is subjected to extensive quality checking including quantification of pluripotency using gene expression microarray data and the software PluriTest, and screening for genetic stability and chromosomal abnormalities using genotyping and a newly developed CNV calling software[1]. Successfully reprogrammed lines are banked for distribution and future use, and further studied to identify major drivers of molecular and cellular heterogeneity in iPSCs. Further HipSci project assays include profiling with methylation array, whole-exome and RNA sequencing, proteomic mass spectrometry and cellular phenotyping. Selected lines are banked at ECACC and EBiSC and the entire quality control and subsequent assay data are deposited to open-access (ArrayExpress, EVA, PRIDE) or managed-access (EGA) repositories, with links to these available via the project coordination web site www.hipsci.org. Initial results from the project are available in a biorxiv preprint[2]. We wish to encourage wide use both of the data sets and cell lines. This poster will provide information about the data processing workflow and data resources available for HipSci, with a summary of major results and up to date numbers for data and sample availability. 1) A Method for Checking Genomic Integrity in Cultured Cell Lines from SNP Genotyping Data, Danecek et al. PLoS ONE, Vol. 11, No. 5. (13 May 2016), e0155014, doi:10.1371/journal.pone.0155014 2) Common genetic variation drives molecular heterogeneity in human iPSCs, Kilpinen et al. biorxiv.org, May 25, 2016. doi: 10.1101/055160.

2128W

Genetic factors and non-genetic mechanisms both contribute to iPSC methylation variation and aberrancy. *E.N. Smith[1], A. Panopoulos[2,3], A.D. Arias[1], P.J. Shepard[4,5], Y. Hishida[2], V. Modesto[6], K. Diffenderfer[6], C. Conner[2], W. Biggs[7], E. Sandoval[7], A. D'Antonio-Chronowska[9], W.T. Berggren[6], J.C. Izpisua Belmonte[2], K.A. Frazer[1,4,8,9].* 1) Department of Pediatrics and Rady Children's Hospital, University of CA, San Diego, 92093, CA; 2) Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA; 3) Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA; 4) Moores Cancer Center, University of California at San Diego, La Jolla, CA, USA; 5) BioSpyder Technologies, Inc. Carlsbad, CA 92010; 6) Stem Cell Core, Salk Institute for Biological Studies, La Jolla, CA, USA; 7) Human Longevity, Inc. San Diego, CA, USA; 8) Bioinformatics and Systems Biology Program, University of California at San Diego, La Jolla, CA, USA; 9) Institute for Genomic Medicine, University of California at San Diego, La Jolla, CA, USA.

Since induced pluripotent stem cells (iPSCs) were first reported, considerable effort has focused on examining the molecular consequences that follow this type of cellular conversion. Several studies have reported that some CpGs differ in methylation status between individual iPSC lines (methylation variation) as well as that some CpGs differ in methylation status between iPSCs and embryonic stem cells (ESCs) (methylation aberrancy). Because iPSCs and ESCs differ in their genetic background, it is currently unclear how much of the perceived methylation variation and aberrancy can be contributed to genetic variants and how much is due to other variables. Here, to examine the roles of genetic background, clone lines and passage in methylation variation and aberrancy at CpG sites in the iPSC lines, we generated 22 iPSC clonal lines from six individuals (3 pairs of older monozygotic twins), and profiled early (passages 5 (p5) and 9 (p9)) and late (passage 20 (p20)) passages as well as fibroblasts (tissue of origin) using genome-wide methylation arrays. Because the vast majority of genetic variants are not assayed by genotyping arrays, we performed whole genome sequencing (WGS) to test whether genetic variants near the CpG could explain genetic background effects. We show that divergent methylation patterns in twin fibroblasts are reset after reprogramming based on genetic background, with samples derived from twins more similar to each other. Using ANOVA analyses, we show that methylation at CpGs is most associated with genetic background rather than clone or passage alone. Genetic-background associated sites are more likely to be close to genetic variants identified through WGS and to be affected by established methylation quantitative trait loci (meQTL), suggesting that CpGs are more similar between twins due to genetic variation. We further show that aberrantly methylated CpGs are also enriched for being near genetic variants and associated with meQTLs, suggesting that a significant proportion of aberrant sites are due to genetic variation among lines. However, methylation aberrancy can also be strongly associated with clone-specific effects independent of genetics. Of note, aberrant sites are enriched in specific chromatin states and recurrent across studies. Our study shows that while genetic variation regulates most genome-wide methylation levels, there appears to also be a non-genetic chromatin-based mechanism underlying some variant and aberrant sites.

2129T

Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *J. Ito[1,2], S. Yamada[1], R. Sugimoto[1], T. Kimura[1], H. Nakaoka[1], I. Inoue[1,2].* 1) Human genetics, National Institute of Genetics, Mishima, Japan; 2) School of Life Science, The graduate university for advanced studies (SOKENDAI), Hayama, Japan.

Human genome contains numerous copies of HERVs (human endogenous retroviruses) harboring their own transcriptional regulatory elements. These HERV regulatory elements are important for evolution of host regulatory networks because these elements are shared among numerous copies of HERVs and can affect not only a single gene but also many genes as a coordinate manner. We systematically identified and characterized HERV regulatory elements based on ChIP-Seq datasets produced by *ENCODE* and *Roadmap Epigenomics*. We particularly focused on regulatory elements that originally existed in HERVs before their integrations, termed HERV conserved regulatory elements (HCREs). Total 1,300 types and 85,000 sites of HCREs were identified from 300 HERV types. HERVs frequently have HCREs working in pluripotent cells, endoderm linages, and hematopoietic cells, besides CTCF binding sites. Some HERV types such as LTR5_Hs and LTR7 have subtype-specific HCREs, indicating regulatory elements of them were dynamically changed during genomic expansions of them. Genomic positions of some kinds of HCREs were significantly enriched in regions around genes associated with specific biological processes such as innate immune response, and furthermore, three-dimensional interactions between those types of HCREs and promoters of such genes were confirmed by using chromatin-interactome data generated by Hi-C method, indicating those types of HCREs cooperatively reinforced or modified regulatory networks concerned with such biological processes. Overall, we identified various HCREs that might have been involved in evolution of host regulatory networks. This study provides not only fundamental information to understand impacts of HERVs on host transcriptions but also new insights into transcriptional modulation systems that HERVs or their ancestral retroviruses originally had.

2130F

Environmentally-induced epigenetic variability is associated with metabolic traits. *E. Carnero-Montoro[1], A. Visconti[1], C. Sacco[1,2], P.C. Tsai[1], T.D. Spector[1], M. Falchi[1], J.T. Bell[1].* 1) Department of Twin Research, King's College London, London, London, United Kingdom; 2) Department of Statistical Sciences, University of Bologna, Italy.

Epigenetic modifications have been proposed to play a key role in mediating individuals' phenotypic responses to environmental exposures. Recent epigenome-wide association studies (EWAS) confirm that environmental exposures, such as smoking, are strongly associated with DNA methylation levels in human populations. To date EWAS have predominantly focused on identifying differences in methylation levels. We hypothesize that environmental factors can increase epigenetic variability at functionally relevant genomic regions in individuals after exposure, and that the environmentally-induced epigenetic variability can in turn trigger metabolic responses relevant to disease. Here, we used a twin-based cohort (TwinsUK) to identify CpG-sites within the Illumina Infinium HumanMethylation450 array where the variance in methylation differed between smokers and non-smokers, and we denote these as *environmental variable methylated sites* (envVMS). We identified more than 80 envVMS using two approaches: i) a population-based approach, in which we compared the methylation mean and variance between smokers and non-smokers, and ii) a smoking-concordant MZ twin study, in which we assessed the methylation differences within twin pairs. Heritability analyses confirmed that higher environmental contribution to epigenetic variance was observed at envVMS in smokers. Population-based envVMS are likely to partly capture gene-environment interactions, as opposed to envVMS obtained from the MZ-twin design, and consistent with this hypothesis we observed an enrichment of genotype-by-smoking interactions effects on DNA methylation levels at CpG-sites that were population-unique envVMS. We also assessed the association between envVMSs and metabolic and healthy ageing traits to explore the hypothesized impact of environmental-induced epigenetic variability on disease-related phenotypes. We detect significant associations between methylation at envVMS and a number of metabolic traits only in smokers. Our results support the hypothesis that epigenetic changes can mediate phenotypic plasticity in response to environmental exposures.

2131W

Enabling high throughput next generation sequencing from low input ChIP and cfDNA. *K. Cunningham[1], A. Heguy[2], P. Meyn[2], D. Horvath[3], J. Laliberte[1], S. Sandhu[1], C. Schumacher[1], L. Kurihara[1], V. Makarov[1], T. Harkins[1].* 1) Swift Biosciences, Ann Arbor, MI; 2) New York University, New York, NY; 3) Beckman Coulter, Indianapolis, IN.

Preparation of large quantities of limiting, difficult samples such as cfDNA and ChIP DNA can be a time consuming, challenging process, with the end result of sequencing data that does not support the desired goals of the project. Here we present enabling the generation of high quality data from these limiting, difficult samples by harnessing the power of Accel-NGS® 2S library preparation kits with high-throughput automation on the Beckman Coulter Biomek® FX*ᴾ* Liquid Handling Workstation to prepare large quantities of samples in a timely, robust manner. Highly efficient library preparations that maximize complexity and uniform representation of the genome make the most efficient use of next generation sequencing (NGS) reads and enable comprehensive analysis of DNA samples of limited quality or quantity. End repair of both the 3' and 5' DNA termini utilized in the Accel-NGS 2S Plus and PCR Free DNA Library Kits allows greater efficiency of adapter ligation. Additionally, sequential adapter attachment requires no adapter titration for lower input quantities. Together, these deliver a more complex library requiring less sequencing, reducing the overall sequencing cost for a given sample and enabling production of high quality libraries from low yield ChIP and 10ng input of circulating, cell-free DNA. As little as 100pg of ChIP starting material can be used to generate confident peak calls that are equivalent to ng input quantities, which decreases the number of cells required without loss of data quality. Low input PCR-Free library construction of cfDNA is enabled by Swift's highly efficient adapter ligation and relatively undamaged cfDNA ends, resulting in even higher conversion efficiency of input DNA into functional library molecules. The minimal sequence-dependent bias of Accel-NGS 2S adapter attachment results in libraries that faithfully represent the cfDNA sample. Data quality was determined by analysis of data generated from sequencing on an Illumina® MiSeq®. Optimized Accel-NGS 2S configuration combined with the flexible workflow options and optimized pipetting of the Biomek FX*ᴾ* Liquid Handler minimize consumable use and sample loss and increase lab efficiencies with a walk-away solution capable of generating up to 96 libraries in 4.5 hours. The technology provides the opportunity to use one kit and one automated program to generate libraries from multiple sample types, regardless of their quality and quantity, in a reproducible and accurate manner.

2132T

Joint allelic imbalance analysis of RNA and ATAC sequencing to explore the genomic regulatory landscape. *C.E. Krebs[1], T. Wu[2], L.M. Olde Loohuis[2], R.A. Ophoff[1,2].* 1) Department of Human Genetics, UCLA, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA.

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a fast and comprehensive way to characterize genomic regulatory variation with regard to chromatin accessibility. Along with other sequencing-based molecular phenotypes, these assays provide precise information about the regulatory landscape of cells, including allele-specific differences between maternal and paternal chromosomes, which can occur as the result of heterozygous *cis*-acting quantitative trait loci (QTL). But, unlike QTL mapping, which requires large sample sizes, detection of allelic imbalance can be done in as few as one individual with simple binomial testing of allelic ratios. In this study, we sought to explore the landscape of allelic imbalance in RNA-seq and ATAC-seq data from lymphoblast cell lines (LCLs) of a HapMap trio, hypothesizing that allele-specific expression (ASE) would coincide with allele-specific chromatin accessibility (ASCA). We were able to determine statistically significant allelic imbalances at heterozygous sites called from whole genome sequencing by using the GATK tool ASEReadCounter, then performing a binomial test at each site using a null allelic ratio of 0.5, followed by a 5% FDR correction of p-values, and finally by filtering out sites with an effect size less than 0.15. Of an average of 2,347,885 heterozygous sites per sample, an average of 16,942 RNA-seq sites and 67,114 ATAC-seq sites per sample passed filters for mapping and base quality and were subject to allele counting by the tool. Of these, an average of 4,556 sites per samples displayed statistically significant ASE and 4,829 displayed ASCA, and an average of 978 genes per sample shared an allelic imbalance in chromatin accessibility and gene expression. Considering that some of these imbalances are caused by base calling errors or by silencing due to imprinting, these numbers are in line with previous studies and indicate that at least some genetic regulation of gene expression is probably mediated by chromatin accessibility. Overall, the joint analysis of allelic imbalance in multiple sequencing-based molecular phenotypes is a powerful way to explore individual genomic regulatory variation and, when scaled up, these assays can be used to elucidate population-level epigenomic variation and could potentially provide insight into the mechanisms of disease-associated genetic variants.

2133F

Global analysis of transcription factor affinity *in vivo.* *A. Stergachis[1,2], H. Wang[2], M. Maurano[2], A. Reynolds[2], L. Yang[3], S. Neph[2], M. MacCoss[2], R. Rohs[3], J. Stamatoyannopoulos[2,4,5].* 1) Internal Medicine, Brigham and Women's Hospital, Boston, MA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Molecular and Computational Biology Program, University of Southern California, Los Angeles, CA; 4) Department of Medicine, University of Washington, Seattle, WA; 5) Altius Institute for Biomedical Sciences, Seattle, WA.

The affinity of a transcription factor (TF) for individual genomic recognition sites in vivo is fundamental to regulatory DNA function, yet remains virtually undefined. We used graded salt perturbation of TF-DNA electrostatic interactions to quantify the relative affinity of the genome regulator CTCF at >20,000 occupancy sites in human cells. We find that the occupancy landscape is fundamentally partitioned between strictly affinity-dependent elements and cooperativity-dependent elements relying on neighboring TFs. The former localize chiefly within compact distal regulatory regions, while the latter are found primarily within promoters and larger distal elements. Cell-selective occupancy is achieved mainly by cooperative utilization of low-affinity binding sites. Affinity appears to be the predominant driver of evolutionary constraint at TF recognition sites, and explains observed occupancy patterns in diverse mammals.

2134W

Epigenomic signatures of high-altitude adaptation. *A. Childebayeva[1,2], D.C. Dolinoy[2], J.M. Goodrich[2], F. Leon-Valerde[3], M. Rivera-Ch[3], M. Kiyamu[3], T.D. Brutsaert[4], A.W. Bigham[1].* 1) Anthropology, University of Michigan, Ann Arbor, MI; 2) Environmental Health Sciences, University of Michigan Ann Arbor, MI; 3) Departamento de Ciencias Biológicas y Fisiológicas, Universidad Peruana Cayetano Heredia, Lima, Peru; 4) Exercise Science, Syracuse University, Syracuse, NY.

High-altitude adapted individuals show distinct circulatory, respiratory, and hematological adaptations to chronic hypoxia. Emerging genetic data support an evolutionary origin and a genetic basis for these observed physiological adaptations to high altitude. However, the epigenetic contribution to adaptation to hypobaric hypoxia has not been characterized. We performed a genome-wide Illumina Infinium MethylationEPIC array on n=28 whole blood samples from individuals of Quechua high-altitude ancestry living in Peru. We performed a differential methylation analysis of the samples between a group recruited at high altitude and a group recruited at low altitude. After normalization of the data, we found n=253 differentially methylated CpG sites between our groups at FDR cutoff 0.05 corresponding to signaling regulation, signaling transduction, cell division, apoptosis, as well as other pathways based on a pathway analysis performed in DAVID. These data show that exposure to high altitude may have an effect on the epigenome, and contribute to our understanding of ways in which the human organism can respond and adapt to environmental conditions. This project was funded by National Science Foundation grant 1132310, the Leakey Foundation and the Department of Anthropology at the University of Michigan.

2135T

Smoking associated methylation quantitative trait loci preferentially map to neurodevelopmental pathways. *M. Dogan[1,2], S.R.H. Beach[3], R. Philibert[1,2,4].* 1) Biomedical Engineering, University of Iowa, Iowa City, IA; 2) Psychiatry, University of Iowa, Iowa City, IA; 3) Center for Family Research, University of Georgia, Athens, GA; 4) Behavioral Diagnostics, Iowa City, IA.

Smoking is the leading preventable cause of morbidity and mortality in the United States. Smoking exerts its effects indirectly by increasing susceptibility to common complex diseases such as coronary heart disease. Our understanding of the molecular mechanisms through which smoking increases vulnerability for these disorders could still be improved. This is especially true for disorders than preferentially involve the central nervous system. Our study was designed to understand the effects of smoking on DNA methylation (DNAm) in the presence and absence of genetic context in the Framingham Heart Study (FHS). Data from 1599 individuals from the FHS Offspring cohort were used. These individuals were in their early to mid-sixties with a self-reported smoking rate of 7.6%. Genome-wide DNAm and SNP data were profiled using the Illumina HumanMethylation 450k BeadChip and Affymetrix GeneChip HumanMapping 500k Array Set, respectively. To understand the effects of smoking on DNAm in the absence of genetic variation, we regressed smoking against DNAm, controlling for age, gender and batch. After correction for multiple comparisons, methylation status at 525 sites was significant at a 0.05 level. Consistent with prior studies, the top-ranking probe was cg05575921 from the *AHRR* gene (p-value of 7.65×10^{-155}). Subsequently, to determine the effects of smoking on DNAm in the presence of genetic variation, cis- and trans-methylation quantitative trait loci (meQTL) analyses were conducted to determine the significant effects of SNP on DNAm given smoking status, controlling for age, gender and batch. A total of 126,369,511 cis and 195,068,554,297 trans analyses were performed. Of those, 5294 (0.00419%) and 422,623 (0.00022%) significant cis- and trans-meQTL were generated after correction for multiple comparisons at a 0.05 significance level. To better visualize and compare the connectivity and gene ontology enrichment between the results of both analyses, we generated protein-protein interaction networks. While the DNAm analysis mapped to inflammatory pathways, the cis- and trans-meQTL analyses mapped to neurodevelopmental pathways. These neurodevelopmental pathways could provide additional insight into the association of smoking to psychiatric disorders. Furthermore, this study demonstrates that combined genetic and epigenetic analyses may be crucial in better understanding the interplay between environmental variables such as smoking and pathophysiological outcomes.

2136F

Systematic detection of spatiotemporal patterns of epigenetic changes. *P. Fiziev, J. Ernst.* Biological Chemistry, University of California, Los Angeles, Los Angeles, CA.

Post-translational histone modifications associate with important regulatory regions such as promoters and enhancers that control the expression of genes. Time-course genome-wide maps of these epigenetic marks have become available in a growing number of biological systems. Here, we present a novel computational method for systematic detection of major classes of spatio-temporal patterns of epigenetic changes. It takes as input data from a series of chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) experiments for a single histone mark that are performed at consecutive time points during a given biological process. The method identifies regions for which the mark either expands or contracts significantly with time or holds steady. The classification is based on a Bayesian network that models explicitly the movements of domain boundaries across time points. We show applications of our method on multiple available datasets that characterize chromatin dynamics in different biological systems. We find that spatial dynamics of specific histone marks are coordinated and occur at key locus control regions that regulate genes involved in establishing and defining cell type identity. Expanding and contracting domains also associate with differential transcription factor binding. In addition, our method enables uniquely to study asymmetric spatial patterns such as unidirectional expansions and contractions and compare them to bidirectional patterns. We find that asymmetric directional dynamics are overrepresented in promoter proximal regions and correlate strongly with direction of transcription. Our method is generally applicable to a wide range of time series ChIP-seq data and thus can help in understanding the chromatin dynamics in many biological systems.

2137W

Effects of socioeconomic status on early life DNA methylation. *S.J. Goodman[1,2], W.T. Boyce[3,4,5], S. Lam[4], J.L. MacIsaac[1], M.J. Jones[1], S.M. Mah[1], A. Zaidman-Zait[4], M.S. Kobor[1,2,3,4].* 1) Centre for Molecular Medicine and Therapeutics, BC Children's Hospital, Vancouver, British Columbia, Canada; 2) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 3) Child and Brain Development, Canadian Institute for Advanced Research, Toronto, Ontario, Canada; 4) Human Early Learning Partnership, University of British Columbia, Vancouver, British Columbia, Canada; 5) Pediatrics, UCSF School of Medicine, San Francisco, California, USA.

   Rationale: Socioeconomic status (SES) is a potent environmental determinant of human health. Childhood experiences of SES alter development and lead to health disparities later in life, suggesting that they become 'biologically embedded'. DNA methylation represents an ideal candidate mechanism of biological embedding as it mediates interactions between genetic background and the environment and is especially plastic during developmental periods. Therefore, it has been proposed that epigenetic patterns established in early life as a result of unique experiences and exposures may influence future health outcomes. One major obstacle to uncovering these differences is that SES is heavily confounded with ethnicity in both Canada and the US; this presents a challenge as genetic background strongly contributes to epigenetic variation. We hypothesize that socially driven differential DNA methylation exists in children from diverse socioeconomic backgrounds and, importantly, these methylation patterns can be teased apart from ethnically driven DNA methylation differences. **Methods:** Gene Expression Collaborative for Kids Only (GECKO) is a cohort of 400 socioeconomically diverse, 8-10 year old children which was established to study the biological embedding of early life socioeconomic-related adversity. Using the Illumina Human Methylation 450K array platform to assay DNA methylation and the Illumina PsychArray BeadChip to evaluate SNPs, we have examined genetic, epigenetic, and environmental interplay as it relates to childhood SES. **Results and implications:** Genetic heritage was calculated from the PsychArray data and used to correct for the unequal division of ethnic groups across SES. Epigenetic correlates of childhood SES, as measured by family income, parent occupation and education, uncovered a set of candidate loci, irrespective of genetic heritage. These loci may underlie biological embedding of early life SES. These findings were supported by a parallel weighted gene co-expression network analysis (WGCNA), applied to uncover networks of co-methylated CpGs associated with SES. The strength of WGCNA is that it not only assesses the relationship between DNA methylation and a variable of interest, but it also uncovers relationships between the CpGs themselves. This tool, when paired with traditional EWAS methods may bring us closer to identifying epigenetic disturbances common to early childhood adversity and their relationship to later health outcomes.

2138T

DNA methylation concordance in pediatric blood and buccal tissues. *S.A. Islam[1,2], S.J. Goodman[1,2], J.L. MacIsaac[1,2], R.G. Barr[3,6], W.T. Boyce[4,6], M.S. Kobor[1,2,5,6].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Centre for Molecular Medicine and Therapeutics, BC Children's Hospital, Vancouver, BC, Canada; 3) Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada; 4) Departments of Pediatrics and Psychiatry, University of California, San Francisco, USA; 5) Human Early Learning Partnership, University of British Columbia, Vancouver, British Columbia, Canada; 6) Canadian Institute for Advanced Research, Toronto, Ontario, Canada.

   Recently, population epigenetic studies have become prominent in the field of developmental origins of health and disease (DOHaD). To date, this research has largely focused on identifying differences in DNA methylation (DNAm) associated with early life exposures, demographic factors, or disease states using accessible human tissues, such as peripheral blood and buccal epithelial cells (BECs). Although it is well established that tissue of origin is a main driver of DNAm variability, along with genetic background, there is relatively little known about intra-individual DNAm variation across tissues particularly in pediatric samples. It is also unclear how genetic variation may influence tissue-specific DNAm. Given the widespread use of these accessible tissues for inferring phenotype associations to DNAm, it is important to ask how these tissues compare in terms of DNAm variability at various developmental stages and to assess how genetic differences impact DNAm variance. In this study, we assessed the concordance of genome-wide DNAm profiles between matched peripheral blood mononuclear cells (PMBCs) and BECs in two independent cohorts of children aged 3-4 years old (n = 16) and 8-9 years old (n = 77). Specifically, we compared DNAm variance across ~430,000 sites, measured using the Illumina Infinium HumanMethylation450 BeadChip. Correlation of DNAm measures across individuals within a tissue revealed an appreciable number of highly-correlated sites with discrete distributions, signifying a potential genetic influence on DNAm at these sites. Using the Illumina Infinium PyschArray, we measured genetic variation at over ~500,000 single nucleotide polymorphisms (SNPs) and identified sites, known as methylation quantitative trait loci (mQTLs), which significantly associated with DNAm within a 5 kb window (FDR < 0.05). We found that although a substantial number of these cis-mQTLs were tissue-specific, there were also many sites which were shared across tissues as well as a number which showed a regional association with DNAm. Taken together, the work presented here offers insight into tissue-specific DNAm variability in pediatric samples and the influence of genetic variation on DNAm variance to help guide decisions on the selection of appropriate tissue samples and the interpretation of findings in population epigenetic studies.

2139F

DNA methylation at birth is associated with prenatal exposure to alcohol, cigarette smoke, and illicit drugs. *M.J. Jones[1], N. Koen[2], J.L. MacIsaac[1], A.M. Morin[1], K. Donald[2], A. Vanker[2], K.C. Koenen[3], H.J. Zar[2], M.S. Kobor[1], D.J Stein[2].* 1) University of British Columbia, Vancouver, British Columbia, Canada; 2) University of Cape Town, Cape Town, South Africa; 3) Harvard University, Boston, Massachusetts, USA.

Rationale: Alterations in DNA methylation at birth have been associated with prenatal exposures, and the potential for these marks to last throughout the lifetime presents a possible mechanism for the developmental origins of health and disease. We assessed DNA methylation changes present at birth in a cohort of 150 children from Drakenstein, South Africa, with high frequency of prenatal exposure to alcohol, cigarette smoke, and illicit drugs, in an effort to understand how these early exposures can be associated with lifelong health. Methods: We evaluated DNA methylation in cord blood collected at birth using the Illumina 450k array. We identified a subset of samples showing contamination with maternal blood, and created a screening protocol to identify and exclude these contaminated samples from this and future analyses. As cord blood cell types had not been measured, we used a reference-free method to control for inter-individual differences in cell type. Results: We identified DNA methylation changes associated with alcohol, smoking, and illicit drug exposure. Cigarette smoke exposure results were consistent with previous reports, exhibiting changes at the *AHRR* and *CYP1A1* genes among others. We also compared all three types of exposure to determine whether they cause independent or overlapping signatures of DNA methylation at birth, and whether they were occurring at similar or different types of genomic features. Conclusions and future directions: Prenatal exposures to alcohol, cigarette smoke, and illicit drugs leave behind a residue of exposure in the form of DNA methylation changes present at birth. These changes have the potential to last and influence health trajectories for the rest of the child's life. It is important to assess the persistence of these marks as well as the molecular-level outcomes to fully understand how DNA methylation may connect early life exposures to later life health. Longitudinal studies with detailed data collection may also be able to identify the timing of windows of opportunity when these exposures may become biologically embedded.

2140W

Identifying differentially methylated regions (methylome changes) in response to heavy, long-term cannabis use, in a large longitudinal cohort. *A.J. Osborne[1], J.F. Pearson[4], L.J. Horwood[5], D. Macartney-Coxson[3], T.A. Hore[2], N.J. Gemmell[2], M.A. Kennedy[1].* 1) Department of Pathology, University of Otago Christchurch, Christchurch, Canterbury, New Zealand; 2) Department of Anatomy, University of Otago, Dunedin, Otago, New Zealand; 3) Environmental Science and Research, Kenepuru Science Centre, Porirua, New Zealand; 4) Dean's Department, University of Otago Christchurch, Christchurch, Canterbury, New Zealand; 5) Psychological Medicine, University of Otago Christchurch, Christchurch, Canterbury, New Zealand.

Cannabis use is an important global public health issue, and a growing topic of controversy and debate. There is strong evidence that heavy cannabis use is associated with increased risk of adverse outcomes in a number of areas, including educational achievement, mental health, other illicit drug use and motor vehicle collisions. Many of these findings are drawn from the work of the Christchurch Health and Development Study (CHDS), a longitudinal study of a birth cohort of 1265 children born in 1977 in Christchurch, New Zealand, who have been studied on 23 occasions from birth to the age of 35 (n=987). In the early 1990s, research was initiated into the development and consequences of cannabis use. Drug metabolism, response and addiction have strong genetic components, and multiple genome-wide association studies (GWAS), including studies on the CHDS cohort, have identified genes and allelic variants that are likely contributors to substance use disorders. However, it is clear that not all biological responses are determined by direct changes in DNA sequence. Environmental factors, including drugs, can lead to epigenetic alterations, which are an important source of variation and regulation in the genome. Methylation is one such alteration, and can influence chromatin structure and alter gene expression. This can result in an epigenetic pattern that is quite specific to individual environmental exposures. For example, differentially methylated regions (DMRs) have been identified that are specifically modified in smokers versus non-smokers. It stands to reason that long-term, heavy cannabis use will also similarly affect the methylation status of users and that there will be epigenetic hallmarks of cannabis use. We are utilising Illumina EPIC 850K methylome arrays to identify DMRs in a subset of CHDS participants. The rich, lifetime psychosocial data available for the CHDS cohort mean we are able to control for cigarette smoking and other important environmental factors, during the discovery phase of this study. We anticipate that DMRs we discover may be located in genes, or regulatory regions, that are important in the biological response to cannabis. By highlighting genes and pathways that are influenced by long term cannabis use, we will better understand the impacts of cannabis on genome function, provide insights into human responses to cannabis, potential adverse and medical effects, and inform current public debates about cannabis.

2141T

Genomic context drives the conservation of enhancer activity across species. *A. Fish[1], L. Chen[2], J. Capra[1,2].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Biological Sciences, Vanderbilt University, Nashville, TN.

Despite broad conservation of gene expression and transcription factor (TF) binding preferences across species, there is substantial turnover in enhancer activity patterns. Analyses of sequence-level differences between orthologous regions do not explain the majority of variance in enhancer activity. In this study, we evaluated whether DNA sequence patterns and functional genomics data could distinguish enhancers active in the livers of ten diverse mammalian species (activity-conserved) from those with orthologous sequence present across the ten species, but active only in human (human-specific). Activity-conserved and human-specific liver enhancers exhibit some differences at the DNA sequence level; e.g., activity-conserved enhancers have twice the density of TF binding motifs. However, a support vector machine (SVM) classifier trained on DNA sequence patterns (6-mer spectra) had only a modest ability to distinguish activity-conserved from human-specific enhancers (ROC AUC: 0.65). Integrating epigenetic modifications, TF binding, and chromatin accessibility significantly improved prediction of enhancer activity across species (ROC AUC: 0.84). Functional genomics features traditionally associated with enhancers, such as H3K27ac and H3K4me1, were most predictive of activity conservation, including those from cellular contexts other than liver. Further, activity-conserved liver enhancers overlapped enhancers active in other human cellular contexts at a significantly higher rate than human-specific enhancers. Our results illustrate that sequence changes alone are likely insufficient to account for the variable enhancer activity of orthologous sequences across species, and that the genomic environment influences conservation of activity patterns.

2142F

Genetic and transcriptional analysis of human host response to healthy gut microbiome. *A. Richards[1], M. Burns[2,3], A. Alazizi[1], L. Barreiro[4], R. Pique-Regi[1,5], R. Blekhman[2,3], F. Luca[1,5].* 1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI; 2) Departments of Genetics, Cell Biology, and Development, The University of Minnesota, Minneapolis, MN; 3) Department of Ecology, Evolution, and Behavior, The University of Minnesota, Minneapolis, MN; 4) Department of Pediatrics, Sainte-Justine Hospital Research Centre, University of Montreal, Montreal, QC, Canada; 5) Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI.

The gut microbiome has been shown to play a role in both healthy and diseased states. Several recent studies have linked human genetic variation to changes in the composition of the microbiome. However, establishing the causality of host-microbiome interactions in humans is still challenging. Here, we present a novel experimental system to characterize the transcriptional response induced by the microbiome in human cells and shed light on the molecular mechanisms underlying host-gut microbiome interactions. We cultured primary human colonic epithelial cells in low oxygen to recapitulate the gut environment together with the gut microbiome of healthy individuals extracted from fecal samples. We sequenced the RNA from each sample and identified over 6,000 host genes that change expression following co-culturing with the gut microbiome as compared to colonocytes cultured alone. The differentially expressed genes are enriched for genes associated with several microbiome-related diseases, such as obesity and colorectal cancer. In addition, we identified 87 host SNPs that show allele-specific expression in 69 genes. For 12 SNPs in 12 different genes, allele-specific expression is conditional on the exposure to the microbiome. Of these 12 genes, eight have been associated with diseases linked to the gut microbiome, specifically colorectal cancer, obesity and type 2 diabetes, suggesting that these genes may link the microbiome mechanistically to these diseases. For example, *LASP1*, a gene involved in the cytoskeleton and cell migration, had an elevated total expression as well as allele-specific expression only following co-culturing with the microbiome. Furthermore, *LASP1* expression is increased in colorectal cancer suggesting that the microbiome may influence colorectal cancer through modulation of *LASP1* expression in a genotype-specific manner. Our study demonstrates a scalable approach to study host-gut microbiome interactions and can be used to identify putative mechanisms for the interplay between host genetics and microbiome in health and disease.

2143W

The analysis of conserved non-coding elements and topologically associating domains reveals strong spatial correlation between them. *B. Ambroise, J.C. Fischer, L.E. Hunt, G. Elgar.* The Francis Crick Institute, London, United Kingdom.

The recent development of genome wide chromosome conformation capture (Hi-C) has permitted the study of chromatin interactions within the nucleus. The resulting interactions maps show that that genomes can be divided into large local chromatin domains termed Topologically Associated Domains (TADs). Within TADs, the genome appears to be organized to favour strong internal chromatin interactions rather than external interactions with neighbouring TADs. It has also been suggested that TADs might help to de-lineate basic genomic functions such as gene regulation. Interestingly, TADs also appear to be conserved across species and cell types. Identifying cis-regulatory elements of genes has been a major challenge. Comparative genomics has revealed the existence of conserved non-coding elements (CNEs) in vertebrates. Multiple lines of evidence suggest that these non-coding units are cis-regulatory elements associated with key developmental genes. Using CNEs clusters extracted from the CONDOR database, we investigated whether CNEs and TADs were spatially correlated in the human genome. We looked at different parameters such as absolute and relative distances, overlap of the regions and correlation of the boundaries. In addition, we performed 10,000 simulations to measure the statistical robustness of our analysis. In some cases, we observed that CNE clusters were split between multiple TADs. As a result, CNEs within such clusters were tested for enhancer activity in GFP-reporter gene assays in zebrafish embryos. Our results show that CNE clusters are not randomly distributed across the genome but are in fact distributed with respect to the TAD architecture. Furthermore, zebrafish reporter assays showed that CNEs within one cluster can be associated with more that one developmental gene. This confirmed that TAD boundaries are true functional boundaries and could potentially help to better identify cis-regulatory elements among CNEs.

2144T

Outlier gene expression networks identify functional rare variants. *J.R. Davis[1], E.K. Tsang[2], B. Strober[3], X. Li[4], Z. Zappala[1], L. Fresard[4], A. Battle[5], S.B. Montgomery[1,4].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Biomedical Informatics Program, Stanford University, Stanford, CA; 3) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD; 4) Department of Pathology, Stanford University, Stanford, CA; 5) Computer Science Department, Johns Hopkins University, Baltimore, MD.

Although well powered to discover common regulatory variants, *cis*-eQTL studies have largely overlooked the impact of rare regulatory variation (MAF < 1.0%). Recent efforts have leveraged gene expression outliers to study the effects of functional, rare regulatory variants; however, these studies have been limited by sample size and mostly focused on individual genes. Existing methods to discover expression outliers across genes in a regulatory module have been limited to single tissues and have focused only on cohorts lacking genome-sequencing data. In this project, we extend methods to detect outlier gene expression across multiple genes in a regulatory network and compare results across tissues. By examining the correlation of outlier effects through a network, we can enrich for high quality expression outliers and rare regulatory variants with broad effects across multiple genes. We applied our detection methods to the 5 most widely sampled tissues in the GTEx dataset: muscle, whole blood, skin, adipose subcutaneous, and tibial artery (minimum sample size of 285). For each tissue, we first constructed gene expression networks using a multivariate normal model with L1-regularization to induce sparsity. Using the estimated covariance matrices, we then identified individuals who show significant outlier expression across a given module. We show that these multi-gene outliers are enriched for potentially functional, rare regulatory variants such as promoter variants and highly conserved non-coding variants. By examining specific outlier modules, we show that these rare variants appear to directly influence only a subset of genes, though their effects can appear throughout the module. Finally, we use our methods in a recently developed dataset of undiagnosed rare disease samples to discover regulatory modules with outlier expression compared to healthy GTEx controls. By combining information across gene modules, we assess our power to discover outlier expression effects that are enriched for disease-relevant genes. Our results highlight the importance of considering gene networks to detect functional, rare regulatory variants.

2145F

Functional annotation of the genome using a high-throughput enhancer assay that retains sequence and chromosomal context. *J.C. Klein[1], J. Shendure[1,2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Functional annotation and variant effect prediction in DNA regulatory elements remain major challenges for the field. ChIP-seq, chromatin accessibility and computational approaches have addressed these questions in part, but fail to provide functional validation of specific elements. Massively-parallel reporter assays (MPRAs) allow the direct read-out of the effects of thousands of sequences on transcriptional activity. However, current MPRAs test the ability of short sequences to increase transcription of a nearby reporter on a plasmid. This design inherently misses several characteristics of *in vivo* enhancer elements: 1) enhancers can be longer than the 200bp limit imposed by array-synthesis, 2) it is unlikely that plasmids reflect native chromatin, and 3) enhancers can be located as far as 1MB from their target promoter. To address these limitations, we developed a new class of MPRAs to screen libraries of assembled synthetic sequences for enhancer function in the context of a native, chromosomal locus.   The assay measures the ability of a given sequence to increase the expression of *LMO2*, a gene whose up-regulation resulted in leukemia-like symptoms in several SCID gene therapy patients, when inserted into *LMO2*'s 3' UTR. Unlike conventional episomal MPRAs, where up to 20% of randomly selected sequences can drive transcription, our assay appears to have high specificity. When negative control and LTR fragments were inserted at a similar ratio into the 3'UTR of *LMO2,* 99.8% of all RNA transcripts contained a LTR fragment. Of the 0.2% of RNA transcripts containing the negative control sequence, none were enriched relative to genomic DNA. We are currently screening libraries of thousands of putative enhancer elements in several cell types, including Jurkat T cells, Hek293T cells, HepG2 and GM12878. Using multiplex pairwise assembly of array-derived oligonucleotides, we are screening the same set of putative enhancers both as 173-mers and 352-mers to directly compare the effects of additional sequence context on function.   We are also using this assay to characterize the effects of thousands of GWAS-implicated variants on enhancer function in their appropriate cell lines, with the goal of identifying causal variants for these associations.

2146W

Pinpointing noncoding sequence variation affecting transcription factor occupancy. *M.T. Maurano[1], J.M. Halow[2,3], R. Byron[4], M. Groudine[4], M.A. Bender[4], J.A. Stamatoyannopoulos[2,3].* 1) Institute for Systems Genetics, New York University Langone Medical Center, New York, NY 10016, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 3) Altius Institute for Biomedical Sciences, Seattle, Washington, USA; 4) Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

The function of human regulatory regions depends exquisitely on their local genomic environment and on cellular context, complicating experimental analysis of common disease- and trait-associated variants that localize within regulatory DNA. Allelic profiling of chromatin accessibility across individuals has shown that a minority of sequence variation affects transcription factor (TF) occupancy, yet the low sequence diversity in human populations means that no experimental data are available for the majority of disease-associated variants. We performed deep allelic profiling of chromatin accessibility in F1 hybrids of highly diverged mouse strains and generated high-resolution maps of variants affecting TF activity at their native locus and chromatin environment. The high density of heterozygous sites in these strains enables a precise quantification of the effect size and cell-type specificity of variants associated with altered chromatin accessibility throughout the genome. We develop a compendium of TF-specific sensitivity profiles to account for genomic context effects and model cell-type specific perturbations of the protein-DNA interface. Taking advantage of the high conservation of TF coding sequence among mammals, we apply these models to the classification of single nucleotide polymorphisms in the human genome. These models enable quantitative prediction of the effect of non-coding variation on TF activity across a variety of cellular contexts, facilitating both fine mapping and systems-level analyses of common disease-associated variation in human genomes.

2147T

Genetics of local gene expression across 44 human tissues. *S. Montgomery[1], C. Brown[2], GTEx Consortium.* 1) Departments of Genetics and Pathology, Stanford University, Stanford, CA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA.

Genetic studies of gene expression provide powerful means to identify functional non-coding variants influencing local gene expression and underlying disease pathogenesis. However, to date, many of these studies have been performed in a limited number of tissues in diverse cohorts and individuals. Moreover, such studies have been biased towards tissues that are more readily obtained but may be less relevant for the interpretation of human disease. We present the identification of cis-eQTLs from 44 distinct tissues examined in 70 to 361 individuals in the latest release of the GTEx project (v6). Relative to the GTEx pilot release, this analysis represents a large expansion that has now surveyed an additional 35 tissues and identified thousands of additional eSNPs and eGenes. We discover that 87% of genes have an eQTL and uncouple multiple signals of association for each gene. This abundance of eQTL highlights the need for increased care in using these data for the interpretation of GWAS; indeed, we find that 94% of all SNPs are nominally associated with the expression of one or more genes across GTEx. Further, we discover an increasing number of tissue-type specific eQTLs that are uncovered due to increases in sample size and highlight a unique identity in each surveyed tissue. By integrating large-scale epigenomics data, diverse genome annotation and state-of-the-art fine-mapping algorithms, we provide high resolution analysis of the relationship between cross tissue gene regulation and genetic control of gene expression and identify the features of tissue-specific and shared causal variants. By integrating broader maps of allele-specific expression across tissues we enhance measures of sharing of eQTL. The broad tissue sampling design of GTEx enables us to assess the relationship between studies with large numbers of samples versus large numbers of tissues and their relative utility in interpreting GWAS signals. Finally, we highlight the growing compendium of data available on the GTEx Portal (from v7 onwards). Together, the GTEx project delivers an increased understanding of the genetic control of gene expression across a broad range of cell types and demonstrates exceptional utility for improving our understanding of gene regulation and human genetics.

2148F

Dissecting the functional architecture of local and distal gene expression regulation in multiple human tissues. *X. Liu[1], H.K. Finucane[1,2], A. Gusev[1], G. Bhatia[1], S. Gazal[1], B. Bulik-Sullivan[3,4], F.A. Wright[5], P.F. Sullivan[6], B.M. Neale[3,4], A.L. Price[1,7].* 1) Department of Epidemiology, Harvard Chan TH School of Public Health, Boston, MA; 2) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA; 6) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Genetic variants that modulate gene expression levels play an important role in the etiology of human diseases and complex traits. Large-scale eQTL mapping studies routinely identify large numbers of local eQTLs, but the molecular mechanisms by which genetic variants regulate expression is as yet unclear, particularly for distal eQTLs, which these studies are not well-powered to detect. In this study, we use a heritability partitioning approach to dissect the functional architecture of local and distal gene expression regulation in 15 human tissues. Using an extended version of stratified LD score regression to analyze large expression data sets, we leverage all variants—not just those that pass stringent significance thresholds—to achieve this goal. For both local (within 1Mb of TSS) and distal gene regulation, we estimate the proportion of SNP-heritability explained by SNPs in each functional category and define enrichment as this proportion divided by the proportion of SNPs in each category. We compute standard errors via a block jackknife on genomic regions. In analyses of gene expression in peripheral blood (N=2,494) (Wright et al. 2014 Nat Genet), the top enriched functional categories in local regulation included super enhancers (5.18x), coding regions (3.73x), conserved regions (2.33x) and four histone marks ($p < 3 \times 10^{-7}$ for all enrichments). These enrichments were more statistically significant (though somewhat smaller) than analogous enrichments previously reported for complex traits, emphasizing the large amount of signal in gene expression data. Very similar enrichments were observed across 14 tissues from the GTEx (GTEx Consortium 2015 Science) and MuTHER (Grundberg et al. 2012 Nat Genet) data sets. We also observed pervasive enrichments for distal regulation of peripheral blood gene expression: super enhancers (1.91x), coding regions (4.47x), conserved regions (4.51x) and two histone marks ($p < 3 \times 10^{-7}$ for all enrichments). These results are the first to systematically identify enrichments of distal regulatory elements. Genetic correlation of gene expression across tissues suggests that local gene expression regulation is largely consistent across tissues (average $r_g=0.738$), whereas distal regulation of gene expression is highly tissue-specific (average $r_g=0.084$). Overall, our study quantifies the importance of super enhancers, conserved regions, coding regions and histone modifications in the local and distal regulation of gene expression.

2149W

Finer analysis of the correlation between SNP, CpG methylation and gene expression. *F. Takeuchi.* Research Institute, Nat Cntr for Global Health & Med, Shinjuku-ku, Tokyo, Japan.

It has been known that CpG methylation level in promoter and gene expression level are negatively (i.e, inversely) correlated. However, recent studies involving genome-wide assays in hundreds of human samples discovered many positively-correlated loci as well as those negatively-correlated. Aiming to understand the molecular mechanism behind the complex relation, I analyzed the correlation between SNP, CpG methylation and gene expression. I used publicly available experimental data and applied statistical analyses including 1) classification of CpG sites according to locations in genes, 2) prediction of transcription factor binding motif disruption by SNP, and 3) extraction of sparse primary relations from the graph representing the correlation between SNP, CpG methylation and gene expression. This approach can help understand inter-individual variation of epigenomes, which likely underlies susceptibility to complex diseases.

2150T

Synthetic long non-coding RNAs (SINEUPs) rescue defective gene expression *in vivo.* A. Indrieri[1], C. Grimaldi[1], S. Zucchelli[2,3], R. Tammaro[1], S. Gustincich[2,4,5], B. Franco[1,6]. 1) Telethon Institute of Genetics and Medicine (TIGEM), 80078, Via Campi Flegrei 34, Pozzuoli (NA), Italy; 2) Area of Neuroscience, SISSA, 34136, Trieste, Italy; 3) Dipartimento di Scienze della Salute, Universita' del Piemonte Orientale, Novara, Italy; 4) Department of Neuroscience and Brain Technologies, Italian Institute of Technology, 16163, Genova, Italy; 5) TransSINE Technologies, 1-7-22 Suehiro-cho Tsurumi-ku, Yokohama, Kanagawa 230-0045 Japan; 6) Medical Genetics Services, Department of Translational Medicine, Federico II University, 80131, Naples, Italy.

Non-coding RNAs provide additional regulatory layers to gene expression as well as the potential to being exploited as therapeutic tools. Non-coding RNA-based therapeutic approaches have been attempted in dominant diseases, however their use for treatment of genetic diseases caused by insufficient gene dosage is currently more challenging. SINEUPs are long antisense non-coding RNAs that up-regulate translation in mammalian cells in a gene-specific manner, although, so far evidence of SINEUP efficacy has only been demonstrated in *in vitro* systems. We now show that synthetic SINEUPs effectively and specifically increase protein levels of a gene of interest *in vivo*. We demonstrated that SINEUPs rescue haploinsufficient gene dosage in a medakafish model of a human disorder leading to amelioration of the disease phenotype. Our results demonstrate that SINEUPs act through mechanisms conserved among vertebrates and that SINEUP technology can be successfully applied *in vivo* as a new research and therapeutic tool for gene-specific up-regulation of endogenous functional proteins.

2151F

Family studies identify transcripts with strong parent-of-origin effects in whole blood. *W.-C. Hsueh[1], J.E. Curran[2], H.H.H. Göring[2], R.L. Hanson[1].* 1) NIDDK, Phoenix, AZ; 2) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX.

A parent-of-origin effect (POE) describes the phenomenon where differential effects are observed depending on which parent transmitted the allele. In model systems and human studies, a growing number of examples of POEs are found to influence phenotypic variation, and some studies suggest that POEs are widespread. In studies of complex traits, the incorporation of POE can increase statistical power to identify risk variants by the use of a more precise genetic model. In this study, we analyzed the role of POEs in transcripts expressed in whole blood measured on the Illumina Human-HT12-v4 BeadChip in a family study including 750 Southwestern Native Americans. This sample is comprised of 3,408 self-reported relative pairs, including 610 full-sib pairs and 203 half-sib pairs. Levels of 15,521 transcripts that were expressed above background and which did not contain SNPs were analyzed. Familial resemblance was modeled using classic pedigree-based variance components methods with models assessing the additive heritability (h^2), dominance (d^2) and POE with separate maternal and paternal effects (h^2_M, h^2_P). Models were adjusted for effects of age, sex, Amerindian heritage and tribal affiliations. A false discovery rate (FDR) of ≤0.05 was considered significant. Of all transcripts analyzed, the mean heritability was 0.26, and 79% had significant additive heritability (mean h^2=0.30±0.09). None was found to have significant dominance (mean d^2=0.07). Over all transcripts mean h^2_M=0.20 and mean h^2_P=0.12. We identified 20 transcripts (from 20 genes) with significant POEs and significant differences in h^2_M and h^2_P; 18 showed significant evidence for increased paternal expression and 2 had significantly increased maternal expression. The h^2 for these 20 transcripts with significant POEs was higher compared to h^2 for those without significant POE (0.41±0.11 vs. 0.26±0.12, p<0.0001). None of these 20 genes are known imprinted genes. Heritability of transcripts expressed in whole blood is substantial and might be mostly explained by additive effects; however, our study identified a subset of transcripts that with strong POEs. Their effects on traits may be more appropriately modeled by incorporating POE, which will increase the statistical power to detect their effects. In addition, statistical modeling of POE may help to identify candidate genes of greater priority for experimental verification of imprinting.

2152W

A high-resolution landscape of transcriptional regulation unraveled by Transcription Initiation QTL (tiQTL). *H. Kwak[1,2], H.M. Kang[2].* 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Uncovering the functional elements of the human genome is accelerating the expansion of our basic biological knowledge of the regulation of gene expression. Recently, transcription at the regulatory elements such as enhancers has been discovered by high depth RNA-seq and nascent RNA sequencing methods, and some were shown to have direct functional role in target gene activation. However, it is still unclear what the impact of genetic variation is on the enhancer transcription, how the variations of the chromatin accessibility and modification states interpret into the transcription of target promoters, and whether the enhancer transcription is a mediator or a byproduct of enhancer activity. Therefore it becomes critical to directly map and measure the transcriptional activities at these regulatory elements. To address these questions, we present a genetic landscape of the transcriptional activity, by sequencing nascent RNAs of lymphoblast cell lines of 69 individuals in the Yoruban population. The nascent RNA sequencing revealed a precise map of enhancer transcription activities and transcription at promoters. We performed a Transcription Initiation Quantitative Trait Loci (tiQTL) analysis, and identified 1,053 unique transcription sites significantly associated with at least one genetic variant at FDR <0.1. While the majority of cis-tiQTLs were <10kb away from the corresponding TSS, 18% of cis-tiQTLs were located >100kb away from the TSS, suggesting long-range interaction between genetic variants and enhancer transcription activities. We observed that 34% of tiQTLs overlap with DNase hypersensitivity QTLs (dsQTLs), and these overlapping tiQTLs show 40% larger effect sizes than corresponding dsQTLs on average. We present examples of tiQTLs that have constantly open chromatin and not dsQTLs but are eQTLs. These findings suggest that enhancer transcription provides specific information on the mechanism by mediating target promoter activation and is not just a byproduct of chromatin accessibility. We also present examples of tiQTLs at upstream antisense transcription sites and the first intronic regions of which the effect sizes anti-correlate with the eQTL, suggesting possible repressive mechanisms of these previously less understood transcriptional activities. Results of our study will further expand the understanding of the genetic variation of active enhancer transcription, and their relationship with gene expression levels.

2153T

Fine mapping gene expression associations in IPS-derived sensory neurons with ATAC-seq. *J. Schwartzentruber[1], J. Rodrigues[1], A. Knights[1], K. Alasoo[1], A. Goncalves[1], M. Patel[1], S. Foskolou[2], R. Ferreira[2], C. Benn[2], A. Wilber[2], M. Bictash[2], E. Impey[2], L. Cao[2], L. Vincente[2], A. Loucif[2], P. Whiting[2], A. Gutteridge[2], D. Gaffney[1], HIPSCI Consortium (http://www.hipsci.org).* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) Pfizer Neuroscience and Pain Research Unit, Pfizer Ltd., Great Abington, Cambridge, United Kingdom.

Sensory neurons relay information from the tissues to the central nervous system, but are extremely challenging to study in humans, with primary cells only available via invasive tissue biopsy. Pain-sensing neurons are of particular interest because pain symptoms are a component of many diseases, and at least 10% of adults suffer from chronic pain. Sensory neurons derived from human induced pluripotent stem cells (hIPSCs) are a potentially promising *in vitro* model for pain. We differentiated sensory neurons from 97 IPS lines derived from healthy individuals by the Human Induced Pluripotent Stem Cell Initiative (HIPSCI) and measured gene expression (RNA-seq N=97), chromatin accessibility (ATAC-seq N=31), and neuron electrophysiology (N=30 to 56). We used single cell RNA-seq from one reference sample to identify markers of sensory neurons and other contaminating cell types, which enabled us to estimate the purity of our cell lines lacking scRNA-seq data. Using an allele-specific method, we detect over 3000 eQTLs, 2000 splice QTLs, and 6000 chromatin accessibility QTLs at an FDR of 10%, and show how the data can be used to fine map putatively causal variants for human neurological disease. Using our electrophysiology data we also find several genes with significant associations between gene expression and neuronal excitability across individuals. Our study illustrates how IPS-based model systems can be used to understand the functions of common genetic variants in neuronal cell lineages.

2154F

Identifying factors which control variable escape from X chromosome inactivation. *B. Balaton, C. Brown.* Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

X chromosome inactivation is the epigenetic silencing of one of the two X chromosomes in female eutherian mammals, to dosage compensate with males who have a single X chromosome. Not all X-linked genes are silenced; 15% escape inactivation while another 15% differ in their inactivation status between individuals, tissues and studies. This latter class of variable escape genes provides an opportunity to study a gene that is escaping inactivation and one which is subject to inactivation in the same genomic context, allowing identification of the underlying genetic and epigenetic causes of escape from inactivation. By combining the results of 3 previous studies using four different methods to obtain a gene's X-inactivation status, we assigned a consensus inactivation status call to 639 genes, including 78% of X-linked protein coding genes which are expressed in females. Eight percent of genes were found to variably escape between individuals and tissues while another 7% were variable between studies. We are interested in identifying which features correlate with a gene's X-inactivation status within a cell line. Our list of consensus X-inactivation status calls identified regions of variable escape from inactivation, suggesting possible co-regulation; however, preliminary results suggest that neighboring variable escape genes are independently regulated. We therefore hypothesized that promoter CpG methylation of variable escape genes would be predictive of their X-inactivation status; however, variable escape genes tended to have a broad range of methylation levels and did not correlate completely with their X-inactivation status. Our results show that methylation may be contributing to the variability of these genes' X-inactivation statuses, but that there must be other unaccounted factors. We are now investigating CTCF binding occupancy, topologically associated domain boundaries, and retrotransposon polymorphisms to determine whether or not they are also contributing to this variable regulation. Genetic or epigenetic features which are strongly associated with a gene's X-inactivation status may be causative of escape or silencing. Understanding the cause of escape from inactivation will allow us to more easily test for and predict which variable escape genes are being expressed in individuals. This knowledge may eventually allow us to change the genetic or epigenetic context surrounding a gene and inactivate or reactivate the copy on the inactive X.

2155W

Role of *Dxz4* in the 3D bipartite structure of the inactive X chromosome.
X. Deng[1], G. Bonora[2], V. Ramani[2], W. Ma[7], R. Qiu[2], Z. Duan[3,4], J. Shendure[2], W.S. Noble[2,5], C.M. Disteche[1,6]. 1) Pathology, University of Washington, Seattle, WA., USA; 2) Genome Sciences, University of Washington, Seattle WA., USA; 3) Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle WA., USA; 4) Hematology, University of Washington, Seattle WA., USA; 5) Computer Science and Engineering, University of Washington, Seattle WA., USA; 6) Medicine, University of Washington, Seattle WA., USA; 7) Statistics, University of California, Riverside CA., USA.

X inactivation is associated with *XIST* RNA coating, epigenetic modifications, and compaction of the X chromosome. To determine the 3D structure of the X chromosomes based on allele-specific chromatin contacts in female cells we employed a new in situ Hi-C assay in mouse F1 hybrid systems (brain and Patski cells) in which alleles can be distinguished based on single nucleotide polymorphisms. We previously reported that the inactive mouse X chromosome has a strikingly different 3D configuration from the active X. Indeed, the inactive X comprises two condensed superdomains of frequent long-range intrachromosomal contacts separated by a hinge region, a configuration conserved in human. Genes that escape X inactivation are preferentially located at the periphery of the inactive X. Both in mouse and human the hinge region contains the conserved long noncoding RNA *Dxz4/DXZ4* locus that binds CTCF on the inactive X. In mouse, the hinge region also contains the minisatellite *Ds-TR* adjacent to a region with strong CTCF binding. To determine the role of the elements located in the hinge region in shaping the inactive X chromosome we have constructed allele-specific CRISPR/cas9 deletions and inversions within the hinge region. We report that, while deletion of *Ds-TR* alone or inversion of CTCF binding sites located near *Ds-TR* have no effect on the structure of the inactive X, a 127kb deletion that include *Dxz4* causes a drastic disruption of the superdomains. Both new contacts between superdomains and loss of contacts within superdomains were detected by Hi-C. Analyses of gene expression detected small changes in X inactivation and escape.

2156T

Identifying the functional domains of XIST and their role in X-chromosome inactivation. *T. Dixon-McDougall, C. Brown.* Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

Early in human embryonic development, the nearly 20kb long non-coding RNA, XIST, coats the X chromosome and triggers the stable inactivation of the nearly 1000 genes along its length. XIST was one of the first long non-coding RNAs identified, and similar to other regulatory long non-coding RNAs, acts as a modular scaffold, binding a range of factors that reorganize chromatin marks, silence gene expression and restructure its chromosome of origin. Many XIST interacting proteins have been identified, but without understanding the functional significance of the various XIST domains, how these factors interact with XIST to inactivate a chromosome is still a mystery. To develop a map of the functional domains of XIST and their contributions to the overall process of chromosome inactivation, the functionality of inducible full-length and partial XIST constructs were compared to each other to determine the effect of removing various domains of XIST. As the unique architecture of the X-chromosome itself contributes to X-chromosome inactivation, all of these XIST constructs were integrated into an autosomal FRT site located on chromosome 8p in a male cell line to avoid confounding variables. We used immunofluorescence to examine recruitment of heterochromatic marks; allele-specific pyrosequencing to measure silencing and RNA FISH to examine the ability to localize their chromosome of origin to the perinucleolar compartment. Initially, an assessment of 3-4 kb deletions revealed that the 3' region of XIST contributes to the recruitment of all heterochromatin marks analyzed, distal gene silencing and chromatin reorganization. An internal segment of XIST spanning two of the more strongly conserved repeat sequences was found to contribute to the recruitment of a subset of heterochromatin marks and the chromatin organization. These two domains possessed redundant systems necessary for localizing to their chromosome of origin and forming a unified XIST RNA coated domain. Subsequent finer-scale ~ 1 kb deletions were generated using CRISPR to specifically excise regions with a focus on the tandem repeat regions. These various constructs were measured using the same methods, creating a list of each segment of XIST's contribution to chromosome inactivation. This represents the first full length map of the functional domains of human XIST and an opportunity to identify novel long non-coding RNA functional domains and the factors with which they interact.

2157F

Human elements regulating escape from X-chromosome inactivation are recognized in mouse, and define new regions to interrogate for mechanism.

S. Peeters[1,2], A. Korecki[1,3], E. Simpson[1,3], C. Brown[1,2]. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Molecular Epigenetics Group, Life Sciences Institute, Vancouver, BC, Canada; 3) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada.

X-chromosome inactivation (XCI) epigenetically silences an X chromosome in mammalian females to provide dosage compensation between the sexes. However, a remarkable 15% of genes on the human inactive X chromosome are able to escape inactivation, with another 15% of genes that vary between individuals and tissues. Although genes escaping from XCI have been implicated in playing a role in sexual dimorphisms between men and women, as well as in manifestations of sex chromosome aneuploidies, no clear and complete mechanism behind this escape phenomenon has been identified and experimentally validated. Additionally, studies on human XCI have been limited due to the lack of a developmental model, as human embryonic stem cells have very unstable epigenetic regulation of XCI. As previous reports have described an intrinsic capacity of mouse gene *Kdm5c* to escape from XCI, we targeted a bacterial artificial chromosome (BAC) containing its human homologue *KDM5C* to the normally inactivated *Hprt* locus on the mouse X chromosome, to investigate the conservation of escape elements between mouse and human. Transgenic mice were bred to a strain carrying a deletion in *Xist* thus ensuring that our integrated BACs would always be on the inactive X. Brain, liver, and spleen from adult mice were analyzed for evidence of *KDM5C* escape from the inactive X by measuring promoter CpG island methylation of bisulphite converted DNA, as well as expression levels of RNA by RT-qPCR. Unexpectedly, human *KDM5C* appeared subject to XCI by expression, as well as showing DNA methylation at its promoter, leading us to hypothesize that either *Hprt* was an atypically strong silencing domain, or that human elements regulating escape from XCI are not recognized by mouse. To test these hypotheses we created two additional transgenic mouse strains, the first carrying an integration of mouse *Kdm5c* as a positive control for escape at *Hprt*, and the second carrying another human escape gene, *RPS4X*. Preliminary examination of tissues from these mice suggest that both of them are escaping from XCI, leading us to conclude that the elements controlling escape in humans are recognized by mouse, and that it was a missing regulatory element in the *KDM5C* integration that resulted in it being silenced. This work has added valuable support to using mouse as a model system for studying how human genes escape from XCI, and has highlighted new regions to interrogate in the search for elements regulating this process.

2158W

Attenuation of hedgehog signaling predisposes to nonalcoholic fatty liver disease. *A. Martinez[1], M. Guillen-Sacoto[1], R. Lipinski[2], Y. Abe[1], P. Kruszka[1], K. Weiss[1], J. Everson[3], R. Bataller[4], D. Kleiner[5], K. Sulik[3,6], B. Solomon[7,8], M. Muenke[1].* 1) Medical Genetics Branch, NIH/NHGRI, Bethesda, MD; 2) Department of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI; 3) Bowles Center for Alcohol Studies, University of North Carolina, Chapel Hill, NC; 4) Division of Gastroenterology and Hepatology, Department of Medicine, University of North Carolina, Chapel Hill, NC; 5) Laboratory of Pathology, National Cancer Institute, Bethesda, MD; 6) Carolina Institute for Developmental Disabilities, University of North Carolina, Chapel Hill, NC; Department of Cell Biology and Physiology, University of North Carolina, Chapel Hill, NC; 7) Division of Medical Genomics, Inova Translational Medicine Institute, Falls Church, VA; 8) Department of Pediatrics, Inova Children's Hospital and Virginia Commonwealth University School of Medicine, Falls Church, VA.

Nonalcoholic fatty liver disease (NAFLD) is the most common form of liver disease and a leading cause of liver transplantation in the United States. While Hedgehog (Hh) signaling has been implicated in liver development and the progression of chronic liver injury, a role in liver lipid metabolism and the early stages of NAFLD has not been established. In this study we examined the prevalence of NAFLD in a cohort of patients with overt or microform holoprosencephaly (HPE), a disorder associated with germline mutations disrupting Hh signaling. Compared to the general population, the incidence of NAFLD was significantly higher in this cohort (54.50% of adults, $P < 0.001$; and 33.33% of children, $P = 0.002$), and affected individuals were younger (30.3 years, $P = 0.026$ in adults, 5.5 years, $P = 0.037$ in children). To test the hypothesis that Hh signaling disruption predisposes to liver steatosis, we used a *Gli2* heterozygous null (*Gli2+/-*) mouse model of diet-induced NAFLD. Gli2 is a vertebrate transcription factor necessary for Hh signal transduction. Compared to their wild-type littermates, mutant mice demonstrated increased weight gain when placed on a high-fat (HF) diet and increased liver steatosis induced by both a HF and a methionine- and choline-deficient (MCD) diet. Because MCD diets induce liver steatosis in the presence of weight loss, the increased susceptibility of mutant mice to accumulate hepatic fat is independent of the effects of systemic adipocity characteristic of obesity. In agreement with previous reports, the livers of mutant mice showed decreased hepatic expression of pro-fibrotic and pro-inflammatory genes. Interestingly, mutant mouse also revealed increased expression of PPAR-γ, a potent anti-fibrogenic and anti-inflammatory regulator. Our results demonstrate a direct involvement of Hh signaling in the early stages of NAFLD pathogenesis independent of obesity. We propose a disease mechanism that involves increased liver steatosis with reduced fibrosis, which may have important implications for the management of NAFLD and the development of new therapies targeting this pathway.

2159T

Interaction of saposin C with GCase mutations leads to mutation-dependent and tissue-specific disease phenotype in mice. *Y. Sun[1,3], H. Ran[1], B. Liou[1], B. Quinn[1], Y. Xu[1], G.A. Grabowski[1], W. Zhang[2].* 1) Div Human Gen, Cincinnati Children's Hosp, Cincinnati, OH; 2) Dept of Pathology, Cincinnati Children's Hosp, Cincinnati, OH; 3) Dept of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH.

The defective lysosomal hydrolysis of glucosylceramide (GC) in Gaucher disease is caused by mutations in *GBA1* that encodes acid beta-glucosidase (GCase). Saposin C activates GCase and has protective effects for GCase against proteolysis. To investigate the *in vivo* function of saposin C on *Gba1* mutants, variant Gaucher disease mouse models were generated by backcrossing point mutated GCase mice into saposin C deficient (C-/-) mice. The mutant GCase activities in the resultant mice were reduced by ~50%. The 9H;C* (D409H/D409H;C-/-) mice developed severe visceral GC accumulation and storage cells that were CD68 positive macrophage. In contrast, 4L;C* (V394L/V394L;C-/-) mice, a neuronopathic Gaucher disease model, accumulated GC and glucosylsphingosine (GS) primarily in the brain. Unlike D409V/D409V mice that developed normally for over a year, 9V;C* (D409V/D409V;C-/-) pups had a lethal skin defect. 0S;C* (N370S/N370S;C-/-) mice like N370S/N370S neonates had a lethal skin abnormality. In 9V;C* and 0S;C* pups, ultrastructural studies revealed irregular lamellar bodies in stratum corneum. Neonatal 9V;C* mice had moderate GC accumulation in the liver, brain and lung, whereas 0S;C* neonates showed slight GC accumulation in the lung. 4L;C* and 9H;C* brains showed GS accumulation, whereas 9V;C* and 0S;C* brains did not. These results suggest that lack of saposin C reduced the residual mutant GCase activity and potentiated substrate accumulation in selected tissues. GS appears to be a poor *in vivo* substrate for V394L and D409H GCases. These variant Gaucher disease mouse models provide insights into interaction of saposin C with specific mutant GCase and contribute to understanding the in vivo functional relationships between saposin C and GCase.

2160F

Studies of the functional role of *LYPLAL1* in Nonalcoholic Fatty Liver Disease (NAFLD). *E. Speliotes, Y. Chen, A. Wright, A. Wall, A. Tai, B. Halligan.* Dept of Internal Med, Gastroenterology & Dept of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

The prevalence of nonalcoholic fatty liver disease (NAFLD) has increased with the rise in obesity and is predicted to become the leading cause of liver disease globally by 2020. NAFLD is a spectrum of disease consisting of steatosis (fat, mostly triglycerides), steatohepatitis (NASH; fat and inflammation), and fibrosis/cirrhosis (scarring). The predisposition to develop NAFLD is partially genetic and in population genetic studies we have identified a genetic variant near the gene *LYPLAL1* that predisposes to human population based NAFLD. The protein encoded by the *LYPLAL1* gene, Lysophospholipase-like protein 1 (LYPLAL1) has strong homology to Lysophospholipase protein 1 (LYPLA1, Acyl-Protein Thioesterase APT1). LYPLAL1 and LYPLA1 are thought to be acyl-lysophospholipases which remove palmitate from s-palmitoylated proteins. To test the role of LYPLAL1 on NAFLD we created human cell line and mouse models of NAFLD. Knockdown and overexpression of *LYPLAL1* increased and decreased hepatic steatosis as measured by total cellular triglyceride accumulation respectively in both human HuH7 and HepG2 cell lines. We expressed and purified GST-LYPLAL1 fusion protein, and demonstrated that it can hydrolyze 4-nitrophenyl acetate and 4-nitrophenyl butyrin substrates consistent with its possible role as a thioesterase. Further, we stained wild type HuH7 and HepG2 cells with an anti-LYPLAL1 antibody and found that LYPLAL1 is localized to the perinuclear endoplasmic reticulum of human hepatocellular carcinoma cell lines, the site of protein palmitoylation/depalmitoylation suggesting a role for LYPLAL1 in the regulation of global protein localization. We have generated several Cas9/CRISPR *Lyplal1* knockout mouse mutants including early in frame stop codons and one large deletion leading to missense mutation. We are currently evaluating their phenotype on both normal and high fat/sugar (western) diets to determine the effect of *Lyplal1* knock out on NAFLD and other metabolic conditions. To test the involvement of LYPLAL1 protein palmitoylation we are using the acyl-biotinyl exchange (ABE) approach to exchange biotin groups for palmitate groups and analyze the results by blotting and MS proteomics. By comparing the spectrum of proteins palmitoylated in *LYPLAL1* normal, overexpression and knockdown cells lines and wild type and knockout mouse livers, we hope to identify the targets of LYPLAL1 and better understand its role in NAFLD. .

2161W

An Ashkenazi Jewish-predominant mutation in *CSF2RB* increases risk for Crohn's disease and attenuates GM-CSF signaling. *L. Chuang[1], N. Villaverde[1], K.Y. Hui[2,3], A. Mortha[4], A. Rahman[1], A.P. Levine[5], T. Haritunians[6], S. Ng[3], W. Zhang[3], N. Hsu[1], J. Facey[1], T. Luong[1], H. Fernandez-Hernandez[1], D. Li[6], M. Rivas[7], E.R. Schiff[5], L. Schumm[9], B.M. Bowen[3], Y. Sharma[1,3], K. Ning[1,3,10], C. Abraham[3], M. Daly[7,12], R.H. Duerr[11], A.W. Segal[5], M. Merad[4], D.P. McGovern[6], I. Peter[1], J.H. Cho[1,8].* 1) Genetics and Genomic sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 3) Section of Digestive Diseases, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT; 4) Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Centre for Molecular Medicine, Division of Medicine, University College, London, UK; 6) F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 7) Department of Medical and Population Genetics, Broad Institute, Cambridge, MA; 8) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 9) Department of Health Studies, University of Chicago, Chicago, IL; 10) Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA; 11) Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh; 12) Department of Genetics, Harvard Medical School, Boston, MA.

Crohn's disease (CD) has the highest prevalence in Ashkenazi Jewish (AJ) populations. We sought to identify rare, CD-associated frameshift variants of high functional and statistical effects. We performed exome sequencing of AJ CD cases, and genotyped 224 frameshift variants using custom array genotyping in an AJ case-control cohort. Quantile-quantile (Q-Q) analysis of the 88 frameshift mutations was evaluated in order to assess the validity of the null distribution assumption. It revealed only two markers demonstrating significantly greater association evidence than predicted, namely the *NOD2* frameshift mutation and a frameshift mutation in the *CSF2RB* gene (P value 8.52 x 10^{-7}), which encodes for the colony stimulating factor 2 receptor, beta chain. The *CSF2RB* mutation involves deletion of a single guanine base, which is predicted to cause a frameshift at codon 708 of *CSF2RB* (colony-stimulating factor 2 receptor, beta chain). This would result in the expression of a 729-amino acid sequence, compared to 897 residues encoded by the canonical transcript. Exome-sequencing and array-based genotyping was performed in 1477 AJ CD cases and 2614 AJ controls. Replication genotyping of a *CSF2RB* (colony-stimulating factor 2, receptor beta) frameshift was performed in 1515 CD cases and 6980 controls. Intestinal *CSF2RB* gene expression and GM-CSF responsive cells were defined by CyTOF. Wild-type and mutant CSF2RB function was compared in cellular transfection and primary monocyte studies. We observed nominal evidence for CD association for the *CSF2RB* frameshift (P value 8.52x10^{-7}) in the discovery cohort, which was significantly replicated (combined P value 3.42x10^{-6}). In the intestine, robust GM-CSF induction of STAT5 phosphorylation is observed, with lesser induction of pERK and pAKT. Co-transfection of wild-type and mutant CSF2RB displayed decreased pSTAT5 with GM-CSF stimulation, consistent with a dominant negative effect. Monocytes from heterozygous frameshift carriers, present in 6% of AJ CD cases, demonstrated diminished GM-CSF responses, with markedly decreased aldehyde dehydrogenase enzyme activity associated with immune tolerance. Our findings define a primary role of diminished GM-CSF signaling and impaired innate immunity in CD pathogenesis.

2162T

Novel variants identified in Chinese ankylosing spondylitis families by whole-exome sequencing. *J. Liu[1], Z. Niu[1], Y. Li[1], Y. Wang[1], Y. Ma[1], H. Zhang[1], L. Jin[1], H. Zou[2], J. Wang[1].* 1) School of Life Science , Fudan University, Shanghai, China; 2) Division of Rheumatology, Huashan Hospital, Shanghai, China.

Ankylosing spondylitis (AS) is the main subtype and a main outcome of spondyloarthritis, the susceptibility to which is mainly determined by genetics. HLA-B27 shows the strongest reported association with AS, while it can only explain about a quarter of the overall genetic susceptibility to AS. On the other hand, previous studies reported that HLA-B27 positive individuals who had a first-degree relative infected with AS have more (6-16 times) susceptibility to developing AS themselves than those HLA-B27 positive ones but without AS family history. It indicated that some other genes have a strong effect on the development of AS in individuals with family history. Therefore, it is required to find high risk variants in patients with family history by next-generation sequencing.   Three unrelated AS Chinese families (6 patients were all HLA-B27 positive) were recruited in the first stage of the research. Additional 994 Chinese subjects with AS and 999 ethnically matched controls were recruited in the second stage. The diagnosis of AS was performed according to 1984 modified New York criteria of AS. After performing whole-exome sequencing in these AS families, we selected five gene variants (*TTN, ADAM2, RNF123, MGP,* and *LRR1*) to be validated in the second stage according to the following rules: (I) It was a functional nonsense mutation in coding region; (II) MAF in 1k genomes data <5%; (III) It was related to inflammation or bone information. Three of these selected variants (located in TTN, RNF123 and ADAM2) showed statistical significance between cases and controls. The odds ratio and chi square p-value of TTN variant were 5.04 and 0.02 respectively. As to RNF123 and ADAM2, their variants were not found in both the 1k genomes data base and all the controls. All individuals carrying the minor alleles were replicated by Sanger Sequencing. In conclusion, we performed a whole-exome sequencing to look for the rare variants from three AS families and validated five candidate variants in a 994 cases vs. 999 controls test. Three rare variants showed significant difference between AS and controls. In addition, all the three variants were nonsense in coding region and would play functional roles in AS.

2163F

BREATH: A web-accessible database of normal human and mouse lung development generated by the Molecular Atlas of Lung Development Program (LungMAP). *R. Clark[1], N.C. Gaddis[1], J.L. Levy[1], H. Pan[1], Y. Xu[2], M. Duparc[1], G.P. Page[1], J.A. Whitsett[2], S.M. Palmer[3], LungMAP Data Coordinating Center.* 1) RTI International, Research Triangle Park, NC; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC.

The lung is a complex organ with high cellular heterogeneity, and research is needed to define the interactive gene networks and dynamic crosstalk among multiple cell types that coordinate normal lung development. Significant knowledge gaps exist in the understanding of lung development from late fetal to perinatal stages and through early childhood, a critical period when the diverse lung cells go through terminal differentiation and maturation and when alveoli form. Coordination and integration of information from disparate assays and experimental approaches is critically needed to help the pulmonary research community uncover the molecular determinants of lung development. The goal of LungMAP is to build an open-access resource providing a comprehensive molecular atlas of late-stage lung development in humans and mice, making otherwise dispersed data, reagents, and protocols freely available to the research community. We have created the Bioinformatics REsource ATlas for the Healthy lung (BREATH) database, applying novel data management and bioinformatics approaches to manage high-throughput multidimensional experimental data. BREATH is built on a triple store database backbone, integrated with anatomical ontologies for lung development, and provides access to novel web-based tools for the analysis and visualization of data generated by the four Research Centers and Human Tissue Core of the LungMAP Consortium. The current version of BREATH contains confocal immunofluorescence images, *in situ* hybridization images, histological images, and 3D confocal and uCT images of developing mouse and human lung at several time points, as well as single-cell RNAseq, proteomic, and lipidomic data from mouse and human lung cells. The LungMAP website (www.lungmap.net) provides an entry portal to the BREATH database and tools for exploring and interacting with the lung images and omics data. Future versions of BREATH will incorporate additional data types, including metabolomic and epigenetic data, and novel tools for cross-datatype analysis. A better understanding of the basic molecular pathways that regulate normal lung development will enable development of innovative treatments that successfully manipulate, enhance and improve lung injury repair and regeneration.

2164W

Pituitary hormone deficiency: Role of chromatin architecture in disease pathophysiology. *A. Daly, L.Y.M. Cheung, S.A. Camper.* Human Genetics, Michigan, Ann Arbor, MI.

Congenital hypopituitarism occurs in about 1/4000 births and can be life-threatening if not treated effectively. There are many genetic causes, including mutations in *PROP1* and *POU1F1*, but most patients have no molecular diagnosis. PROP1 induces POU1F1 expression, and POU1F1 organizes chromatin and drives the differentiation of cells that express thyrotropin, growth hormone or prolactin. To understand the mechanism of POU1F1 action, we utilized two POU1F1-expressing cell lines that represent sequential steps in pituitary development: hormone negative and hormone positive. We identified changes in chromatin architecture and factors that are exclusively associated with hormone expression using the Assay for Transposase Accessible Chromatin-Sequencing (ATAC-seq) and RNA-seq. There were numerous changes in chromatin architecture. *Nr5a1* encodes a transcription factor specific for the gonadotrope lineage, and it shifts to closed chromatin in more differentiated cells. This is consistent with the idea that competition between POU1F1 and NR5A1 establishes separate cell fates. *Rax* encodes the retinal and anterior neural fold homeobox transcription factor necessary for eye and hypothalamic development, and it also coverts to closed chromatin in the more differentiated cells. This may represent differentiation of a basic head placode towards pituitary gland fate instead of the eye fate and/or distinguishing hypothalamus from pituitary. This idea is supported by the discovery that *Prop1* deficiency permits ectopic expression of genes in the pituitary gland that are normally expressed in other head placodes and/or the hypothalamus. Application of Gene Set Enrichment Analyses (GSEA) to differentially accessible chromatin revealed that genes involved in cell proliferation and cell migration shift from open to closed chromatin as differentiation ensues. This suggests that the cell lines represent the *Pou1f1* expressing cells near the stem cell niche and those that have migrated into the parenchyma of the gland and differentiated to produce hormones. The RNA-Seq analysis will reveal candidate genes that work together with POU1F1 to alter chromatin architecture and promote differentiation and expression of the hormone genes. In conclusion, we have advanced our understanding of the mechanism of PROP1 and POU1F1 action and identified candidate genes for CPHD of unknown etiology.

2165T

The role of FREM/FRAS proteins in the development of diaphragmatic sac hernias. *V.K. Jordan[1], T.F. Beck[2,7], A. Hernandez-Garcia[2], P. Khundert[2], B.J. Kim[2], S.N. Jhangiani[2,3], T. Gambin[2], M. Starkovich[2,3], A.H. Li[4], D. Muzny[2,3], K. Lally[6], R. Gibbs[2,3], E. Boerwinkle[4], J.R. Lupski[2,3,5], D.A. Scott[1,2].* 1) Molecular Physiology and Biophysics, Baylor College of medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Department of Pediatric Surgery, University of Texas Health Science Center, Houston, TX, USA; 7) Current affiliation: NHGRI, National Institutes of Health, Bethesda, MD, USA.

Congenital diaphragmatic hernia (CDH) occurs when the abdominal viscera—the liver, gallbladder, stomach, and/or intestines--enter into the thoracic cavity through an abnormal opening in the diaphragm. This life-threatening birth defect affects 1 in 3,000 newborns and accounts for ~8% of all major congenital anomalies. In ~20% of human CDH cases, the herniated viscera are encapsulated by a thin layer of membranous tissue referred to as a hernial sac. We have previously shown that isolated sac CDH in humans and mice can be caused by recessive pathogenic variants in *FREM1*, which encodes an extracellular matrix protein. In the extracellular matrix, FREM1 forms a self-stabilizing ternary complex with two related proteins—FREM2 and FRAS1—which plays an important role in cell adhesion and intercellular signaling. To determine if pathogenic variants in *FREM2* and *FRAS1* can also cause CDH, we performed exome sequencing on 69 patients with CDH. Four of these individuals (6%) carried rare, putatively deleterious compound and/or double heterozygous changes in these genes. Three of these individuals carried a c.4031G>A, p.Arg1344His change in *FREM2*. The frequency of this allele among Caucasians in our CDH cohort was statistically greater than that seen in ethnically matched control populations P < 0.0001. In support of *FREM2*'s role in the development of CDH, we also found sac hernias in 8.2% (5/61) of FREM2-deficient mice. We also found one rare FRAS1-deficient mouse with sac CDH (1/102, 1%). The sac hernias in these mice were indistinguishable from those seen in FREM1-deficient mice. In all cases, these hernias were located in the anterior region of the diaphragm and contained the gallbladder and a pedunculated mass of liver tissue. Further analysis revealed that these sac hernias form when the migrating muscular components of the diaphragm fail to meet at the midline behind the sternum. We went on to demonstrate that GATA4, a retinoic acid responsive transcription factor implicated in the development of sac CDH, modulates the transcriptional expression of *Fras1* in diaphragmatic mesothelial cells. We conclude that pathogenic variants in *FREM1*, *FREM2* and *FRAS1* may increase the risk of developing sac CDH in both humans and mice. Our data also suggests an important link between retinoic acid signaling and the maintenance of the extracellular matrix during diaphragm development.

2166F

Cushing's Disease associated with *USP8* mutations. *F.R. Faucz[1], A. Tirosh[1], A. Berthon[1], L.C. Hernández-Ramírez[1], N. Settas[1], A. Angelousi[1], R. Correa[1], P. Chittiboina[2], N. Pankratz[3], J. Lane[3], A. Dimopoulos[4], J. Mills[4], M. Lodish[1], C.A. Stratakis[1].* 1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA; 2) Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA; 3) Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, MN, USA; 4) Epidemiology Branch, Division of Intramural Population Health Research, NIH, Rockville, MD, USA.

   Introduction: Somatic mutations in the ubiquitin-specific protease 8 gene (*USP8*), resulting in increased activity of the epidermal growth factor receptor (EGFR) signaling pathway, constitute the most common genetic alteration in patients with Cushing disease (CD). However, the occurrence and possible pathogenic role of germline *USP8* variants has not been assessed before. **Methods**: To examine the prevalence and clinical correlations of mutations in *USP8*, the complete *USP8*-coding and surrounding intronic regions were sequenced in 42 pediatric CD patients, followed up at the NIH, by Sanger sequencing of germline and tumor DNA. Clinical characteristics were compared among patients with a predicted by in silico as damaging germline variant of unknown significance (VUS) and/or somatic *USP8* mutations, vs. those without *USP8* changes (controls). **Results**: Mean age at diagnosis for the full cohort was 13.7±3.4 years (range, 6.1-18.7). Twenty one patients (50%) had *USP8* mutations, 8 with germline VUS and 13 had somatic mutations. All the somatic mutations that were not present at the germline level were located in the previously described hotspot region. Patients harboring a germline VUS had longer disease duration (3.1±1.6 vs. 1.8±1.6 years, p=0.02) and required treatment following operation more often (50% vs. 14.3%, p=0.045) compared with the controls. Patients with somatic mutations were predominantly females (84.6% vs. 47.6%, p=0.03). The biochemical profiles of somatic or germline mutations were similar to those of the controls. Adenoma parameters, including size and invasiveness were also similar. The time between disease onset and clinical diagnosis was longer for patients with germline *USP8* VUS (Breslow test, p=0.04). **Discussion**: Herein we report clinical data of pediatric CD patients seen in the NIH, according to presence and type of *USP8* mutations and/or germline VUS. We showed that patients harboring somatic *USP8* mutations were predominantly female. Moreover, patients with germline *USP8* VUS were diagnosed later. This may suggest that the disease is more indolent among CD patients with germline VUS of *USP8*. Further studies are needed to clarify how germline *USP8* VUS could affect EGFR signaling, and whether they could possibly contribute to pituitary tumorigenesis.

2167W

***Nr2e1* influences exploratory behavior but it is not necessary for the gonadotroph and somatotroph pituitary differentiation of zebrafish (*Danio rerio*).** *C. Caetano da Silva[1], P. S. Page-McCaw[2], W. Chen[2], L. R.S. Carvalho[1].* 1) Internal Medicine Department, Endocrinology Discipline, University of São Paulo Medical School, São Paulo, SP, Brazil; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States.

   Nr2e1, also known as *Tlx,* is an orphan nuclear hormone receptor that acts as a transcriptional repressor through physical association with transcriptional corepressors. Mutations in *Nr2e1* in mouse demonstrate that it is important for neural stem cell maintenance and proliferation. The loss of *Nr2e1* results in retinopathy, severe limbic defects, reduction in the volume of superficial cortical layers, also mice mutant for *Nr2e1* are aggressive and famously difficult to handle, previously being given the gene name *fierce*. Due to the fact that zebrafish is an animal model of easy handle, fast development and external reproduction, we used the genome editing tool, CRISPR/Cas9, to produce a permanent, heritable mutation in *nr2e1*. As seen in mice, *nr2e1* homozygous mutants are viable and both male and female mutants mate and produce viable offspring. In order to probe the behavioral phenotype of zebrafish, we have developed a novel assay to assess exploratory behaviors in this model. In the larval stages zebrafish are a non-aggressive and non-territorial species. However, aggressiveness is correlated with fearlessness, this fact prompted us to developed an assay that measures exploratory behavior. We observe that mutant animals are more exploratory than wildtype controls. Such as *Nr2e1*, *Sox2* is essential for neural stem cell maintenance and neogenesis in the mouse brain and it is well characterized stem cell marker in the pituitary. *nr2e1* is also expressed in the zebrafish pituitary, but its role in the pituitary development and terminal hormone expression has not been characterized. We analyzed Sox2 and Nr2e1 expression by immunofluorescence in the whole zebrafish brain at 10 dpf, 45 dpf, and in adults. While both proteins were present in the pituitary during development, surprisingly, they do not co-localize. We are currently investigating the role of *nr2e1* in zebrafish pituitary development.

2168T

Elevated testosterone levels in mice deficient for the chromatin modifying enzyme Kdm5c. *H. Weimar[1], K. Lewallen[1], Y. Shi[2], J. Xu[1].* 1) Dept. of Integrative Physiology and Neuroscience, Washington State University , Pullman, WA; 2) Boston Children's Hospital, Division of Newborn Medicine, Boston, MA.

The X-linked gene *Kdm5c* encodes a histone demethylase that is involved in epigenetic regulation of gene expression. By removing the methylation modifications from lysine 4 on histone H3, the Kdm5c enzyme suppresses the transcription of its targeted genes. When *Kdm5c* is mutated, its protein products are either absent or function poorly, resulting in up-regulation of genes normally repressed by this demethylase. Mutations of *KDM5C* cause many symptoms, including intellectual disability, aggression and disruptive behavior, as well as compromised male fertility. The goal of this ongoing project is to uncover the molecular basis for the disease phenotypes. We examined mice with a germline transmitted constitutive deletion mutation of *Kdm5c*, i.e. *Kdm5c* deficient mice (Kdm5c[Def]). They underwent behavioral testing, including a test of aggression. In comparison to wild type (WT) mice, Kdm5c[Def] mice displayed higher levels of aggression, reduced sociability, and memory deficits. Consistent with the finding of high aggression, the serum and testicular concentrations of testosterone were elevated in the Kdm5c[Def] mice, which in turn led to the up-regulated expression of androgen responsive genes in the brain, as shown by RNA-seq analysis. To investigate the cause for the increase of testosterone, expressions of steroidogenic genes were quantified in testes with RT-qPCR and found to be either *decreased* in the Kdm5c[Def] mice (e.g. *Hsd3b6*, *Hsd17b1,* and *Hsd17b3*) or similar between the two genotypes, making these genes less likely to be responsible for the increased levels of testosterone. Only one steroidogenic gene, *Srd5a1*, showed higher levels of mRNA expression in Kdm5c[Def] mice, which was concordant with the higher levels of H3K4 methylation at its promotor in the mutant mice. Histological examination revealed fewer testosterone producing Leydig cells in the Kdm5c[Def] mice as well as irregularities in spermatogenesis. The findings of this study help us to better understand the role of Kdm5c and epigenetic chromatin re-modeling in steroid production with direct relevance to spermatogenesis, male infertility, and behavior. Ultimately, this study might suggest chromatin targeted new treatments for male infertility as well as neuropsychiatric symptoms, such as aggression, that are related to abnormal testosterone levels. (Supported by an NIH R01 MH096066 to YS & JX).

2169F

Identification and characterization of a novel variant of human lysyl oxidase-like 3 (*LOXL3***).** *Y. Kim, A. Sharma-Bhandari, C. Li.* Dept Biochem, Wonkwang Univ Sch Med, Iksan, Jeonbuk, South Korea.

The human lysyl oxidase-like 3 (*LOXL3*) functions as a copper-dependent amine oxidase toward collagen and elastin. The LOXL3 protein contains four scavenger receptor cysteine-rich domains in the N-terminus in addition to the C-terminal characteristic domains of the lysyl oxidase (LOX) family, such as a copper-binding domain, a cytokine receptor-like domain and residues for the lysyl-tyrosyl quinone cofactor. Using BLASTN searches, we identified a novel variant of *LOXL3* (termed *LOXL3-sv2*), which lacked the sequences corresponding to exons 4 and 5 of *LOXL3*. The *LOXL3-sv2* mRNA is at least 2,398 bp in length, encoding a 608 amino acid-long polypeptide with a calculated molecular mass of 67.4 kDa. The deletion of exons 4 and 5 do not change the open-reading frame of LOXL3 but results in deletion of the SRCR domain 2. The recombinant LOXL3-sv2 protein showed a BAPN-inhibitable amine oxidase activity toward collagen type I. In RT-PCR analysis, *LOXL3-sv2* was detected in all human tissues tested, along with *LOXL3* and *LOXL3-sv1*, a previously identified variant of *LOXL3*. These findings indicate that the human *LOXL3* gene encodes at least three variants, LOXL3 LOXL3-sv1 and LOXL3-sv2, all of which function as amine oxidases.

2170W

Analysis of the interplay of lysosomal/autophagosomal pathways in osteoclasts towards correction of osteopetrosis. *C.E. Bauer[1,2], M. Pata[2], J. Vacher[1,2].* 1) Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada; 2) Clinical Research Institute of Montreal, Department of Medicine, University of Montreal, Montreal, Quebec Canada.

Osteopetrosis is a disease characterized by the failure of osteoclasts (OCs) to resorb bone. As a consequence, bone mass is increased and bone marrow space is reduced; leading to premature death. The grey-lethal (*gl/gl*) Ostm1 null mouse model mimics severe human forms of osteopetrosis, and although the exact function of Ostm1 in osteoclasts is not fully understood, it appears to influence OC ruffled border (RB) formation. The RB is a platform for proton/protease excretion as well as bone material uptake and is formed through fusion of secretory lysosomes with the bone-apposed plasma membrane. Interestingly, our results implicate Ostm1 in autophagocytic clearance via the lysosomal pathway in neurons. Thus, it is imperative to assess Ostm1 role within the OC and the interplay of autophagy and lysosome biogenesis in OC RB formation. To assess these facets, we designed a transgene to stimulate the lysosomal/autophagy pathways in mature OCs. The expression of the Tfeb transcription factor, a master regulator of the autophagocytic/lysosomal pathways, was targeted to mature OCs with the Capthepsin K (Ctsk) promoter. To determine whether Ostm1's function is critical in the cellular clearance of autophagosomes, the expression of key autophagosomal/lysosomal trafficking and flux molecular regulators have been assessed in transgenic and non-transgenic wild-type (wt) mice. *In vitro* analyses of mature OC samples, cultured from transgenic wt mice, indicate that there is a significant upregulation of autophagosomal genes (*Atg5* and *Rab7*) when compared to non-transgenic controls. This correlates with the LC3II/I protein expression ratio, a marker of autophagy. Additionally, transgenic mice have higher levels of *Atp61v1*, a gene associated with OCs activity. Together, these data demonstrate that Tfeb upregulation can stimulate autophagy/lysosomal pathways in OCs, thereby furthering the understanding of Ostm1's role in RB function. Concomitantly, assessing correction of OC maturation/activation and the osteopetrotic phenotype is ongoing in Tfeb *gl/gl* transgenic mice, *in vivo*. Consequently, these studies will add to our understanding of the lysosomal/autophagocytic interplay in osteoclasts and could result in the development of new therapeutic treatments for osteopetrosis.

2171T

Loss of the lipid phosphatase Inpp4b in mature osteoblasts affects bone homeostasis. *W. Liu[1,2], L. Saad[1,2], M. Pata[2], J. Vacher[1,2].* 1) Division of Experimental Medicine, McGill University, Montreal, QC, Canada; 2) Department of Medicine, University of Montreal, Clinical Research Institute of Montreal, Montreal, QC, Canada.

Bone homeostasis relies on the tightly coordinated control of the bone-forming osteoblasts and the bone-resorbing osteoclasts. Disruption in the crosstalk between osteoblasts and osteoclasts can lead to different bone diseases such as osteoporosis and osteopetrosis, which affect over 1 million Canadians. Previous findings in our lab revealed a dysregulation in inositol-polyphosphate-4-phosphatase type II (Inpp4b) in the murine grey-lethal model of severe osteopetrosis. *Inpp4b*, expressed in both osteoblasts and osteoclasts, has been shown to be a negative regulator of osteoclastogenesis and a prognostic locus of osteoporosis. In order to elucidate the function of *Inpp4b* in mature osteoblasts and in bone homeostasis, we generated a floxed *Inpp4b* allele that would be deleted specifically in mature osteoblasts using human Osteocalcin-Cre in *Inpp4b*[lox/lox]; hOc-Cre (CKO) mice. Initial *in vivo* analyses of CKO mice show an increase in bone density at 8 weeks of age, reaching a control level by 4 months. *Ex vivo,* differentiated osteoblasts from calvaria of neonatal CKO mice display a decreased mineralization level compared to controls. Subsequent gene expression analyses reveal a downregulation of the mature osteoblast marker osteocalcin, concomitant with an increase in the mineralization inhibitor, osteopontin. These findings prompted us to determine the mechanism by which *Inpp4b* can regulate mineralization. Since autophagy has previously been described to be involved in mineralization, we examined the expression of the autophagy-related genes, *Atg5* and *Atg7*, in mature CKO osteoblasts. Interestingly, we observed a decrease in expression of these genes as well, suggesting an autophagy-related role for *Inpp4b* in mineralization. Taken together, these results demonstrate that the specific deletion of *Inpp4b* in mature osteoblasts can modulate bone homeostasis. In the future, we will characterize how *Inpp4b* can affect signalling pathways regulating osteoblast/osteoclast crosstalk. As a result, this study may yield future therapeutic targets in the prevention and treatment of different bone pathologies.

2172F

ERAP1 deficient mice phenocopy the major findings of humans with ankylosing spondylitis, including sacroiliitis, osteoporosis, bridging osteophytes, and sacral and vertebral fusions. *D.P.W. Rastall[1], S. Raehtz[2], Y. Pepelyayeva[1], F.L. Collins[2], P. O'Connell[1], S. Godbehere-Roosa[1], M. Henderson[3], L.R. McCabe[2,3], A. Amalfitano[1,4].* 1) Department of Microbiology and Molecular Genetics, College of Osteopathic Medicine, Michigan State University, East Lansing, MI 48824, USA; 2) Department of Physiology, Michigan State University, East Lansing, MI 48824, USA; 3) Department of Radiology, Michigan State University, East Lansing, MI 48824, USA; 4) Department of Pediatrics, College of Osteopathic Medicine, Michigan State University, East Lansing, MI 48824, USA.

In humans, several genome-wide association studies (GWAS) have identified variants in the endoplasmic reticulum aminopeptidase 1 (ERAP1) gene that alter the risk for development of ankylosing spondylitis (AS). We have demonstrated that high-risk AS ERAP1 variants directly stimulate human immune cells to cause secretion of IL-1β, IL-6, and TNF-α. Additionally ERAP1-/- mice had increased activation of NK and NK-T cells and increased production of IL-12 and MCP-1 in response to pathogen recognition and stimulation with TLR agonists. We also demonstrated a critical role for ERAP1 in adaptive immunity *in vivo* by demonstrating that ERAP1 could completely alter the immunodominant T-cell response, by creating and destroying peptides. Importantly, we also demonstrated that high-risk AS ERAP1 variants could over-trim peptide ligands resulting in decreased amounts of MHC-I on the surface of immune cells. As pro-inflammatory states have been linked with bone loss, we hypothesized that ERAP1 may also play a role in bone homeostasis. To test this hypothesis we investigated the role of ERAP1 in bone homeostasis using ERAP1-/- mice. Utilizing micro-CT based imaging, we demonstrate that aged ERAP1-/- mice (as compared to control mice) progressively develop many of the hallmark signs of human AS, including increased sacroiliitis, osteoporosis, erosions in the axial spine, bridging syndesmophytes (ankylosis), and bony fusion of the sacrum and vertebrae. These findings were confirmed with blinded scoring of both 2D radiographs, blinded scoring of 3D isosurface reconstructions, and quantitative measures of bone morphology. ERAP1-/- mice also accurately reproduce the global microscopic bone remodeling observed in AS patients including significantly decreased trabecular bone mineral content, decreased bone volume fraction, decreased trabecular thickness, and increased trabecular spacing in the sacrum, spine, and femur as confirmed by quantitative morphometry (utilizing GE Healthcare Micro View software). Together, these results demonstrate that ERAP1-/- mice can be used to pave the way to future studies exploring interventions that may prevent these bony pathologies from occurring. Furthermore, these results confirm the findings of GWAS that first identified the possible involvement of ERAP1 in AS, by identifying an animal model that demonstrates a role for ERAP1 in bone pathologies of the axial skeleton.

2173W

Mutations in *MBTPS2* cause X-linked osteogenesis imperfecta and demonstrate a fundamental role for regulated intramembrane proteolysis in bone development. *V. Shotelersuk[1], U. Lindert[2], W.A. Cabral[3], S. Ausavarat[1], S. Tongkobpetch[1], K. Ludin[4], A.M. Barnes[3], P. Yeetong[1], M. Weis[5], B. Krabichler[6], C. Srichomthong[1], E. Makareeva[7], A.R. Janecke[6,8], S. Leikin[7], B. Röthlisberger[4], M. Rohrbach[2], I. Kennerknecht[9], D.R. Eyre[5], K. Suphapeetiporn[1], C. Giunta[2], J.C. Marini[3].* 1) Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; 2) Division of Metabolism, Connective Tissue Unit and Children's Research Center, University Children's Hospital Zurich, Zurich, 8032 Switzerland; 3) Bone and Extracellular Matrix Branch, National Institute of Child Health and Human Development, National Institutes of Health, 20892 Bethesda, Maryland, USA; 4) Center for Laboratory Medicine, Department of Medical Genetics, Kantonsspital Aarau, 5001 Aarau, Switzerland; 5) Department of Orthopedics and Sports Medicine, University of Washington, 98195 Seattle, Washington, USA; 6) Division of Human Genetics, Medical University of Innsbruck, 6020 Innsbruck, Austria; 7) Section on Physical Biochemistry, National Institute of Child Health and Human Development, National Institutes of Health, 20892 Bethesda, Maryland, USA; 8) Department of Pediatrics I, Medical University of Innsbruck, 6020 Innsbruck, Austria; 9) Institute of Human Genetics, Westfälische Wilhelms University, 48149 Münster, Germany.

Osteogenesis imperfecta (OI) is a collagen-related heritable bone dysplasia. Dominant OI is caused by defects in type I collagen or *IFITM5*, while recessive forms result from deficiency of proteins that interact with collagen for modification, folding or crosslinking. We have identified the first X-linked form of OI caused by a defect in regulated intramembrane proteolysis (RIP). One type of RIP involves sequential cleavage of substrates by site-1 (S1P) and site-2 protease (S2P), encoded by *MBTPS1* and *MBTPS2*, respectively. S1P and S2P are Golgi membrane proteases that cleave regulatory proteins transported from the ER membrane in response to stress or decreased sterol metabolites, releasing mature N-terminal fragments that enter the nucleus to activate gene transcription. In two pedigrees with moderately severe OI, linkage analysis and next generation sequencing identified novel missense mutations in *MBTPS2*, predicting p.N459S and p.L505F substitutions located in or near the S2P NPDG metal-coordinating motif. Neither *MBTPS2* transcripts nor S2P protein were decreased in proband fibroblasts and osteoblasts. However, proband and *Mbtps2*-deficient CHO cells co-transfected with mutant *MBTPS2* and reporter constructs indicate deficient cleavage or activation of RIP substrates OASIS, ATF6 and SREBP. Consistent with diminished OASIS signalling, proband fibroblasts have significantly reduced type I collagen secretion (20-73% of normal control cells). Also, proband fibroblast matrix contains collagen with decreased mature crosslinks. Critically, Proband bone type I collagen contained less than half the normal level of hydroxylated Lys87, the residue crucial for crosslinking, and is associated with decreased Lysyl Hydroxylase 1 (LH1) levels in proband osteoblasts. In addition, proband urinary LP/HP crosslink ratios are increased. These findings suggest abnormal collagen crosslinking undermines bone strength in X-linked OI. S2P-deficient osteoblasts also demonstrated broadly defective differentiation with decreased expression of transcripts related to osteoblast maturation, including *ALPL*, when induced to differentiate in culture. Transcripts encoding OASIS and SMAD4, which together upregulate expression of matrix-associated genes, were also significantly reduced in proband versus normal control osteoblasts. These are the first human studies to demonstrate the fundamental role of RIP in bone development, in addition to its function in cholesterol metabolism.

2174T

Unraveling the *TBX5* regulatory landscape. *F. Petit[1,2], S. Manouvrier-Hanu[2], N. Ahituv[1,3].* 1) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, CA, USA; 2) Univ. Lille, CHU Lille, EA 7364 - RADEME, F-59000 Lille, France; 3) Institute for Human Genetics, University of California San Francisco, CA, USA.

Human limb malformations are frequent congenital abnormalities with a prevalence of 1/500 births. So far, very few gene mutations leading to isolated (non-syndromic) limb malformations have been found. There is growing evidence that disruption of gene regulatory elements may play a major role in their pathogenesis. Discovering these elements and understanding their function still remains a major challenge. Here, we set out to characterize the role of *TBX5* regulatory elements in human limb malformations. *TBX5* encodes a transcription factor that plays a crucial role in forelimb bud initiation. Its inactivation in animal models abolishes forelimb skeletal formation. In humans, *TBX5* heterozygous mutations or rearrangements are responsible for the association of preaxial upper limb malformations and congenital heart disease called Holt-Oram Syndrome (HOS, MIM 142900). However, screening for *TBX5* coding mutations remains negative in 15-30% of typical HOS patients. Isolated bilateral amelia of the upper-limbs (MIM 601360) is a very rare condition reminding of the phenotype for *Tbx5* inactivation in animal models. Recurrent cases have been described in some families, suggesting a genetic component. We hypothesize that in these two conditions, disruption of *TBX5* regulatory element/s could cause these phenotypes. To gain insights into *TBX5* regulation, we used various genomic datasets (ChIP-seq, ChIA-PET, comparative genomics) coupled with mouse transgenic enhancer assays to identify and characterize novel *TBX5* regulatory elements. In addition, using *in situ* hybridization and CRISPR/Cas9 genome editing, we are characterizing the function of a *Tbx5* antisense non-coding RNA. To better understand how *Tbx5* regulates other genes, we are performing ChIP-seq for it on mouse E10.5 forelimbs. Finally, to translate our work to human limb malformations, we are screening patients with HOS that have no *TBX5* coding mutations and patients with amelia of the upper limbs using a candidate sequencing approach and whole-genome sequencing. Combined, our work will unravel the gene regulatory landscape of *TBX5* and how its alteration can lead to human disease. **Acknowledgments:** Sequencing was provided by the University of Washington - Center for Mendelian Genomics (UW-CMG) and was funded by the National Human Genome Research Institute and the National Heart, Lung and Blood Institute grant 2UM1HG006493 to Drs. Debbie Nickerson, Michael Bamshad, and Suzanne Leal.

2175F

***Sc65*-null mice provide evidence for a novel endoplasmic reticulum complex regulating collagen lysyl hydroxylation.** *R. Morello[1,5], M.E. Heard[1], R. Besio[1], M.A. Weis[2], J. Rai[2], D.M. Hudson[2], M. Dimori[1], S.M. Zimmerman[1], J.A. Kamykowski[1], W.R. Hogue[3], F.L. Swain[3], M.S. Burdine[4], S.G. Mackintosh[4], A.J. Tackett[4], L.J. Suva[3], D.R. Eyre[2].* 1) Department of Physiology & Biophysics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA; 3) Department of Orthopaedic Surgery, Center for Orthopaedic Research, University of Arkansas for Medical Sciences, Little Rock, AR; 4) Department of Biochemistry & Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR; 5) Division of Genetics, University of Arkansas for Medical Sciences, Little Rock, AR.

Sc65 (Synaptoneal Complex 65) is an endoplasmic reticulum protein that belongs to the Leprecan family which include the prolyl 3-hydroxylases (P3H1, P3H2, P3H3) and cartilage associated protein (CRTAP). We and others have shown that mutations in both *CRTAP* and *LEPRE1* (encoding P3H1) cause recessive forms of Osteogenesis Imperfecta. Sc65 and CRTAP are both non-enzymatic proteins and share high homology suggesting Sc65 may also function in bone homeostasis. We have demonstrated in two different mouse genetic models that loss of Sc65 results in low bone mass. To study Sc65 function in bone, co-immunoprecipitation (co-IP) of Sc65 candidate interactors in mouse fibroblasts followed by mass spectrometry was performed. These experiments identified several fibrillar procollagen α-chains as likely substrates of Sc65 supporting the idea that Sc65 plays a role in collagen modification, similar to other Leprecans. At the biochemical level, mass spectrometry of type I collagen peptides showed severe under-hydroxylation at helical cross-linking sites K87 and K930/933 in collagen α1(I) and α2(I) chains both from bone and skin which are known LH1 preferred substrate residues but with no effect on sites of prolyl 3-hydroxylation. Direct co-IP assays showed Sc65 interaction with lysyl-hydroxylase 1 (LH1, Plod1), prolyl 3-hydroxylase 3 and cyclophilin B. Western blot revealed dramatic reduction of LH1 and P3H3 in primary osteoblasts and skin fibroblasts from *Sc65*-KO mice. Size exclusion chromatography confirmed that Sc65 and P3H3 form a stable complex in the ER that affects the activity of lysyl-hydroxylase 1 potentially through interactions with the enzyme and/or cyclophilin B. Further testing showed that *Sc65*-KO mice also have fragile skin with less tensile strength than control mice, consistent with a collagen cross-linking abnormality. Collectively, these results indicate that Sc65 is a novel adapter molecule that stabilizes a unique ER-resident complex that is essential for proper collagen lysyl-hydroxylation. Loss of Sc65 leads to complex instability and defective fibrillar collagen modifications which negatively impacts bone, skin and likely other connective tissues. This phenotype and the underlying collagen molecular defect resemble those seen in EDS type VIA patients carrying *PLOD1* mutations; therefore mutations in either *Sc65* or *LEPREL2* may be responsible for a similar or milder connective tissue disease that is yet uncharacterized.

2176W

Pigmented 3D skin models made from patient-specific iPS cell-derived melanocytes. *K.M. Eckl[1,2,3], D. Lima Cunha[2,3,4], T. Saric[5], H.C. Hennies[2,3,4,6].* 1) Dept. for Biology, Edge Hill University, Ormskirk, Lancashire, United Kingdom; 2) Center for Dermatogenetics, Div. of Human Genetics and Dept. for Dermatology, Medical Univ Innsbruck, Innsbruck, Austria; 3) Dept. of Biological Sciences, School of Applied Sciences, Univ. of Huddersfield, Huddersfield, UK; 4) Center for Dermatogenetics, Cologne Center for Genomics, Univ. of Cologne, Cologne, Germany; 5) Inst. for Neurophysiology, Univ. Hospital Cologne, Cologne, Germany; 6) CECAD, Univ. of Cologne, Cologne, Germany.

Skin model systems are reaching international acclaim to further patient-specific drug development and to minimize the use of animals in applied medical research. Lately we developed 3D skin models for various monogenic severe genodermatoses with and without changes in pigmentation. Up to date most 3D full-thickness skin models mimic the multi-layered epidermis and a simplified dermis, but are void of melanocytes, the body's first-line defense against UV-radiation induced skin cancer. Patient-derived, pigmented 3D skin models help us to understand disease mechanisms and allow safe and patient-directed drug development for genetic orphan diseases. Induced pluripotent stem cells (iPSC) generated from dermal fibroblast are an excellent source for fate-specific differentiation of cells into any desired tissue. Here we present first results from a new approach to differentiate human feeder-cultured iPSC into melanin expressing pure cultures of melanocytes. From roughly five iPSC colonies we have been able to cryopreserve more than 40 million cells per line within the first six passages. Melanocytes derived from iPSC underwent extensive characterization at the final end point and during all differentiation processes to compare and contrast similarities and differences between healthy and diseased cell lines, which encompass cells from a late infant patient with severe photosensitive trichothiodystrophy (TTD) with a homozygous mutation in *ERCC2* as well as cells from a patient diagnosed with congenital ichthyosis caused by a homozygous nonsense mutation in *TGM1*. To proof in-vitro functionality of the cells in regard to melanin-transfer and viability in a tissue environment, we used iPSC-derived melanocytes for patient-specific 3D skin modelling. With success we showed iPSC-derived melanocyte integration into the basal layers of the epidermal sheets as well as synthesis of melanin and darkening of 3D skin models after exposure to UV light. Most monogenic genodermatoses either with or without pigmentation changes are rare. Treatment is often only symptomatic and rarely sufficient. Individualized pigmented 3D full-thickness skin models allow academia and industry to identify those patient and disease specific drugs, which further melanocyte integrity allowing the skin to defend UV light caused skin damage.

2177T

Functional modeling of Bainbridge-Ropers syndrome (BRS) mutations suggests a dominant-negative mechanism of pathogenesis. *H. Nistala[1], J. Dronzek[1], J. McClain[2], D. Kajimura[3], S. Wolf[1], J. Overton[1], C. Shoenherr[2], B. Zambrowicz[3], J. Lupski[4], A.N. Economides[1,2].* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Genome Engineering Technologies, Regeneron Pharmaceuticals, Tarrytown, NY; 3) Velocigene Operations, Regeneron Pharmaceuticals, Tarrytown, NY; 4) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Objective Heterozygous *de novo* truncating mutations in *ASXL3* are associated with BRS, characterized by neurologic abnormalities including intellectual disability, severe psychomotor retardation, and microcephaly. However, phenotype in affected individuals varies both in presentation and severity. Truncating mutations occurring in the 5' end of the penultimate exon are associated with more severe phenotype whereas truncating mutations upstream or at the extreme 3' end the gene result in benign phenotype. The goal of this study is to define molecularly the observed phenotypic polarity in BRS mutations. **Methods** Biallelic gene disruption at the *Asxl3* locus was achieved using CRISPR/Cas9-mediated gene editing in mES cells. Recombinants harboring truncating mutations either within the mutational hotspot (representing severe variants) or upstream (benign) were screened using Sanger sequencing and TOPO-PCR cloning. Targeted mES and parental mES (no editing detected) were differentiated into neuronal progenitors using established neural induction protocols. Transcript diversity and mRNA stability were compared between benign and severe variants using RT-PCR, PacBio sequencing and Northern-blot analyses. Furthermore, a recombinant chimeric splicing construct harboring *mAsxl3* locus (with or without the BRS mutations) flanked by constitutive splicing rat *Ins2* exon 2 and 3 was used to validate findings observed in the CRISPR-edited mES clones. **Results** Findings to date in CRISPR-edited mES and heterologous splicing reporter assays suggest that benign and pathogenic variants undergo nonsense-associated altered splicing (NAS). Consequently, mutant transcripts derived from the benign variant activated NMD whereas the stable truncated mRNA derived from pathogenic variants escape NMD with potential dominant-negative function. **Conclusion** Functional modeling of BRS mutations suggests that NAS and resulting activation or escape from NMD contributes to variable neurological phenotypes observed in patients with allelic truncating mutations in *ASXL3*. These observations are consistent with published clinical reports that truncating mutations occurring in the 5' end of Ex 11 or the novel splicing mutation abolishing splice donor site of Ex 11 result in severe BRS phenotype. Taken together, these findings crucially highlight how protein-truncating variants within the same gene can result in variable phenotypes depending on how the substrate transcript is post-transcriptionally processed.

2178F

Development of a screening platform to identify small molecules that modify mRNA splicing in familial dysautonomia. *M. Salani[1], A. Brenner[1], F. Urbina[2], E. Morini[1], E. Law[1], M. Nibratt[1], S.A. Slaugenhaupt[1].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) University of North Carolina, Chapel Hill, NC.

Familial Dysautonomia (FD) is an autonomic and sensory neuropathy caused by a mutation in the splice donor site of intron 20 of the *IKBKAP* gene. Variable skipping of exon 20 in the *IKBKAP* mRNA leads to a tissue-specific reduction in the level of IKAP protein. Patients produce normal mRNA and protein albeit at low leverls.Therefore, we hypothesized that we could increase cellular IKAP through splicing modification. We previously identified the plant cytokinin kinetin as an efficacious *IKBKAP* splicing modulator, able to restore normal levels of IKAP protein expression in patient cells as well as modify splicing *in vivo* in our transgenic mouse model of FD. Armed with a promising lead compound and unmet medical need, our FD project was chosen to be part of the NIH Blueprint Neurotherapeutics Network. Acting as a collaborative framework, Blueprint gives access to NIH-funded resources and consultants with expertise in various aspects of drug discovery. We developed a primary splicing assay by inserting Renilla (RLuc) and Firefly (FLuc) reporters into our previously well-characterized *IKBKAP* mini-gene constructs. Evaluation of FLuc/RLuc expression in 293T cells transfected with the construct enables a determination of the percent exon inclusion. Here we demonstrate the quality and reproducibility of our screening method. The 8-point dose-response curves generated in the primary assay have reproducible EC_{50} and efficacy values. To avoid false positives we also developed a counter screen in order to identify molecules that might interfere directly with the reporter. Further, we developed a secondary assay using human FD fibroblasts in which we quantify both IKBKAP expression and IKAP protein. After validation of our screening platform, we initiated medicinal chemistry and have tested 520 compounds using our primary assay, yielding 210 compounds that are more potent or efficacious than kinetin. Activity of a subset of these was verified in the secondary assay, and active compounds with favorable ADME and pharmacokinetic properties were evaluated in our transgenic mouse model of FD. *IKBKAP* splicing changes were observed *in vivo*, and detailed evaluation of tissue-specific activity is underway. Development and implementation of this moderately high throughput screening platform for splicing modification has greatly enhanced the likelihood of translating our findings into a new therapy that directly targets the molecular mechanism of disease in FD patients.

2179W

Dissection of the role of PARK9 (ATP13A2) in a clinical case with Kufor-Rakeb syndrome and implications for phenotype modification. *M. Kousi[1], J. Willer[1], V. La Bella[2], R. Spataro[2], N. Katsanis[1].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) ALS Clinical Research Center, University of Palermo, Palermo, Italy.

ATP13A2 is a member of the P5 ATPase family, recessive mutations in which have been reported to cause a recessive form of parkinsonism. The condition is termed Kufor-Rakeb syndrome and manifests juvenile-onset parkinsonism with associated spasticity, supranuclear gaze palsy, and dementia. Extensive inter- and intra-familial clinical variability have been reported. We recruited a family in which the 30-year-old proband presented with juvenile-onset amyotrophic lateral sclerosis (ALS) and dementia. The diagnosis was supported by the clinical observation of leg spasticity and extremity weakness with characteristic electromyography findings, as well as frontotemporal and cerebellar atrophy observed upon brain imaging evaluation. Subsequent to whole-exome sequencing, we identified a homozygous nonsense change that segregated with the phenotype and was absent from all control databases. *In vitro* studies in patient cells indicated that the mutant message is not subject to nonsense-mediated decay, thus indicating that this allele removes the last 567 amino acids of ATP13A2. To understand how the expression of mutant *ATP13A2* gives rise to an ALS pathology, we sought to establish the direction of effect of the allele identified in our proband and compare it to the effect of other previously published causative alleles, by generating an *in vivo* zebrafish surrogate model that recapitulates both the motor neuron deficits and cerebellar atrophy described. Armed with prior knowledge that supports that *ATP13A2* dysfunction is due to proteasomal defects, we complemented our studies by evaluating a range of components, including proteasomal agonists and genetic determinants such as *TFEB* that augment mitochondrial function and mediate misfolded a-synuclein aggregate exocytocis, as putative therapeutic approaches. Taken together, we expand the mutational spectrum of *ATP13A2* and highlight the importance of the *in vitro* and *in vivo* tools we generated towards the genetic and/or chemical amelioration of a core biological process that can be of putative therapeutic benefit to several other movement disorders that are due to proteasomal dysfunction, such as ALS and Parkinson disorder.

2180T

Expansion of the clinical phenotype of the distal 10q26.3 deletion syndrome to include ataxia and hyperemia of the hands and feet. *M. Lacaria[1], M. Srour[2,3], J. Michaud[4,5,6], A. Doja[7,8,9], E. Miller[10], J. Schwartzentruber[11], J. Majewski[11], K. Boycott[1,7,9] FORGE Canada Consortium.* 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada; 3) Department of Pediatrics, McGill University, Montreal, Quebec, Canada; 4) Research Center, Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada; 5) Department of Pediatrics, Universite de Montreal, Montreal, Quebec, Canada; 6) Department of Neurosciences, Universite de Montreal, Montreal, Quebec, Canada; 7) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 8) Division of Neurology, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 9) Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; 10) Department of Medical Imaging, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 11) Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

Distal deletion of the long arm of chromosome 10 is associated with a dysmorphic craniofacial appearance, microcephaly, behavioral issues, developmental delay, intellectual disability, and ocular, urogenital, and limb abnormalities. Herein, we present clinical, molecular and cytogenetic investigations of four patients, including two siblings, with nearly identical terminal deletions of 10q26.3, all of whom have an atypical presentation of this syndrome. Their prominent features include ataxia, mild-to-moderate intellectual disability, and hyperemia of the hands and feet, and they do not display many of the other features commonly associated with deletions of this region. These results point to a novel gene locus associated with ataxia and highlight the variability of the clinical presentation of patients with deletions of this region.

2181F

Impact of *SNX14* mutations on endocytic trafficking and autophagy. *P. Stanier[1], D. Bryant[1], E. Peskett[1], C. Demetriou[1], M. Ishida[1], M. Seda[1], J. Hurst[2], R. Scott[2], S. Sousa[3], D. Jenkins[1], M. Bitner-Glindzicz[1,2], G.E. Moore[1].* 1) University College London, Institute of Child Health, London, UK, United Kingdom; 2) Great Ormond Street Hospital, London, United Kingdom; 3) Hospital Pediátrico de Coimbra, Portugal.

Many conditions have been described where severe intellectual disability and ataxia are found in patients with cerebellar hypotrophy. Individually, most types are rare and without a known molecular pathology. Mutations in *SNX14* have recently been reported to cause a distinct autosomal-recessive cerebellar ataxia with moderate to severe intellectual disability, early-onset cerebellar atrophy, sensorineural hearing loss and coarsened facial features. The SNX14 protein contains a Phox (PX) domain and a regulator of G protein signalling (RGS) domain. Patient-derived fibroblasts were acquired from individuals of three unrelated consanguineous families in Turkey and Portugal. These cell lines have mutations that result in either truncation or loss of the SNX14 protein and in one patient, loss of a single exon (containing the PX-domain) alone was sufficient to result in the associated pathology. This suggests that the PX domain is critical for normal SNX14 function. *SNX14* mutations lead to the accumulation of vesicular inclusions in patient cells, indicating that there is a disturbance in protein metabolism and/or vesicle mediated transport. We show that cellular processes associated with autophagy provide additional evidence of this. However, the precise function of SNX14 remains unknown. We are investigating the role of SNX14 in subcellular trafficking by monitoring these processes in cell-based models of the disease. Loss of SNX14 appears to impact on cholesterol distribution and other lipids. Additionally, we are examining *snx14-/-* zebrafish to establish a research platform with which to explore the neurological deficits associated with the disease.

2182W

Relationship between *MAGEL2*, **a gene implicated in Prader-Willi syndrome, and RING-zinc finger-type E3 ubiquitin ligases.** *V. Carias, R. Wevrick. Medical Genetics, University of Alberta, Edmonton, Alberta, Canada.*

Background: Prader-Willi syndrome (PWS) is a neurodevelopmental disorder characterized by developmental disabilities, obesity, excessive daytime sleepiness and night-time wakening, distinct facial features and autism spectrum disorder. PWS is a sporadic multigene chromosome deletion disorder involving the 15q11-q13 region. *MAGEL2* is one of several genes inactivated in children with PWS and is expressed throughout the brain during development. Mutations in *MAGEL2* alone cause Schaaf-Yang syndrome, a neurodevelopmental disorder associated with many clinical manifestations that are also common in PWS. Members of this family of proteins interact with and modify the function of RING-zinc finger-type E3 ubiquitin ligases (E3 ligases). E3 ligases are involved in ubiquitination, a posttranslational modification in which proteins are targeted for modification and degradation. Ubiquitination is involved in various cell processes such as cell cycle, regulation of transcription, circadian rhythm, and neuronal development. The specific protein function and proteome of MAGEL2 has yet to be elucidated. We investigated the relationship between MAGEL2 and various E3 ligases implicated in PWS. Specifically, we focussed on the E3 ligase TRIM32, which is implicated in Bardet-Biedl syndrome, another genetic neurodevelopmental obesity syndrome, and is also mutated in one form of limb girdle muscular dystrophy. Methods and Results: Using the Gateway Cloning vector system, various constructs were created for expression in mammalian cells (HA-MAGEL2, SF-MAGEL2, V5-TRIM32, V5-TRIM55, V5-RNF41, and V5-RBX1). Co-transfections of each E3 ligase and *MAGEL2* at various ratios were performed in U2OS cells to determine if there are effects on the abundance of either protein via western blot analysis. Co-expression of MAGEL2 modified the stability of several E3 ligases as determined by western blot analysis. Immunofluorescence was used to determine subcellular localization of MAGEL2 and various E3 ligases. Interestingly, MAGEL2 co-localizes with several E3 ligases in the perinuclear region in U2OS cells. Interactions between MAGEL2 and E3 ligases have also been detected. Conclusions: The identification of interactions between MAGEL2 and E3 ligases and their effects on their ubiquitination protein substrates will be useful in determining the underlying cellular mechanisms disrupted in children with PWS and related disorders.

2183T

Variants in *FAT2* **and** *FAT3* **are associated with autism and childhood apraxia of speech.** *S. Klein[1], S. Mullegama[2], U. C.L.A Clinical Genomics Center[2], J.A. Martinez-Agosto[1,3].* 1) Department of Human Genetics, UCLA, Los Angeles, CA; 2) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 3) Department of Pediatrics, Division of Medical Genetics, UCLA, Los Angeles, CA.

The genetic heterogeneity of autism spectrum disorders (ASDs) poses a challenge in identifying distinct developmental and cellular processes responsible for its etiology. Stratification of phenotypes within ASDs may help to identify distinct genetic subgroups, uncovering patterns of connectivity and signaling pathways responsible for its overall etiology. Amongst ASDs, the presence of specific comorbidities, such as childhood apraxia of speech (CAS), helps to define endophenotypes that may share common genetic etiologies. In this study we surveyed a cohort of ASD cases that had undergone clinical exome sequencing (CES) and focused on a subset that presented with CAS (5%). In 20% of cases we have identified variants in *CNTNAP2,* which is known to regulate speech development. Here we report novel variants in *FAT2* and *FAT3* identified in an additional 40% of ASD-CAS cases. *FAT* genes encode for transmembrane proteins that contain tandem cadherin-type repeats as well as epidermal growth factor (EGF)-like repeats. CES identified *de novo* rare heterozygous (c.85T>C) and compound heterozygous variants (c.[3475 C>A]+[5726C>T]) in *FAT2* and *FAT3* (c.[1868A>C] + [12863 A>C]+[12413 A>C]; c.[11332C>T]+[1081 G>C]). Variants in *FAT2* and *FAT3* were also associated with esotropia and seizures. The presence of distinct phenotypic findings further suggests the important role of FAT genes in brain function where they are highly expressed. Specifically, while *FAT3* is expressed in the cerebral cortex, *FAT2* is highly expressed in the cerebellum, reinforcing that connectivity between cortical and cerebellar function may be important in the pathogenesis of ASD and CAS. Pathway analysis of the FAT2 and FAT3 proteins identified genetic links to key regulators of speech development including FOXP1, FOXP2, FOXP4 and CNTNAP2, strengthening their potential role in language development. Our findings demonstrate a novel association for FAT2-3 with ASD-CAS. We propose that ASD-CAS represents a distinct genetic subgroup of ASD cases and provide novel insights into the genetic etiology of ASD-CAS.

2184F

X-linked histone demethylase Kdm6a regulates cerebellar development and motor coordination. *J. Xu[1], H. Weimar[1], T. Driessen[1], K. Ge[2].* 1) Dept. of Integrative Physiology and Neuroscience, Washington State University, Pullman , WA; 2) National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD 20814, USA.

Kdm6a is an X-linked histone demethylase that activates gene expression via removal of the repressive methylation mark at histone H3 lysine 27 (H3K27). In humans, *KDM6A* mutations cause Kabuki syndrome, a disorder characterized by intellectual disability and motor coordination deficits. To assess the role of Kdm6a in brain development and behavior, we generated a neuron-specific Kdm6a deficient mouse model (Kdm6a[Def]). The Kdm6a[Def] and wild type (WT) littermates have comparable body weight, reproductive behavior, and lifespan, different from the constitutive *Kdm6a* KO mice which die prenatally. The Kdm6a[Def] and WT mice also scored similarly in behavioral tests such as fear conditioning, locomotor activity, and grip strength. However, the mutant mice were impaired for motor coordination when tested with a rotarod. Gene expression analysis with RNA-seq and RT-qPCR identified mis-regulated genes, including up-regulated expression of cerebellar granule cell-specific genes (e.g. *Zic1*, *Etv1*, *Gabra6*, and *Grin2c*), in the Kdm6a[Def] mice. Some of these genes, *Zic1* and *Etv1* for instance, are known to be crucial to the proliferation and differentiation of granule cell precursors. We thus hypothesized that, in the Kdm6a[Def] granule precursors, gene mis-regulation leads to an altered cellular development and subsequently circuitry abnormality in terms of the inhibition-excitation balance, resulting in the behavioral manifestation of motor incoordination. Supportive evidence has been obtained from our ongoing studies including BrdU tracing of the newly born cerebellar granule cells. We believe that this series of analysis is shedding new light on the etiology of Kabuki syndrome as well as the epigenetic basis of brain development and motor behavior.

2185W

Novel role for the small GTPase Bardet-Biedl Syndrome protein 3 in the regulation of glucose transport in the brain. *S.C. Huang[1], C.S. Carter[2], N. Nuangchamnong[3], V. Buffard[2], C. Searby[2], Q. Zhang[2], V.C. Sheffield[2].* 1) Department of Molecular and Cellular Biology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) Department of Obstetrics and Gynecology, University of Iowa, Iowa City, IA.

Glucose is the obligate fuel source for the brain and is essential for normal neurological function. This nutrient depends on glucose transport protein type 1 (GLUT1) for its passage through the endothelial cells of the blood brain barrier. Dysregulation of such transport underlies GLUT1 Deficiency Syndrome [MIM: 606777], a debilitating neurological disease characterized by seizures and cognitive impairment. In spite of the importance of glucose delivery to the brain, little is known about the mechanisms that regulate its passage across the blood brain barrier. Therefore, the purpose of this study is to better understand glucose transport across the blood brain barrier and how glucose regulation affects brain function. Our approach to studying this problem was by using a genetic disease model. The blood brain barrier endothelial cells contain primary cilia to which transporter proteins and their downstream effectors often localize. Emerging evidence suggests that cilia play a central role in coordinating signaling pathways in development and cell homeostasis, with defects causing diseases called ciliopathies. One of the most studied ciliopathies is Bardet-Biedl Syndrome (BBS), a genetically heterogeneous and pleiotropic disorder with cardinal features that include cognitive impairment, obesity, retinal degeneration, and polydactyly. Proteins encoded by these BBS genes form a protein complex called the BBSome, which traffics cargo to and from cilia. Among these BBS-associated proteins is BBS3 [MIM: 600151], an ADP-ribosylation factor-like small GTPase. We have recently found that *Bbs3* knockout mice show diminished glucose levels in the brain with hallmarks of GLUT1 Deficiency Syndrome. We have identified low glucose levels in the cerebral spinal fluid in *Bbs3* knockout mice despite normal blood glucose. Through assessment of glucose uptake in the brain by fludeoxyglucose positron emission tomography, we have found that *Bbs3* knockout mice also have low glucose uptake in the brain, and this decreased uptake varies across different brain regions. Our findings indicate that BBS3 plays a role in regulating glucose transport across the blood brain barrier.

2186T

Impaired ER-Golgi trafficking as a novel therapeutic target for Pelizaeus-Merzbacher disease caused by PLP1 amino acid substitutions. *K. Inoue[1], H. Li[1], P.R. Mangalika[1], A. Nishizawa[1], H. Saya[2], Y. Goto[1].* 1) Dept MR & BD Res, Natl Inst Neurosci, NCNP, Kodaira, Tokyo, Japan; 2) Div Gene Regulation, Inst Advanced Medical Research, Keio Univ, Tokyo, Japan.

Pelizaeus-Merzbacher disease (PMD) is the most common, but incurable inherited hypomyelinating leukodystrophy caused by various mutations in the *PLP1* gene, which encodes a major myelin membrane lipoprotein. PLP1 amino acid substitutions often cause severe form of PMD. In addition to the well-recognized activation of the apoptotic pathway of unfolded protein response (UPR) as an underlying mechanism for oligodendrocyte cell death in PMD brains, we recently identified that mutant PLP1, which accumulates in ER, also interferes with intracellular transport of other normal membrane and secretary proteins, possibly affecting the cellular homeostasis. To explore this cellular pathology further at the molecular level, we employed temperature-sensitive VSVG-YFP system for real-time visualization of the secretary trafficking. We found that, while WT PLP1 has no effect, mutant PLP1 drastically delayed the processing time of VSVG-YFP coming out of ER. We have also explored if drugs that can enhance the transport of secretary and membrane proteins may rescue the cellular phenotype caused by mutant PLP1. To identify such drugs, we screened a library of 1400 existing medicine and identified 8 potential compounds by luciferase-based screening and immunostaining of KDEL receptor, which is normally present in Golgi apparatus but is mislocalized in ER when co-expressed with mutant PLP1. Current studies by treating PLP1 mutant mice (Msd) with these compounds may lead to a discovery of novel medicines for the treatment of PMD. We propose that impaired ER-Golgi trafficking by mutant PLP1 may serve as a novel therapeutic target for PMD caused by PLP1 amino acid substitutions.

2187F

Role of autophagy in the pathology of Snyder-Robinson Syndrome. *M.C. Malicdan[1,2], M. Goheen[1], J. Albert[1], C. Li[3], C.E. Schwartz[4], Y. Huang[1], J. Cassady[1], L. Wolfe[1,5], G. Zhai[3], P. Lee[1,5], D.R. Adams[1,2,5], W.A. Gahl[1,2,5].* 1) NIH Undiagnosed Diseases Program, NIH, National Human Genome Research Institute, NIH, Bethesda, MD; 2) MGB, NHGRI, National Institutes of Health, Bethesda, MD; 3) University of Miami, Medical School, Department of Molecular and Cellular Pharmacology, Miami, FL; 4) JC Self Research Institute, Greenwood Genetic Center, Greenwood, SC; 5) Office of the Clinical Director, National Human Genome Research Institute, NIH, Bethesda, MD.

Snyder-Robinson Syndrome (SRS) is an X-linked intellectual disability syndrome and results from loss of function mutations in the spermine synthase (*SMS*) gene. SMS is an enzyme involved in the polyamine biosynthesis pathway that catalyzes the conversion of spermidine to spermine. Two brothers presented to the NIH Undiagnosed Diseases Program (UDP) exhibiting SRS clinical features: facial asymmetry, hypotonia, intellectual disability, marfanoid habitus, muscular atrophy, and osteoporosis. Whole exome sequencing and follow-up Sanger sequencing revealed that both patients inherited an NM_004595.4:c.443A>G transition on chromosome X from their heterozygous mother, causing a missense Gln148Arg mutation in the SMS protein. Further analysis of patient cell lines showed a decrease in SMS protein levels, and a corresponding increase of spermidine and decrease of spermine, compared to control cells. Several studies have suggested that changes in polyamine levels can modulate autophagy and alterations of autophagy have been implicated in a number of neurodegenerative disorders as a major contributor to neurodegeneration. We hypothesized that modulation of autophagy in the two UDP SRS patients may be a contributor to their neurological phenotype. To investigate this connection, we designed a series of experiments to assess the macroautophagocytic process in the SMS mutant patient cell lines. Western blot analysis suggested increased steady-state levels of the autophagosome protein markers, LC3B-II and p62, in fibroblasts. Similarly, using immunofluorescence, increased autophagosome formation was detected in patients' cells. Our results indicate that macroautophagy is upregulated in SMS patient cells, suggesting that spermidine accumulating in cells due to spermine synthase deficiency may be targeted for autophagic degradation.

2188W

Phenytoin-responsive epileptic encephalopathy with a intragenic tandem duplication involving *FGF12.* A. Kikuchi[1], R. Shi[2], T. Kobayashi[1,3], R. Sato[1], M. Uematsu[1], K. An[4], S. Kure[1]. 1) Department of Pediatrics, Tohoku University Hospital, Sendai, Japan; 2) Department of Neonatology, First Affiliated Hospital, Medical College of Xi'an Jiaotong University, Xi'an, China; 3) Division of Genomic Medicine Support and Genetic Counseling, Department of Education and Training, Tohoku Medical Megabank Organization (ToMMo), Tohoku University, Sendai, Japan; 4) Clinical Technology Department, Tohoku University Hospital, Sendai, Japan.

Epileptic encephalopathies are heterogenous entity of epilepsies characterized by reccurent seizures with developmental slowing or regression. The genetic etiology of epileptic encephalopathies has gradually been revealed by the identification of causative genes, which include mutations in sodium channel subunits. Recently, non-secreting fibroblast growth factors (FGFs), sodium channel-binding proteins, have been associated with neurodevelopmental disorders including epileptic encephalopathy. Moreover, Siekierska et al. have reported a family of patients with epileptic encephalopathies with a *de novo* mutation in *FGF12*, which have a gain-of-function effect on sodium channel gating, predicting increased neuronal excitability (Neurology, 2016). Here, we report another case of *FGF12*-related epileptic encephalopathy. Contrast with the previous report, the phenotype was relatively mild. The patient had normal development prior to the onset of seizures at three years of age, and his seizure was respond to phenytoin, a sodium channel blocker. Genomic analysis revealed a intragenic tandem duplication involving the gene, probably thorough L1 family long interspersed nuclear elements-mediated nonallelic homologous recombination. cDNA analysis showed biallelic expression of FGF12 with abnormal transcripts. Although it remains unclear that how the products from the mutant allele work, the absence of loss-of-function mutations except in last exon or isoform-specific exon in ExAC database implicates that haploinsufficiency of FGF12 may be enough to invoke a disease phenotype as well as gain-of-function effect as described by Siekierska et al.. This report provides further evidence for as a causative gene for epileptic encephalopathy and expands the phenotypic spectrum of the *FGF12*-related epileptic encephalopathy. We also propose a therapeutic approach for *FGF12*-related epileptic encephalopathy by a sodium channel blocker.

2189T

Identification of muscle segment homeobox gene 1, *MSX1,* **as a candidate susceptibility gene for congenital heart disease in individuals with Wolf-Hirschhorn syndrome.** E.R. Harward[1], S. Venkatasubramanian[3], R.J. Vanzo[1], C.H. Hensel[1], M.R. Sdano[1], M.M. Martin[1], A. Lortz[4], J.C. Carey[2], E.R. Wassman[1], K.S. Ho[1,2]. 1) Lineagen, Inc, Salt Lake City, UT; 2) Department of Pediatrics, University of Utah; 3) School of Computing, University of Utah; 4) 4p- Support Group.

Wolf-Hirschhorn syndrome (WHS) is a well-described contiguous gene deletion syndrome caused by varying sized deletions of the short arm of chromosome 4 with highly variable phenotypic features and disability. We used high-resolution genotype-phenotype correlation to define genetic loci within the 4p region that are likely causative for individual features and recently described a novel candidate gene associated with seizures in these individuals (Ho, et al, 2016). In the same cohort of thirty-four individuals with WHS, with deletions of 4p ranging from 1.3–32.5 MBp in size and containing 28-200 genes, we further correlated deletion breakpoints on a custom, ultra-high resolution chromosomal microarray with over 20 other specific phenotypic features of WHS. A novel statistical technique was used to analyze this dataset for non-obvious correlations between specific clinical features and genomic regions to identify candidate genes of likely pathogenicity. Congenital heart defects are highly prevalent in our WHS cohort (71%) and our statistical approach suggested muscle segment homeobox 1 (*MSX1*) as a relevant candidate gene within a 3 Mbp region associated with congenital heart defects (p=0.0048 Fisher's exact test, two-tailed). Within this region, *MSX1* is the likely candidate gene supported by its role in cardiac development. *MSX1* acts along with Wnt and BMP signaling in early vertebrate cardiac development (Rao et al, 2016), and dual disruption of both *Msx1* and its paralog *Msx2* in mice results in cardiac outflow tract abnormalities. *MSX1* appeared hyper-methylated in a human fetus with double outlet right ventricle, VSD, and hypoplasia of the ascending aorta (Serra-Juhe et al 2014). Aortic valve dysplasia was reported in a woman with a *de novo* duplication of a 3.8 Mbp containing the *MSX1* gene (Hitz, 2012). Two independent genome-wide association studies and one replication study identified SNPs in *MSX1* associated with atrial and ventricular septal defects. (Li et al 2015; Cordell et al 2013). *MSX1* had been proposed as a candidate gene involved in the oligodontia and cleft lip/palate associated with WHS. Our analysis did not support *MSX1* as the best candidate gene association for either, rather finding FGF pathway members involved in both. This computational approach will be applied to other traits associated with WHS as well as other genetic syndromes to search for the best predictive medical management.

2190F

Using zebrafish to model the role of *CLEC16A* in neurological disease. *M. Bakay[1], C. Seiler[2], S. Yoeu[1], C. Kao[1], R. Pandey[1], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Research Core Services, Children's Hospital of Philadelphia, Philadelphia, PA.

CLEC16A SNPs are associated with multiple autoimmune and inflammatory diseases, including type 1 diabetes, multiple sclerosis primary adrenal insufficiency, Crohn's disease, primary biliary cirrhosis, juvenile idiopathic arthritis, rheumatoid arthritis and alopecia areata; however, the primary function of this gene and the protein its encodes remains largely unknown. Recent studies in different model systems revealed that CLEC16A play an important role in autophagy/mitophagy. Zebrafish are now established as a valuable tool for disease modeling, especially for mitophagy and neurodegeneration. We hypothesize that using a CLEC16A specific morpholino (MO) mRNA knock down in zebrafish will reveal a neurological phenotype and help to understand the role of CLEC16A in nervous system.We designed a splice blocking MO which generates a frameshift mutation after exon 4 and introduced a STOP-codon soon after. RT-PCR was used to identify and quantify new transcripts. Introduction of the mutation was confirmed by sequencing. To perform MOs titration, three different concentrations (0.5, 0.75 and 1.0mM) of CLEC16A specific MOs were injected into the yolks of 1–8-cell-staged embryos. Since all of them had similar efficiency we decided to use 0.75mM MO for all further experiments. We confirmed stable CLEC16A KD as long as day 7 by CR-PCR. We studies 3, 5 and 7 day old zebrafish by microscopy. Loss of CLEC16A expression resulted in significantly reduced branching of motor neurons when compared to control MO injections.Our data suggest that CLEC16A may play an important role in motor neuron branching and development. We are currently investigating the effect of CLEC16A loss on disease pathogenesis in CLEC16A-mutant zebrafish focusing on neurogenesis, immune cell function and autophagy with preliminary results from those studies supporting an important role of CLEC16A in neuronal development.

2191W

Differential gene expression of sensory and motor neuron in embryonic spinal cord with single cell analysis. *J. Kim, M. Lin, T. Souaiaia, V. Wolseley, C. Walker, A. Camarena, J. Nguyen, J. Herstein, W. Mack, O. Evgrafov, R. Chow, J. Knowles.* Zilka Neurogenetic Institute, USC, 1501 San Pablo st. #401 Los Angeles, CA.

The gray matter in spinal cord is composed of dorsal and ventral horns. Dorsal horn contains sensory neurons and ventral horn contains motor neurons. Gradient of SHH protein expression determines the fate of these compartment during brain development. However, details of differentiation of these compartments of human embryo remains unclear. Our group has performed single-cell RNA sequencing of 17 interneurons, 34 sensory neurons and 226 motor neurons from human embryonic spinal cord of gestational age 9-16 weeks. Distinct sensory and motor neuron clusters were observed in PCA plot. 1,674 genes were found to be differentially expressed using DESeq2. This list includes several G protein coupled receptor genes, which are involved in pathways downstream of SHH (Sonic Hedgehog) gradient. We found some sensory markers such as POU4F1 on the list of overexpressed genes in sensory neurons. .

2192T

LGI1 mutations linked to lateral temporal epilepsy allow protein secretion and impair interaction with ADAM22/23 receptors. *C. Nobile[1], E. Dazzo[1], E. Leonardi[2], E. Belluzzi[3], E. Greggio[3], L. Vitiello[3], S. Malacrida[4], S. Tosatto[4].* 1) CNR-Neuroscience Institute, Padova, PD, Italy; 2) Woman and Child's Health Dept, University of Padua, Padova, Italy; 3) Biology Dept, University of Padua, Padova, Italy; 4) Biomedical Sciences Dept, University of Padua, Padova, Italy.

Autosomal dominant lateral temporal epilepsy (ADTLE) is a focal epilepsy syndrome caused by mutations in the LGI1 gene, which encodes a secreted protein. Most ADLTE-causing mutations inhibit LGI1 protein secretion, and only a few secretion-positive missense mutations have been reported. We describe the effects of four disease-causing nonsynonymous LGI1 mutations, T380A, R407C, S473L, and R474Q, on protein secretion and extracellular interactions. Expression of LGI1 mutant proteins in cultured cells shows that these mutations do not inhibit protein secretion. This likely results from the lack of effects of these mutations on LGI1 protein folding, as suggested by 3D protein modelling. In addition, immunofluorescence and co-immunoprecipitation experiments reveal that all four secretion-positive mutations significantly impair interaction of LGI1 with the ADAM22 and ADAM23 receptors on the cell surface. Thus, rather than inhibiting protein secretion, some ADLTE-causing LGI1 mutations reduce affinity of extracellular LGI1 for ADAM22 and ADAM23, which are both involved in the molecular mechanisms leading to ADLTE.

2193F

A combined approach of exome sequencing and genome editing identified *WDR62/MCPH2* mutations in patients with primary microcephaly. *T. Miyamoto[1], Y. Masatsuna[1], A. Fukumitsu[1], S.N. Akutsu[1], K. Hosoba[1], H. Morino[2], H. Kawakami[2], T. Yamamoto[3], K. Shimizu[4], H. Ohashi[4], S. Matsuura[1].* 1) Genetics & Cell Biol, RIRBM, Hiroshima University, Hiroshima, Japan; 2) Epidemiol, RIRBM, Hiroshima University, Hiroshima, Japan; 3) Math & Life Sci, Grad. Sch. Sci, Hiroshima University, Higashi-hiroshima, Japan; 4) Saitama Children's Medical Center, Saitama, Japan.

Primary microcephaly (MCPH) is an autosomal recessive neurodevelopmental disorder characterized by congenital reduction of head circumference. So far 13 causative genes have been identified. It was reported that *WDR62* gene, which encodes centrosomal protein to control mitotic spindle assembly, is responsible for *MCPH2* gene locus (Nature 2010, Nat Genet 2010). Here we identified compound heterozygous mutations c.731 C>T (p.Ser 244 Leu) and c.2413G>T (p.Glu 805 X) in the *WDR62/MCPH2* gene of the two siblings in a Japanese microcephalic family using whole exome sequencing. The molecular and cellular pathology of *WDR62/MCPH2* mutation-caused microcephaly remains unclear. Here we used CRISPR/Cas9 system-mediated genome editing technology and single strand oligonucleotide (ssODN) as a point-mutation targeting donor to generate human *WDR62/MCPH2* c.731 C>T (p.Ser 244 Leu) missense mutation knock-in cell lines in order to elucidate the physiological role of WDR62. *WDR62/MCPH2* mutated cells did not show aneuploidy, suggesting that WDR62 is dispensable for chromosomal stability. In the normal metaphase, mitotic spindle forms parallel to the substratum to ensure symmetric cell division. *WDR62/MCPH2* mutated cells exhibited randomized spindle orientation caused by the impaired microtubule assembly activity. During early neurodevelopment, neural stem cell expands its population by symmetric cell division, and the following asymmetric cell division produces neural stem cells and neurons to control the proper size of brain. Our results suggest that *WDR62/MCPH2* mutations disrupt symmetric cell division of neural stem cell to induce premature asymmetric cell division through the instability of spindle orientation to cause microcephaly.

2194W

Sequencing of exon 5 Chat gene in Iranian congenital myasthenic syndrome patients. *Z. Farjami[1,2], F. Shariatmadari[3], M. Houshmand[1].* 1) National Institute of Genetic Engineering and Biotechnology, Tehran,Iran, Tehran, Iran; 2) Department of Biology, Damghan Branch, Islamic Azad University, Damghan Iran; 3) Amir Kabir Hospital, Department of Pediatric Neurology, Faculty of Medicine, University of Medical science, Arak, Iran.

Congenital myasthenic syndrome (CMS) is a heterogeneous group of inherited disorders, which caused by handicapped signal transmission at the neuromuscular junction. Fatigable muscle weakness is a diagnosis symptom of syndrome. Phenotype analysis showed a uniform pattern of clinical features including mild to moderate fatigable weakness of ocular and bilateral ptosis, facial, bulbar, and limb muscles Mutations in several genes that encoding proteins expressed at the neuromuscular junction may cause CMS. Aim of present study is assessment of mutations in 5 exon of CHAT gene, which encodes choline acetyltransferas. The important Mutations in this gene cause congenital myasthenic syndromes. Utility of deletion/duplication analysis for this gene is unknown .The choline acetyltransferas CHAT gene might be frequent in European congenital myasthenic syndrome patients of Gypsy ethnic origin. Proportion of CMS Attributed to Mutations in CHAT Gene is 5%. We selected 23 Iranian patients with congenital myasthenic syndrome. DNA extracted from peripheral blood and polymerase chain reaction (PCR) and DNA cycle sequencing operated by automated sequencers ABI 3700 for exon 5 of CHAT gene. Our results in 23 Iranian patients with CMS, showed no missense mutation in 5 exon of CHAT Gene. It seems that Iranian CMS population has different genetic variations in compare with European population. Also different exons in CHAT or other genes could be hot spots in Iranian CMS population, which needs further studies.

2195T

Effectively modeling rare neurodevelopmental disorders: A pilot project identifying a novel chromatin remodelling disease. *S.C. Bell[1], H. Peng[1], I. Kolobova[1], C. Ernst[1,2,3].* 1) Neuroscience, McGill University, Montreal, Quebec, Canada; 2) Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Psychiatry, McGill University, Montreal, Quebec, Canada.

Developing effective models of rare neurodevelopmental disorders (NDDs) presents significant challenges due to the rarity of patients and the difficulties associated with acquiring human neuronal cells. This commonly forces researchers to rely on substandard models to study rare NDDs, which can obfuscate results. Therefore, developing a more effective methodology for modeling rare NDDs represents a key step towards improved therapeutics for these disorders. **Here we present our methodology for modeling rare NDDs, using a family with a rare mutation in a chromatin remodelling complex as a pilot study.** We recruited a family with two children suffering from severe seizures and developmental delay, with an unknown causative mutation. Exome sequencing of twenty-three family members identified a novel and potentially disease-causing variant in the gene *ACTL6B*, which encodes a key targeting component of an important regulator of neuronal development, the *SWI/SWF* chromatin remodeling complex. In order to model the disease present in the family, we converted fibroblasts from affected and healthy patients into induced pluripotent stem cells (iPSCs) and generated cortical neural progenitor cells. Using this disease model, patient cells were found to have a significantly reduced level of expression of *ACTL6B*, which correlated with a dysregulation of neuronal differentiation. The methodology utilized to investigate this variant in *ACTL6B* demonstrates how researchers can develop efficient and effective models of neural cells to model rare NDDs. The research shown here is also the first description of a disease-causing variant in the *ACTL6B* gene.

2196F

Interactions between the Prader-Willi syndrome candidate protein MAGEL2 and neuronal proteins. *M. Sanderson, H. Cortes, M. Zoeteman, R. Wevrick.* Medical Genetics, University of Alberta, Edmonton, Alberta, Canada.

   Inactivating mutations in the *MAGEL2* gene results in a neurodevelopmental disorder that causes intellectual disability called Schaaf-Yang syndrome. *MAGEL2* is one of the genes inactivated in Prader-Willi syndrome, which shares many clinical findings with Schaaf-Yang syndrome such as autism spectrum disorder. *MAGEL2* is predominantly expressed in brain and muscle, but is also seen in other tissues. The MAGEL2 protein contains a MAGE homology domain that is shared with ~40 other mammalian MAGE proteins. The MAGE protein necdin is encoded by *NDN*, which is co-deleted with *MAGEL2* in people with PWS. Both MAGEL2 and necdin interact with E3 ubiquitin ligases through their MAGE homology domain. In mice, loss of *Magel2* affects brain development and function, disrupts circadian cycles and causes other phenotypes that recapitulate symptoms of Schaaf-Yang and Prader-Willi syndromes. Mice lacking *Magel2* have decreased volume in specific regions of the brain and altered brain chemistry. However, little is known about the role that MAGEL2 plays in brain function and development. Identifying the proteins that interact with MAGEL2 in neuronal cells will allow us to connect the underlying molecular processes that MAGEL2 is involved in with the neurodevelopmental deficiencies in Prader-Willi syndrome and associated disorders. We used a combination of approaches to identify novel interactions between MAGEL2 and other proteins, including affinity purification-mass spectrometry, proximity-dependent biotin identification, yeast two hybrid analysis and candidate protein screening. We expressed recombinant epitope-tagged human MAGEL2 or mouse Magel2 in human or mouse cell lines that endogenously express MAGEL2/Magel2. We identified MAGEL2-interacting proteins involved in a variety of intracellular processes, including axon guidance, axonal transport, cellular responses to growth factor stimulus, nervous system development, protein ubiquitination and deubiquitination, retrograde transport, and autophagy. Expression of MAGEL2 affects the abundance of co-expressed proteins, suggesting that it plays an important role in protein stability through ubiquitin-mediated processes. This study provides additional insight into the normal cellular and developmental role of *MAGEL2*, and mechanisms through which the loss of *MAGEL2* in children with Schaaf-Yang or Prader-Willi syndromes impairs brain development and function.

2197W

Disruptions to the miRNA regulatory pathway may cause an increased rate of schizophrenia in individuals with 22q11.2 DS. *W. Manley, S. Siecinski, S. Ryan, V. Coulibaly, S. Buyske, L. Brzustowicz.* Genetics, Rutgers University, Piscataway, NJ.

   Schizophrenia is a complex and poorly understood disease caused by the interplay of many environmental and genetic factors. The 22q11.2 deletion syndrome (DS) is a disorder that is caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia in affected individuals versus the general population. The missing 22q11.2 region contains DGCR8, which is required for the initial step of miRNA biogenesis. However, the 22q11.2 deletion itself is not sufficient to cause schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11.2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via depletion of several miRNAs. This network may serve as a protective buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have expanded our sample of human neural stem cells (NSCs) from individuals with the 22q11.2 deletion and schizophrenia. The NSCs were generated from iPSCs using Neural Induction Media (Life Technologies). In order to ensure the presence of the 22q11.2 deletion, a FISH analysis was conducted (Cell Line Genetics) using a probe for 22q11.2 (TUPLE). Also, the levels of DGCR8 gene expression were quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Gene Expression Assays and Copy Number Assays to ensure DGCR8 reduction in 22q11.2 DS NSCs relative to otherwise healthy control NSCs. Unexpected chromosomal abnormalities associated with iPSC culture methods have made routine detection of the deletion essential. We have characterized disruptions to the miRNA regulatory network in the NSC lines using Taqman Array Human Microarray Cards Version 3.0 and Illumina Small RNA sequencing (RNA-Seq) using ten lines (5 control and 5 22q11.2 DS). Here we will present the miRNAs that we have identified to be differentially expressed in otherwise healthy control NSCs versus NSCs with 22q11.2 DS and schizophrenia. Our preliminary data shows that of the 377 tested miRNAs, nearly 30% have at least a 2 fold expression change in deletion samples. We predict that these miRNAs could be involved in the elevated risk of developing schizophrenia in individuals with 22q11.2 deletion syndrome versus the general population. Future studies will focus on determining the biological targets of these miRNAs using miRNA target prediction software.

2198T

Mouse models of human disease available from The Jackson Laboratory. *S. Rockwood, M. Sasner, The Genetic Resource Science Team.* The Jackson Laboratory, Bar Harbor, ME.

As our understanding of basic disease mechanisms grow, so does our appreciation for the complexities that stand as barriers to developing effective therapies. The advent of new genetic engineering technologies, tractable in the mouse, has spawned promising preclinical approaches to overcome these obstacles. To facilitate these developments, The Jackson Laboratory (JAX) Mouse Repository provides access to one of the largest sets of well-characterized genetically engineered mutant mouse strains available. Hundreds of new mouse lines generated by both the international scientific community and JAX researchers are added to the Repository each year. A growing selection of models with applications as hosts for cancer xenograft modeling, primarily on the NSG platform, is readily available. Strains have been optimized for specific purposes: support of human and murine hematopoietic cell engraftment, reduced xenogeneic graft-versus-host disease response, engraftment of human hematopoietic stem cells without irradiation and stable engraftment of primary human hepatocytes. Mouse strains that recapitulate aspects of specific diseases, such as Alzheimer's disease and Huntington's are available as well. To exploit the potential of CRISPR/Cas9 technology to generate precise models of disease, a large variety of Cas9-expressing lines are offered, including constitutive expressing and *cre*-inducible lines. Cas9 strains are available on multiple genetic backgrounds. In response to the threat posed by the spread of Zika virus, colonies of *Ifnar1*-deficient mice have been expanded. Several recent studies demonstrate that *Ifnar1* knockout mice represent a model that facilitates the study of Zika virus pathogenesis and should prove to be useful in the development of related therapies. In order to promote reproducibility in research, the Repository maintains a robust quality control and assurance program. Newly arriving strains are assayed to confirm the presence of expected genetically engineered mutations. Equally important, strains are also routinely screened for undesirable contaminating alleles and genetic backgrounds. Donating a strain to the Repository is an easy way to fulfill the NIH's requirements for sharing mice. Researchers wishing to have strains considered for inclusion in the Repository are encouraged to submit their strains: www.jax.org/donate-a-mouse. This work is supported by NIH, The Howard Hughes Medical Institute, and several private charitable foundations.

2199F

IImpaired transferrin receptor recycling in neurodegeneration with brain iron accumulation (NBIA). *A. Drecourt[1], M. Dussiot[1], J. Babdor[1], M. Garfa-Traoré[1], N. Goudin[1], F. Habarou[1], C. Bole-Feysot[1], P. Nitschké[1], C. Ottolenghi[1], V. Serre[3], I. Desguerre[2], N. Boddaert[2], O. Hermine[1], A. Munnich[1,2], A. Rotig[1].* 1) Institut IMAGINE, INSERM U1163, 24 Boulevard du Montparnasse, 75015 Paris, France; 2) Departments of Pediatrics, Radiology and Genetics, Hôpital Necker-Enfants Malades, 149 Rue de Sèvres, 75015 Paris, France; 3) Institut Jacques Monod, CNRS (UMR7592), 15 rue Hélène Brion, 75013 Paris, France.

Neurodegeneration with brain iron accumulation (NBIA) encompasses a group of rare neurogenerative disorders with different clinical, brain MRI and molecular features, underlined by progressive extrapyramidal dysfunction and iron accumulation in the brain. Eleven disease genes have been hitherto identified (*PANK2* [MIM 606157], *PLA2G6* [MIM 603604], *COASY* [MIM 609855], *FA2H* [MIM 611026], *ATP13A2* [MIM 610513], *C2orf37* [MIM 612515], *WDR45* [MIM 300526], *C19orf12* [MIMM 614297], *CP* [MIM 117700], *FTL* [MIM 134790] and *GTPBP2* [MIM 607434]). Few of these genes are known to be related to iron homeostasis. Exome sequencing in two independent families identified two novel genes for NBIA, *CRAT* [MIM 600184] and *REPS1* [MIM 614825]. *CRAT* encodes the carnitine acetyltransferase and b-oxidation of U-$^{13}C_{16}$ palmitate in cultured fibroblasts of the patient detected decreased levels of C_4 and C_2 suggesting a reduced β-oxidation that was rescued by overexpression of wt *CRAT* cDNA. REPS1 is involved in endocytosis and fibroblasts of the patient showed significant increase of iron content that was rescued by overexpression of wt *REPS1* cDNA. Internalization of transferrin (Tf)-bound iron by transferrin receptor 1 (TfR1) mediated endocytosis is the major route of iron uptake. Iron homeostasis is regulated by a post-transcriptional mechanism involving iron responsive proteins (IRPs) that represent the main iron regulators in vertebrates. This post-transcriptional regulation is the only known mechanism of iron homeostasis regulation. Interestingly, post-transcriptional regulation of TfR1 as well as ferritin was normal in fibroblasts of patients with *REPS1*, *CRAT*, *PANK2*, *PLA2G6*, *C19ORF12* and *FA2H* mutations, yet TfR1 and ferritins proteins were dramatically increased accounting for the abnormal iron uptake and major iron overload in NBIA . We also observed abnormal Tf and TfR1 recycling in NBIA fibroblasts, increased amount of TfR1 at cell surface and iron overload when cells were grown in high iron condition. Therefore, whatever the disease causing gene, NBIA fibroblasts are characterized by abnormal TfR1 recycling and iron overload. Moreover, the apparent escape of TfR1 and ferritin to IRPs regulation suggested alteration of an as yet unknown post-translational regulation in NBIA fibroblasts.

2200W

Immune cells emerging as major source of neurodegeneration in Parkinson's disease. *M.K. Pandey[1,2], D. Magnusen[1], M.M. McKay[1], M.A. Gurtler[1], M. Caballero[1], T.A. Burrow[1,2].* 1) Divisions of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; 2) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229.

GBA1 mutations and increases of neuronal α-synulcein protein (α-syn) are the hallmark of Parkinson's disease (PD). Using *Gba1[D409/D409]* mouse model of PD, we have observed α-syn connection with brain recruitment of inflammatory subsets of dendritic cells (DCs) and T cells and their role in increased production of pro-inflammatory cytokines. These cytokines trigger neuronal loss and cognitive dysfunction in *Gba1[D409/D409]* mouse. However, the mechanism by which α-syn induces immune cells activation and increased generation of pro-inflammatory cytokines in PD is still not clear. To explore this, DCs and CD4[+]T cells were purified from brain of WT and *Gba1[D409/D409]* mouse and used for performing their ex-vivo stimulation with α-syn. These later cells were used for measurement of any alteration in phosphorylation of signaling cascades, (e.g., ERK1/2 and AKT), which are critical for generation of pro-inflammatory cytokines and induction of tissue inflammation. Our data showed that α-syn stimulated immune cells cause upregulation of pERK1/2 and pAKT and increased generation of pro - inflammatory cytokines (e.g., IFNg, TNFα, and IL6) in *Gba1[D409/D409]* mouse. Targeting indicated signaling could be potential therapy to discontinue inflammation and neurodegeneration in PD. .

2201T

The molecular phenotypes of *CHRNA7* copy number changes in human cells. *M.A. Gillentine[1], J.J. Kim[2], C.P. Schaaf[1,3].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Stem Cell Core, Baylor College of Medicine, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute, Houston, TX.

Chromosome 15q13 is one of the least stable regions in the genome due to low copy repeat (LCR) elements clustering into breakpoints (BP1-BP6), making the region vulnerable to non-allelic homologous recombination. At 15q13.3, recurrent copy number variants (CNVs) are observed in neuroaffected populations utilizing three sets of breakpoints, BP3-BP5, BP4-BP5, or the distal-*CHRNA7*-LCR (D-CHRNA7-LCR)-BP5. Deletions have been established as pathogenic, with probands presenting with autism spectrum disorder (ASD), epilepsy, and intellectual disability/developmental delay (ID/DD). 15q13.3 duplications, one of the most common CNVs identified in clinical microarray samples, have only been recently characterized, with probands presenting with ID/DD, ASD, and attention deficit hyperactivity disorder. Encoding for the α7 nicotinic acetylcholine receptor (nAChR) with high expression in the brain, *CHRNA7* has been suggested as a candidate gene for these CNVs. Additionally, the variable expressivity observed with these CNVs has been suggested to be a result of possible modifiers affecting the function of *CHRNA7*. One likely modifier is the human specific fusion gene, *CHRFAM7A*. In HEK293T cells, we have found that when overexpressed, α7 and CHRNA7-FAM7A colocalize and colocalize with the α7-specific antagonist α-bungarotoxin (α-bgtx), indicating that they form functional receptors. Additionally, when *CHRNA7* is coexpressed with *CHRFAM7A* or *CHRFAM7AΔ2bp*, we observe a decrease in α-bgtx fluorescence intensity, suggesting a possible dominant negative effect. In concert, utilizing patient derived human induced pluripotent stem cells (iPSCs) and iPSC-derived neurons, we are determining the molecular phenotypes and functional consequences of *CHRNA7* CNVs. In these iPSCs, mRNA levels of genes within 15q13.3 CNVs correlate to their genomic copy number. For smaller *CHRNA7* duplications spanning D-CHRNA7-LCR to BP5, mRNA overexpression suggests that the whole gene is duplicated. Additionally, duplications that may include part of *OTUD7A* do not appear to disrupt its expression. These changes in expression likely result in alterations of α7 protein expression and α7-nAChR specific calcium flux, which may be further impacted by the presence of *CHRFAM7A*. Overall, this suggests that changes in *CHRNA7* copy number and the presence of *CHRFAM7A* have functional consequences that are likely contributing to neurological phenotypes observed in patients.

2202F

Neurons derived from individuals with autism spectrum disorder show early impairments in synaptic functionality and network activity. *D.M. Dykxhoorn[1,3], C. Garcia_Serje[1], B.A. DeRosa[1,3], E. Artimovich[2], M.W. Nestor[2], D. Van Booven[1], K.C. Belle[1], H.N. Cukier[1], M.L. Cuccaro[1,4], J.M. Vance[1,3], M.A. Pericak-Vance[1,3].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Hussman Institute for Autism Research, Baltimore, MD; 3) John T. Macdonald Foundation Department of Human Gnenetics, University of Miami, Miami, FL; 4) Department of Neurology, University of Miami, Miami, FL.

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental condition characterized by social-communicative difficulties and the presence of restricted repetitive behaviors and interests. Recent studies have shown that genes harboring ASD risk loci are highly enriched in genes expressed during early cortical development and genes that function in pathways involving the regulation of transcription, chromatin remodeling, cell adhesion, signaling complexes, and synapse function. However, the impact that these genetic variants have on ASD pathophysiology is still being revealed. This is due in part to a lack of genetically-relevant human disease models. Advances in human induced pluripotent stem cell (iPSC) technology and neural differentiation techniques have made it possible to study the molecular mechanisms that underlie ASD pathology in a model system that mimics the human genetics. To that end, we generated patient-specific iPSC lines from 6 unrelated, non-syndromic ASD individuals with rare variants identified through exome sequencing and 5 gender and ethnicity matched control lines. The ASD and control iPSCs were differentiated into cortical neurons and whole transcriptome analysis was performed. In addition, since the early establishment of functional cortical neural networks is an important step in overall brain development, and may be impaired in individuals with ASD, we examined the electrophysiological characteristics of these iPSC-derived neurons. Transcriptome analysis implicate disturbances in the regulation of transcription, WNT signaling, chromatin remodeling, cell adhesion and migration, and synapse development across all time points analyzed. Over the course of neuronal differentiation, we observe key changes in genes associated with neural development and synaptic functionality, including WNT/MET signaling ($p = 4.95 \times 10^{-36}$), cell migration ($p = 1.34 \times 10^{-6}$), GABAergic neuron signal transmission ($p = 2.86 \times 10^{-9}$), and WNT-mediated axon guidance ($p = 4.71 \times 10^{-5}$). Consistent with these patterns of gene expression, cortical neurons derived from ASD individuals demonstrated significantly decreased network spiking activity from MEA recordings ($p < 0.001$) as well as decreased numbers of calcium transients ($p < 0.05$). The results of this study suggest that there may be early deficits in network activity and morphology in ASD-specific iPSC-derived neurons that is detectable at both the transcriptomic and functional levels.

2203W

Metabolic consequences of dietary interventions in the Magel2-null mouse model of Prader-Willi syndrome. *J. Bischof, C. Smolarchuk, R. Wevrick.* Department of Medical Genetics, University of Alberta, Edmonton, Alberta, Canada.

Children with Prader-Willi syndrome (PWS) or with Schaaf Yang syndrome and a deficiency of only one of the PWS genes, namely MAGEL2, have prenatal hypotonia, motor delays, and reduced muscle mass. Their myopathy contributes to the development of scoliosis, decreases caloric need, and impairs quality of life. Interventions that build muscle mass or improve exercise tolerance are key to controlling body weight in PWS. Our study focuses on dietary interventions and treatment with readily available compounds so that positive preclinical results can be readily translated into practice in the management of muscle tone and endurance in children with PWS. We found that *Magel2* is necessary for normal maintenance and function of skeletal muscle. Mice lacking *Magel2* have increased fat, decreased muscle, and reduced strength and endurance. We hypothesized that the muscle cells in mice lacking *Magel2* have a primary defect in their capacity to use carbohydrates vs. fat as fuel, and a reduced ability to reuse nutrients from the breakdown of cellular components. We tested whether dietary interventions ameliorate or worsen the muscle-related phenotypes in *Magel2*-null mice. Cohorts of 6-week old mice were fed diets that contained standard, high sucrose, high fat, or high fat then high fat-low carbohydrate (ketogenic) food. After treatment, we measured body composition and metabolic parameters in the mice and compared the findings between genotypes and among diets. These studies will inform us about the underlying bases for hypotonia and reduced muscle mass in PWS and Schaaf-Yang syndrome (MAGEL2 deficiency). These phenotypes have not yet been investigated at a genetic or molecular level. This study will also set the stage for future preclinical and clinical trials in PWS that repurpose drugs already in development for the treatment of other myopathies or forms of muscle atrophy.

2204T

Towards a complex genetic interaction model to explain phenotypic variability associated with rare copy-number variants. *M. Jensen, L. Pizzo, D.M. Singh, Q. Wang, S. Girirajan.* Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA.

Rare copy number variants (CNVs) have been associated with several complex neurodevelopmental disorders, including autism, schizophrenia, intellectual disability (ID), and epilepsy. One class of rare CNVs is characterized by phenotypic heterogeneity and the lack of known causative genes. For instance, the 16p11.2 deletion contributes to 1% of sporadic autism cases, and has also been linked to cases of ID, epilepsy, obesity, and congenital malformations. We hypothesize that interactions between genes within the CNV (in *cis*) and at second sites outside the CNV (in *trans*) determine the neurodevelopmental phenotypes associated with these CNVs, instead of only one causative gene within the CNV. In order to identify important interactions and underlying molecular pathways related to these CNVs, we performed RNA sequencing of 13 *Drosophila* knockdown models of rare CNV orthologs: *CDIPT, CORO1A, DOC2A, KCTD13, MAPK3,* and *PAGR1* in the 16p11.2 deletion; *C16orf52, CDR2,* and *UQCRC2* in the 16p12.1 deletion; and *DLG1, NCBP2, PAK2* and *PIGZ* in the 3q29 deletion. We analyzed the lists of differentially-expressed genes in each knockdown model for enrichment of gene ontology and KEGG Pathway annotations, functional relevance to neurodevelopment, and identified disease-associated genes. Among the 16p11.2 ortholog knockdowns, we found four knockdown models that resulted in a large number of differentially-expressed genes (1,084 genes in *CDIPT*, 527 in *KCTD13*, 457 in *CORO1A*, and 327 in *MAPK3*). Similarly high disruption of expression was observed in the 16p12.1 and 3q29 knockdown models. The differentially expressed genes in the four 16p11.2 knockdown models are enriched for gene ontology terms related to core cellular processes, such as amino acid metabolism and oxidoreductase activity, as well as immune response. In addition, each set of differentially expressed genes contains genes involved in embryonic development, post-synaptic density, and FMRP-binding targets, all functions important to normal neurodevelopment. The top genes in each of the four knockdown models generally show a high degree of connectivity based on protein-protein interaction and co-expression data. We will use these results to generate an interaction-based genetic network that connects genes within CNVs to known disease-associated genes in relevant pathways.

2205F

Brain effects of APOE in Down syndrome revealed by MRI. *L. Dai[1], J. Anderson[2], M. Burback[1], J.R. Korenberg[1].* 1) Department of Pediatrics, Brain Institute, University of Utah, Salt Lake City, UT; 2) Department of Radiology, University of Utah, Salt Lake City, UT.

The most significant genetic risk factor for Alzheimer's disease (AD) is the presence of allelic variants, ε4 and ε2, of APOE. It is known that about 25-70% of people with Down syndrome (DS) will develop features of AD by 60 y/o. However, it is unclear whether APOE variation is associated with developmental alterations of the brain in DS that may predispose to AD or may significantly increase the risk of AD in DS beyond trisomy 21. Previous reports of APOE phenotype-genotype in DS are conflicting. Here we report the APOE genotype and associated structural MRI (FreeSurfer) in a young age adult cohort of 17 subjects with DS and 11 controls (age 14-33), two identical twins discordant for DS and two subjects with DS caused by partial trisomy 21, one that includes, and the other that excludes APP located on HSA21. The results indicate that, whereas the APOE allele frequencies are similar in DS and controls, there is strong association of a subset of brain regions with the ε4 vs ε2 genotypes. These include ε4 associated with decreased volumes of total cortical white matter (WM), area of the medial anterior cingulum WM, and dorsal anterior cingulum WM and similar to findings in AD. The MRI data from the identical twins (22 yo, ε3/ε3), provide a unique test of the APOE-brain associations found in the population based DS cohort. The WM volumes of the normal twin is at the mean of the control cohort, and of the twin with DS is at the mean of the ε3/ε3 DS cohort. These findings strongly suggest that APOE ε4 affects brainwide WM development, particularly the WM circuitry of the medial frontal region of the left hemisphere, the anterior cingulate gyrus. We note that these results suggest a prenatal effect of APOE, ε2 and ε4 on brain development that may or not affect cognition in DS but may significantly increase the risk for AD in DS. The well-established roles of the medial anterior cingulate in error detection, conflict monitoring, emotional modulation and attention reflect a close correspondence with the cognitive and behavioral deficits of DS. Disturbed connectivity of the ACG, its association with APOE4 and its possible relationship to Von Economo neurons provide testable human cellular and organismal models and a truly fresh look at Alzheimer's disease.

2206W

Parsing the genetic brain connectome using rare identical twins discordant for Down syndrome. *J.R. Korenberg[1], J. Anderson[2], M. Burback[1], J. Edgin[3], G. Gerig[4], L. Dai[1].* 1) Department of Pediatrics, University of Utah, Salt Lake City, UT; 2) Department of Radiology, University of Utah, Salt Lake City, UT; 3) Department of Psychology, University of Arizona, Tucson, AZ; 4) New York University Tandon School of Engineering, Brooklyn, NY.

Down syndrome (DS) is a major cause of mental retardation and congenital anomalies. Understanding the genetic or neurocognitive effects of trisomy 21 has always been severely limited by the heterogeneity of genes located on chromosome 21 and by the enormous variations in the remainder of the 30,000 genes located on the other chromosomes. Identical twins are therefore nature's unique opportunity to see the dramatic effects of identical genes on phenotype. Here we report the structural and functional MRI, DTI, and extensive cognitive and genetic measures of a pair of identical twins discordant for DS and compared them to the typical DS and TC cohort. We determined chromosomal features of DS in one, >96% trisomy 21 in cord blood, blood and fibroblasts, and in the other a normalized phenotype, >92% normal karyotype in these tissues. Structural MRI showed distinguishable patterns in most but not all regions between the identical twins. Normalized volumes of the 96% mosaic twin were very close to the DS group, and the 8% mosaic twin volumes were similar to controls. An exception was DS-like volume in the 8% mosaic twin of the right fusiform gyrus, involved in visual memory and language. Only the 96% mosaic twin demonstrated increased internetwork synchrony across networks seen in typical DS during functional MRI. The twins are a graphic vision of how a simply trisomy for 21 disturbs brain wiring and the results implicate the hippocampal and fusiform gyrus formation in the neural systems responsible for visual-spatial and linguistic deficits of DS. .

2207T

Role of a circadian-relevant gene, *NR1D1*, in brain development: Possible involvement in the pathophysiology of autism spectrum disorder. *M. Goto[1], M. Mizuno[2], A. Matsumoto[1], Z. Yang[1], E. Jimbo[1], H. Tabata[2], H. Osaka[1], K. Nagata[2,3], T. Yamagata[1].* 1) Pediatrics, Dept, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Molecular Neurobiology Dept, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan; 3) Neurochemistry Dept, Nagoya University Graduate School of Medicine, Nagoya, Japan.

Autism spectrum disorder (ASD) is frequently accompanied by comorbid conditions, and associated with problems in the early developmental period including hyperactivity, panic, self-injury, and sleep disturbance. Among them, sleep disruption (such as insomnia or a short sleep cycle) is one of the most common and distressed problems. In our previous study (Yang *et al*. 2015), we screened ASD patients with and without sleep disorders for mutations in the coding regions of circadian-relevant genes, and detected mutations in several clock genes including *NR1D1*. Thus, circadian-relevant genes likely represent impaired molecular clock mechanisms that potentially contribute to the etiology of ASD. Here, we further screened ASD patients for *NR1D1* mutations and identified three new mutations, including a *de novo* heterozygous mutation, c.1499G>A (p.R500H). We then examined the role of NR1D1 in development of the mouse cerebral cortex. Acute knockdown of mouse *NR1D1* by *in utero* electroporation caused abnormal positioning of cortical neurons during corticogenesis. This aberrant phenotype was rescued by wild type NR1D1, but not by the c.1499G>A mutant. Moreover, knockdown of *NR1D1* also suppressed axon extension, while proliferation of neuronal progenitors and stem cells at the ventricular zone was not affected. Taken together, NR1D1 plays a pivotal role in corticogenesis via regulation of excitatory neuron migration and synaptic network formation. Consequently, functional defects in NR1D1 may relate to ASD pathophysiology.

2208F

Effects of altered 7q11.23 copy number on neural stem cell physiology and cortical organization during murine embryonic development. *H.M. Oh[1], F. Miller[1,2,3], L. Osborne[1,2].* 1) Institute of Medical Sciences, University of Toronto, Toronto, Ontario, M5S 1A8; 2) Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 1A8; 3) Developmental and Stem Cell Biology, Hospital for Sick Children, ON, M5G 1X8.

Williams Syndrome (WS) and 7q11.23 Duplication Syndrome (Dup7) are two rare neurodevelopment disorders caused by deletion and duplication respectively, of 25 genes on chromosome 7q11.23. Individuals with WS or Dup7 present with an array of cognitive and behavioural phenotypes, including anxiety, autism, speech and language delay, and intellectual disability, but the biological basis for these neuropsychiatric symptoms remain unknown. Structural and functional abnormalities of the cortex have been identified in both WS and Dup7, suggesting that cortical development is perturbed in these disorders. Our lab has generated mouse models with altered gene dosage of two candidate genes from the WS commonly deleted region, General Transcription Factor 2 I (*Gtf2i*) and GTF2I Repeat Domain containing protein 1 (*Gtf2ird1*). These models include mice with hemizygous deletion of both genes (*Gtf2i/Gtf2ird1*[+/-]) and mice with duplication of *Gtf2i* (*Gtf2i*[+/Dup]). We have used these mice to determine whether altered expression of *Gtf2i* and/or *Gtf2ird1* affects neural stem cell growths and cell-fate, and potentially contributes to the neurological features of WS and Dup7. Our results showed that *Gtf2i/Gtf2ird1*[+/-] mice with a hemizygous deletion had an overall reduction in the number of neuronal precursors and neurogenesis with an increase in astrogenesis whereas mice with an additional copy of *Gtf2i* had an increase in the number of precursors and neurogenesis along with a decrease in astrogenesis. These early changes in neural development translated into perturbed cortical organization with changes in cortical thickness and cell packing density in mid-gestation developing brain. These data will help us understand how the developing brain is perturbed in people with WS and Dup7 and may give insight into an answer the pathological causes and possibly help identify therapeutic interventions to treat the cognitive disorders involving autism and social dysfunction.   .

2209W

***ARID1B* disorders of intellectual development: Expanding the phenotype and insights from** *in vitro* **and** *in vivo* **models.** *C. Dias[1,2], M. Levitin[1], G. Sanchez-Andrade[1], S.S. Gerety[1], P.H. Tate[1], F. Guillemot[2], W.C. Skarnes[1], M.E. Hurles[1], D.D.D. Study[1], D.W. Logan[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) The Francis Crick Institute, Mill Hill Laboratories, London, United Kingdom.

Whole exome and genome sequencing of large cohorts of individuals with intellectual and developmental disability (IDD) have revealed that heterozygous mutations in *ARID1B*, of the BAF swi/snf chromatin-remodeling complex, are associated with IDD. Originally identified in individuals with Coffin-Siris Syndrome (CSS), *ARID1B* mutations are associated with both syndromic and non-syndromic IDD. We show that in a cohort of individuals ascertained for developmental disorders, *ARID1B* mutations are identified in individuals with variable clinical features, most commonly with intellectual, developmental or speech and language delay. De novo mutations in *ARID1B* are a leading cause of IDD. We explore the phenotypic spectrum of ARID1B-IDD, identifying inconstant but highly discriminatory features, such as agenesis of the corpus callosum and laryngeal anomalies. To further investigate the role of *ARID1B* in development and cognition, we examined the effect of heterozygous loss of function in mouse models. We show that heterozygote mutant mice recapitulate morphological and behavioral features seen in patients, namely growth delay and long-term memory defects. To examine the effect of *Arid1b* haploinsufficiency on transcriptional regulation, we performed genome-wide RNASeq in cellular and *in vivo* models of *Arid1b* haploinsufficiency. Heterozygous loss of *Arid1b* does not significantly affect transcription in mouse ES cells, but induces significant gene expression deregulation during neural stem cell differentiation. Likewise, haploinsufficient embryonic mouse brain at E14.5 shows an altered transcriptional profile suggestive of a role for *Arid1b* in axonal guidance and cell-cell interaction. Mutant adult hippocampus shows an abnormal gene expression profile: it implicates axon guidance pathways previously identified in another model of BAF complex-associated-IDD, and genes involved in synaptogenesis and long-term potentiation. Dysregulation of genes associated with other neurodevelopmental disorders are over-represented in post-natal mutant hippocampus. Hence, we demonstrate a crucial role for *ARID1B* in brain development and long-term memory. In the mouse model, we identify pathways affected by *Arid1b*-dependent BAF complex disruption at different developmental stages. Our study underscores the relevance of mammalian models to further elucidate IDD and related disorders.

2210T

Reciprocal brain size defects and intellectual disability in patients with activating mutations in the mTOR pathway. *M.R.F. Reijnders[1,8], M. Kousi[2,8], G.M. van Woerden[3,8], J. Bralten[1], G.M. Mancini[4], T. van Essen[5], C.M. Marcelis[1], M. Proietti-Onori[3], E. Smeets[6], M. van Gastel[7], A.P.A. Stegman[6], S.J.C. Stevens[6], C. Gilissen[1], R. Pfundt[1], P.L. Tan[2], T. Kleefstra[1], Y. Elgersma[3,8], N. Katsanis[2,8], H.G. Brunner[1,6,8].* 1) Human Genetics, Radboudumc, Nijmegen, The Netherlands; 2) Center for Human Disease Modeling, Duke University, Durham, North Carolina, USA; 3) Department of Neuroscience and ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) Department of Clinical Genetics, Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands; 5) University of Groningen, University Medical Center of Groningen, Department of Genetics, Groningen, The Netherlands; 6) Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; 7) Department of medical care, SWZ zorg, Son, The Netherlands; 8) These authors contributed equally.

Activating *de novo* or somatic mutations in the mTOR signaling pathway have been associated with a host of neurodevelopmental traits, hallmarked by megalencephaly and intellectual disability (ID). To test systematically the contribution of this pathway to sporadic ID, we evaluated 102 genes from the mTOR pathway in 820 ID patients. We identified 14 *de novo* mutations within 9 different genes: 4 known ID genes (*PIK3R1; PIK3R2; RAF1; PPP2R5D*) and 5 candidate ID genes (*RHEB; RAC1; PPP2R5E; PPP2CA; ERK1*). Aberrant brain size was present in the majority of patients (67%) and this percentage was significantly increased compared to patients with mutations in known ID genes that are not involved in the mTOR pathway (28%). We focused further on the pathomechanism of 4 germline missense mutations in 2 candidate ID genes: *RHEB*, an mTORC1 component, and *RAC1*, that signals through mTORC2 and the MAPK cascade. Whereas patients with *RHEB* mutations show severe ID, hypotonia, epilepsy and macrocephaly, patients with *RAC1* mutations have ID, epilepsy and microcephaly. Functional modeling of the discovered mutations in zebrafish and mouse embryos, as well as in primary neuronal cultures, showed all alleles to be hyperactivating and to induce defects in neuronal migration, proliferation and soma size. A direct opposite effect to overall brain size for *RAC1* and *RHEB* mutants, concordant to the human phenotype, was present and could be ameliorated by the administration of rapamycin *in vivo*. Motivated by the high frequency of aberrant brain size in ID syndromes within the mTOR pathway, we queried the Discovery cohort of the ENIGMA brain volume genome-wide association dataset; we found a significant association of mTOR signaling components with brain volume in the population ($p_{self-contained} = 0.003447$ and $p_{competitive} = 0.041912$). Taken together, our data measure and expand the contribution of mTOR signaling components to syndromic ID. They also suggest the presence of a continuum of effect in this pathway, wherein highly deleterious *de novo* alleles drive pathological conditions whereas likely milder effects contribute to normal variation in brain volume.

2211F

Genomic and functional studies provide evidence for two-hit model for neurodevelopmental disorders. *L. Pizzo, A. Polyak, M. Jensen, E. Huber, J. Yoon, A. Kubina, D. Mayanglambam, S. Girirajan.* Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

We present detailed phenotypic, genomic, and functional analysis of the 16p12.1 deletion, provide evidence for causality of specific genes within 16p12.1, and assess their genetic interactions within specific cellular and developmental pathways. We performed deep phenotyping of 100 families with the deletion and CNV and exome sequencing analysis in 30 families, and identified several loss-of-function variants within functional gene classes associated with various developmental pathways (*SETD5, LAMC3, MIER2, MAPK4* and *CEP135*), and large disease-associated CNVs including 3q29 deletion, 15q13.3 deletion, and 22q11.2 duplication. Clinical data suggested increased severity of phenotypes in probands with additional hits affecting various domains of development and behavior, compared to those with single hits. We further performed functional evaluation by decreasing the expression of the four (*UQCRC2, C16ORF52, POLR3E*, and *CDR2*) 16p12.1 orthologs in *Drosophila melanogaster*. Using a series of tissue-specific promoters and RNA-interference transgenic lines, we assessed dosage sensitivity of the orthologs by testing developmental and molecular phenotypes. Neuronal-specific knockdown led to increased head sizes for *CDR2* (6.5 SD, p<0.0010) and *POLR3E* (5.2 SD, p<0.0010), and a severe morphological eye defects for *POLR3E* and *C16orf52* (p<0.05). Ubiquitous knockdown resulted in early lethality for *POLR3E* and *UQCRC2* fly lines, and neuron-specific knockdown of *POLR3E* resulted in severe motor defects and early adult lethality. To further map specific pathways associated with these phenotypes, we performed RNA sequencing of *Drosophila* brain samples with neuron-specific knockdown and identified several genes belonging to core biological pathways. To fully characterize the contribution of 16p12.1 orthologs to developmental phenotypes, we performed interaction studies using pair-wise knockdown of genes within the CNV. Variable phenotypes were detected for all interactions containing *POLR3E*, suggesting a major role for *POLR3E* for the observed phenotypes, while other genes in the region potentially modify the phenotype. We also tested *trans* interactions of 16p12.1 orthologs with second hits identified in patient cohorts, and found several enhancers and suppressors of the phenotype. Our results provide functional support for the proposed two-hit model and potentially identify novel interacting genes that modulate the 16p12.1 phenotype.

2212W

A human-specific duplicated gene in the *1q21* **locus has a potential function in brain development.** *G.A. Lodewijk[1], F.M.J. Jacobs[1], A.D. Ewing[5], A.M. Novak[2,4], S. Katzman[2], S.R. Salama[2,3], D. Haussler[2,3].* 1) Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands; 2) UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, United States of America; 3) Howard Hughes Medical Institute, Bethesda, MD, United States of America; 4) Biomolecular Engineering, University of California, Santa Cruz, CA, United States of America; 5) Mater Research Institute, University of Queensland, Brisbane, Australia.

The human brain is characterized by a remarkable expansion over the last million years. However, the underlying genetic events related to specifying human brain development remain largely unknown. We hypothesize that gene duplication events have contributed significantly in changing gene regulatory pathways of the brain. In previous work (Jacobs *et al.*, submitted) we cultured human and rhesus monkey embryonic stem cells (ESCs). Under defined conditions, these ESCs were grown into cortical organoids, representing the early embryonic neocortex *in vitro*. At different time points during differentiation, cortical organoids were profiled by RNA-sequencing. In these data we found a human-specific duplicated gene (*HUSD1*) that was differentially expressed between human and rhesus monkey cortical organoids. The *HUSD1* gene cluster consists of several duplications located in the *1q21* locus. The instability of the *1q21* locus has led to many genomic rearrangements of this locus in recent evolutionary history. Specifically in the human genome the *1q21* locus harbors many copy number variations (CNVs) related to brain disorders. To test the significance of *HUSD1* in brain development, we have done two key experiments: 1) Ectopic expression of *HUSD1* in the neocortex of embryonic mice via *in utero* electroporation, revealing it can modulate brain development. 2) RNA-sequencing of mouse cortical organoids, derived from mouse ESCs stably expressing *HUSD1*, which show enhanced expression of early forebrain markers. These data indicate that *HUSD1* is a novel human factor regulating brain development. Our future efforts will largely focus on unraveling the molecular mechanisms of *HUSD1* in human brain development, as well as its potential involvement in *1q21* CNV related brain disorders.

2213T

Exome sequencing of 311 trios confirms the importance of *de novo* **mutations in intellectual disability.** *T. Schwarzmayr[1], H.J. Luedecke[2], A.M. Zink[3], E. Graf[1], N.C. Bramswig[2], T. Wieland[1], J. Becker[3], R. Berutti[1], J.C. Czeschik[2], K. Cremer[3], T. Meitinger[1,5], H. Engels[3], D. Wieczorek[2,4], T.M. Strom[1,5].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Institute of Human Genetics, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; 3) Institute of Human Genetics, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany; 4) Institute of Human Genetics, Heinrich-Heine-University, Medical Faculty, Düsseldorf, Germany; 5) Institute of Human Genetics; Technische Universität München, Munich, Germany.

Intellectual disability (ID) has an estimated prevalence of 1-2%. Recently, exome sequencing of parent-child trios has shown that a large proportion of ID are caused by de novo mutations. We sequenced exomes of 311 children with ID and their parents to an average read depth of 114. CNVs have previously been excluded in most patients. Variant data was generated through an automated pipeline using multiple variant and CNV callers and stored in a custom database. Variants were annotated in a cooperative way through a Web-based frontend and confirmed by Sanger sequencing. We identified 556 de novo variants (116 synonymous, 330 missense, 9 in-frame indels, 101 loss-of-function), with an exomic mutation rate of 1.83 per individual per generation. 52 patients (17%) had at least one protein altering de novo variant in a gene already associated with a developmental disease by monoallellic mutations (291 genes from the DDG2P database version July 2015). 60 patients (19%) had no de novo variant. These variants in combination with data from collaborating groups were instrumental in identifying two genes not yet associated with ID, HIVEP2 and CHAMP1. Additionally, we extended the phenotypic spectrum of protein altering variants in seven genes: ASXL3, KCNH1, SETD5, CTNNB1, ECHS1, IQSEC2 and TCF20.

individuals	de novo	synonymous	missense	indel	LoF
311 cases	556	116 (0.21)	330 (0.59)	9 (0.02)	101 (0.18)
50 controls	58	22 (0.38)	31 (0.53)	0	5 (0.09)

We compared these de novo variants to a set of 58 de novo variants identified in 50 healthy control trios, corresponding to an exomic mutation rate of 1.16 per individual per generation. 17 controls (34%) had no de novo variant. Interestingly, the difference in the synonymous mutation rate was not significant (p=0.51; Mann-Whitney-Wilcoxon test, two-sided), however the difference in the protein altering mutation rate was highly significant (p=2.37×10^{-5}). When comparing only the 253 cases, excluding cases with de novo variants in genes already associated with developmental disorders or in one of the nine genes described above, to the 50 controls, the difference between protein altering mutation rates remains highly significant (p=7.11×10^{-4}). In summary, we show that exome sequencing in a multicenter setting with an appropriate IT environment can efficiently be used to generate clinical grade diagnoses by integrating the advantage of standardized central sequencing and distributed evaluation and annotation of the resulting data in the clinical context.

2214F

Discovery of Vimentin as a novel target under the dynamic regulation of FMRP. *H. Gong[1], G.M. Gafford[1], F. Zhang[1], S.T. Warren[1,2].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA.

Fragile X syndrome (MIM 309550) is the most common cause of intellectual disability and a well-known cause of autism spectrum disorder. Fragile X syndrome is caused by the absence or dysfunction of fragile X mental retardation protein (FMRP), a RNA-binding and translation regulation protein. There are several comprehensive RNA studies that have identified putative FMRP bound sequences and genes through various methods. However, no studies have particularly focused on the identification of the nascent peptides that are stalled by FMRP on the ribosomes. We applied the puromycin-associated nascent chain proteomics (PUNCH-P) method to post-runoff hippocampal slices. By comparing wildtype (WT) and *Fmr1* knockout (KO) mice with and without DHPG stimulation, we discovered Vimentin as the DHPG inducible translation suppression target of FMRP. We also found that in the absence of FMRP, the mRNA and protein levels of Vimentin are elevated both *in vivo* and *in vitro*. Additionally, further experiments indicated that FMRP interacts with Vimentin mRNA and affects its stability. These results together reveal that Vimentin is under the dynamic regulation of FMRP. Identification of the FMRP binding site on Vimentin mRNA is currently ongoing.

2215W

Identification of alternative *FMRP* translation start sites by non-ATG codons. *Y. Wen[1], S.T. Warren[1,2].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA.

Mutations in *FMR1* gene cause Fragile X syndrome (MIM 309550), a frequent cause of inherited intellectual disability. A trinucleotide $(CGG)_n$ repeat expansion in the 5'-untranslated region of *FMR1* gene is the most common mutation, which results in the epigenetic silencing of the gene and the absence of the encoded protein FMRP. The aim of this project is to identify non-ATG-initiated translation of FMRP that could be directed by the hairpin-forming trinucleotide repeats. In our study, we have demonstrated that there are several novel non-canonical translation initiation codons located upstream of the main open reading frame (ORF) of the *FMR1* gene. We also confirmed the existence of these non-ATG start sites for FMRP synthesis in the mouse hippocampus by immunoprecipitation. Our functional analysis for two possible non-ATG start codons (-219ACG & -69CTG), located before and right after the $(CGG)_n$ repeats, showed that in addition to initiating the expression of FMRP at alternative start sites, these potential non-canonical start codons can also affect FMRP expression initiated from classic ATG start. Interestingly, removing the $(CGG)_n$ repeats increased the expression of the alternative FMRP products but decreased the total amount of FMRP. Taken together, our results identify alternative FMRP translation start sites by non-ATG codons in both human and mouse *in vitro* and *in vivo*, providing new evidence of non-ATG initiated translation.

2216T

Driver genes for 2p15p16.1 microdeletion syndrome identified using clinical, genomic and functional analysis. *E. Rajcan-Separovic[1], H. Bagheri[1], C. Badduke[1], Y. Qiao[1], R. Colnaghi[2], D. Alcantara[2], I. Abramowicz[2], C. Dunham[3], J. Wen[1], R. Wildin[4], M. Nowaczyk[5], J. Eichmeyer[6], A. Lehman[3], B. Maranda[7], S. Martell[1], X. Shan[8], S. Lewis[3,9], M. O'Driscoll[2], C. Gregory-Evans[8].* 1) Dept Pathology (Cytogenetics), University of British Columbia, Vancouver, Canada; 2) Human DNA Damage Response Disorders Group, Genome Damage & Stability Centre, University of Sussex, Brighton, United Kingdom; 3) BC Children's Hospital, Vancouver, Canada; 4) National Human Genome Research Institute, Bethesda, Maryland, United States; 5) Department of Pathology and Molecular Medicine, McMaster University Medical Centre, Hamilton, ON, Canada; 6) St. Luke's Children's Hospital, Boise, ID, United States; 7) Medical Genetics, Universite de Sherbrooke, Sherbrooke,Canada; 8) Department of Ophthalmology, University of British Columbia, Vancouver, Canada; 9) Department of Medical Genetics, University of British Columbia, Vancouver, Canada.

   Introduction: The 2p15p16.1 microdeletion syndrome has a core phenotype consisting of intellectual disability, microcephaly, hypotonia, delayed growth, common craniofacial features, and digital anomalies. So far, more than 30 cases with this syndrome have been reported in literature; however, the size of the deletions and their breakpoints vary making the identification of candidate genes challenging. Recent reports pointed to four genes (*XPO1*, *USP34*, *BCL11A* and *REL*) which were included, alone or in combination, in the smallest deletions causing the syndrome.   **Results**: Herein, we describe 8 new cases with 2p15p16.1 deletion and review all published cases. We demonstrate functional deficits for the above4 candidate genes using patient lymphoblast cell lines (LCLs) and knockdown of their orthologs in zebrafish. All genes were dosage sensitive based on reduced protein expression in LCLs. In addition, deletion of XPO1, a nuclear exporter, co-segregated with nuclear accumulation of one of itscargo molecules (rpS5) in patient LCLs. Other pathways associated with these genes (e.g. NF-kB and Wnt signaling as well as DNA damage response) were not impaired in patient LCLs. Knockdown of *xpo1a*, *rel* , *bcl11aa* and *bcl11ab* resulted in abnormal zebrafish embryonic developmentincluding microcephaly, dysmorphic body, hindered growth, small fins as well as structural brain abnormalities. **Conclusions**: Our multifaceted analysis strongly implicates *XPO1*, *REL*, and *BCL11A* as candidate genes for the 2p15p16.1 microdeletion syndrome. It also demonstrates the benefits of using zebrafish for identification of driver genes for genomic disorders. Our current work includes overexpression of the candidate genes in zebrafish, to determine if they cause mirror image phenotypes (e.g. macrocephaly).

2217F

Defects of adult hippocampal neurogenesis unify genetically distinct causes of Kabuki syndrome. *G. Carosso[1,2], J. Weissman[4], J. Fahrner[1,3], L. Zhang[1], H.T. Bjornsson[1,3].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA; 2) Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD, USA; 3) Department of Pediatrics at the Johns Hopkins University School of Medicine, Baltimore, MD, USA; 4) Kennedy Krieger Institute, Baltimore, MD, USA.

   Kabuki syndrome (KS) is a genetically determined cause of intellectual disability resulting from loss-of-function (LOF) mutations in two histone-modifying enzymes, lysine-specific methyltransferase 2D (KMT2D) and lysine-specific demethylase 6A (KDM6A/UTX). These enzymes alter methylation states on histone tails at H3K4 and H3K27, respectively, and promote transcriptionally permissive open chromatin states despite opposing enzymatic activities. However, it remains poorly understood how these distinct genetic causes produce overlapping neurobehavioral phenotypes in KS. Previously, we demonstrated a reversible deficiency of adult hippocampal neurogenesis in the subgranular zone (SGZ) of the dentate gyrus (DG) in a *Kmt2d* LOF mouse model, suggesting that intellectual disability in KS may be amenable to postnatal therapy. We now show reduced cellular proliferation in the SGZ, visualized by EdU incorporation, in a second, genetically distinct mouse model of KS with *Kdm6a* LOF. This points to deficient adult neurogenesis as a unifying pathogenic feature of KS, leading us to hypothesize that heterozygous LOF mutations in both KS genes perturb chromatin accessibility and alter transcription of shared neurogenic gene expression programs. Using confocal microscopy, we see reduced activation of radial glia-like (RGL) neural stem cells (NSCs) in SGZ (25% decrease, p<0.05), indicating disruption of the first discrete neurogenic step. Similar reduction is seen in migrating neuroblasts expressing doublecortin (DCX) in the SGZ (22% decrease, p<0.01), culminating in fewer mature neurons (NeuN+ cells) in the DG granule cell layer (9% decrease). Interestingly, neuroblast deficiency is associated with reduced fluorescence intensity for the open chromatin mark H3K4me3 within DCX+ cells (p<0.05), supporting a mechanistic link between altered histone modifications and neurobehavioral phenotypes. To identify target genes dysregulated secondarily to KS mutations, we are performing next-generation RNA sequencing and genome-wide chromatin profiling in hippocampal NSCs harboring *Kmt2d* and *Kdm6a* LOF mutations. Our results demonstrate that genetically distinct KS types lead to common defects of an ongoing, postnatal neurogenic process accompanied by cell type-specific histone methylation alterations. Targeting this unified pathogenic mechanism at precise stages may enable a generalized rescue strategy for intellectual disability in genetically distinct forms of KS and related disorders.

2218W

Rapid *in vitro* evaluation of variants of uncertain significance from patients with developmental delay and/or intellectual disability. *K.L. Engel[1], J.N. Cochran[1], A.A. Hardigan[1], M.D. Amaral[1], K.M. Bowling[1], C.R. Finila[1], S.M. Hiatt[1], M.L. Thompson[1], D.E. Gray[1], J.S. Whittle[1], W.V. Kelley[1], M.E. Cochran[1], K.M. East[1], N.E. Lamb[1], C.A. Rich[1], K.B. Brothers[1], E.J. Lose[1], S.B. Simmons[2], E.M. Bebin[2], G.S. Barsh[1], G.M. Cooper[1], R.M. Myers[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Alabama at Birmingham, Birmingham, AL.

Developmental delay and intellectual disabilities (DD/ID) include devastating phenotypes and comprise a large fraction of rare undiagnosed conditions in children. Unfortunately, little is known about the cellular mechanisms that lead to disease, and as a consequence, therapeutic advancements have suffered. Previous studies have indicated a strong genetic component to these phenotypes, and successful identification of causal genetic variants through exome or whole genome sequencing sometimes leads to clinical diagnoses, revision of treatment strategies, community and support network building, and increased quality of life. However, sequencing-based diagnostic efforts typically solve only a subset of cases; while precise numbers vary according to ascertainment and study enrollment criteria, large fractions of children cannot be given a precise genetic diagnosis even after whole genome sequencing. In our own work as part of the Clinical Sequencing Exploratory Research (CSER) consortium, we have found diagnostic genetic variants in ~27% of children with conditions refractory to standard diagnostic tests. Rigorous experimental evaluation of variants and genes implicated in DD/ID is needed to increase diagnostic rate, particularly for the subset of the undiagnosed cases where a potential genetic cause is identified but not confirmed, termed to be variants of uncertain significance (VUSs). VUSs typically arise as a result of a lack of information about the relevance (or lack thereof) of a given gene to disease, the impact (or lack thereof) of a given variant on gene function, or both. In our studies to date, ~15% of affected children harbor a VUS. Introduction of these VUSs, a portion of which are in gene expression regulators, into human neurons derived from neural precursor cells through gene editing technology will provide evidence for or against the association of these sequence variants on key molecular and cellular phenotypes including global gene expression at the levels of both transcription and translation, neuronal excitability, and synapse and/or ion channel composition profile. The studies proposed here will provide critical insights regarding the biological roles of genes and variants associated with DD/ID and establish a framework for future mechanistic interrogation of genetic variation. Furthermore, the insights gained from this work will inform future studies aimed at therapeutic intervention for under– and undiagnosed DD/ID disorders.

2219T

Functional investigation of Early B-Cell Factor 3 (*EBF3*) haplo-insufficiency as a novel genetic cause of developmental delay and intellectual disability. *A. Hardigan[1,2], D. Savic[2], B. Roberts[2], J. Prokop[2], R. Ramaker[2], K. Bowling[2], S. Hiatt[2], C. Holden[2], G. Cooper[2], R. Myers[2].* 1) University of Alabama at Birmingham, Department of Genetics, Birmingham, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL.

Developmental delay and intellectual disability (DD/ID) is characterized by limitations in cognitive and adaptive behavior and represents a heterogeneous group of syndromic and non-syndromic disorders. DD/ID is estimated to have a worldwide prevalence of 1-3% and can occur in combination with congenital malformations and other neurologic disorders such as epilepsy. Genetic mutations represent a major cause of established DD/ID cases and over 700 genes have been linked to various forms of DD/ID. The identification and functional characterization of new genes involved in DD/ID has important implications for diagnosis and clinical management. As part of the Clinical Sequencing Exploratory Research (CSER) program, our lab has employed trio-based whole-exome and whole-genome sequencing (WES/WGS) of almost 250 patients and unaffected parents to identify the potential underlying genetic causes of their respective DD/ID conditions. We have identified two DD/ID patients with different VUS in the atypical helix-loop-helix (HLH) transcription factor (TF), early B-cell factor 3 (*EBF3*). Both *EBF3* VUS are SNVs, causing a p.P177L missense mutation in the DNA-binding domain and an IVS11+1G>T mutation disrupting a D-1 splice site in the HLH dimerization domain, respectively. Although not a previously documented DD/ID gene, *EBF3* is a member of the EBF TF family that has been shown to control migration of Cajal-Retzius cells from the cortical hem and drive expression of genes critical for neuronal differentiation, making it a prime candidate. However, relatively little is known about how *EBF3* fits into the intricate gene regulatory circuitry underlying neural development In order to identify a causal mechanistic basis for the identified *EBF3* mutations in DD/ID etiology, we have comprehensively assessed the role of each VUS on DNA binding affinity and global genome-wide occupancy using ChIP-seq, and determined alterations in the gene expression programs regulated by *EBF3* with RNA-seq. These experiments demonstrate that the VUS disrupt *EBF3* function via reducing DNA-binding activity and disrupt critical neurodevelopmental gene expression networks. In conjunction with *EBF3* VUS patient data, this work supports the establishment of *EBF3* as a new DD/ID gene and serves as a framework for its further functional validation.

2220F

Zebrafish models of Marinesco-Sjögren syndrome. *G. Kawahara, Y. Hayashi.* Pathophysiology, Tokyo Medical School, Tokyo, Japan.

Marinesco-Sjögren syndrome (MSS) is an autosomal recessive disease characterized by progressive myopathy, cerebellar ataxia, mental retardation, and congenital cataracta. MSS is caused by mutations in the *SIL1* gene. SIL1 is known as a nucleotide exchange factor for the endoplasmic reticulum chaperone BiP. Zebrafish disease models are useful to analyze the mechanism of various diseases, including neuromuscular disorders. We analyzed zebrafish sil1 function using antisense morpholino oligos. To create MSS model fish, we have started to create MSS knocked-out fish by CRISPR-Cas9 system. Two different morpholinos were injected at 1-2 cell stage. At 4 days post-fertilization (dpf), thirty percent of fish injected with morpholino 1 and 2 showed reduced birefringence. Morphants had small diameter of eyes and reduction of purkinje cells in cerebellar area. Co-injection of zebrafish *sil1* mRNA along with the morpholino restored normal development the morphants. Additionally, BiP (a ER stress marker protein), LC3 (an autophagy marker protein) and activated caspase 3 (an apoptosis marker) were increased in these morphants. With CRISPR/Cas9 system, we successfully identified founders that could pass mutations in the *sil1* gene through the germline. Their offspring carried indel mutations, suggesting that these founders carrying heritable mutations. These findings suggest that it will be feasible to create a MSS model fish for a therapeutic chemical screening.

2221W

Aberrant splicing of *ACADM* exon 5: A balance between enhancers and silencers. *L.L. Christensen, G.H. Bruun, M.R. Larsen, B.S. Andresen.* Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark.

Correct splicing of vulnerable constitutive exons is dependent on exonic splicing enhancers (ESE) and exonic splicing silencers (ESS). Therefore sequence variations that create or disrupt ESS/ESEs may cause missplicing and disease, irrespective of the predicted consequences according to the genetic code. Sequence variants in *ACADM* may cause MCAD deficiency (MCADD). Because MCADD is frequent, newborns are screened by MS/MS and diagnosis is confirmed by mutation analysis. Therefore, the potential deleterious effect of identified variants needs to be correctly assessed. We previously identified a frequent missense variation, c.362C>T in exon 5, which abolishes a fundamental ESE, causing exon skipping, loss of functional protein and consequently MCADD. The c.362ESE functions both by antagonizing a juxtaposed c.351ESS, and it is required to define the weak exon 5 3' splice site. Notably, a silent polymorphic variation, c.351A>C inactivates the c.351ESS, and makes splicing immune to the deleterious c.362C>T. We speculated that *ACADM* exon 5 is vulnerable, and investigated presumed silent and missense variants identified in different parts of exon 5 in newborns with MCADD. Minigene studies revealed that several of the variants cause exon skipping by abolishing different ESEs. Employing siRNA/overexpression of splicing regulatory factors (SRFs) and RNA pull-down, we characterized new ESEs and ESSs and the binding SRFs. We show that the ESEs and ESSs need to be in a finely tuned balance to allow exon 5 inclusion. This balance is not only dependent on steric competition of SRFs for binding to juxtaposed ESE/ESSs, as previously demonstrated for the c.362ESE and c.351ESS. Instead the overall balance between ESEs/ESSs located throughout the exon needs to be maintained. In line with this we show that a c.331G>A missense variation abolishes an ESE to cause complete exon skipping. Remarkably, the polymorphic c.351A>C, which inactivates the distant c.351ESS, also makes splicing immune to the deleterious effect of c.331G>A. Moreover, inactivation of an ESS at pos. c.325, abolishes the splicing defect of the distant c.362C>T variant as well as that of the juxtaposed c.331G>A. We conclude that splicing of weak exons can be disrupted by several different missense/silent variations. Moreover, weak exons are critically dependent on the overall balance between ESE/ESSs located throughout the exon. It is thus essential to evaluate effects on splicing in the context of the haplotype.

2222T

lincRNA gene expression in the mammalian inner ear. *K. Ushakov[1], T. Koffler-Brill[1], A. Rom[2], I. Ulitsky[2], K.B. Avraham[1].* 1) Human Molecular Genetics and Biochemistry, Tel Aviv University, Tel Aviv, Israel; 2) Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel.

Mammalian genomes are highly complex and regulated at multiple levels. A novel class of RNA molecules, the long intervening non-coding RNAs (lincRNAs), imposes one of these levels of regulation. lincRNAs are transcripts of >200nt in length and similarly to messenger RNA (mRNA), they are capped, polyadenylated and spliced. However, unlike mRNAs, they are less abundant and possess a higher tissue specificity. lincRNAs are versatile and exert their regulation at several cellular levels and have been linked to development and disease. The mammalian inner ear is a complex organ comprised of the cochlea, responsible for hearing and the vestibule, responsible for balance. To reach such tissue intricacy, several levels of regulation are required and thus the development of these organs involves complex genetic programs, orchestrated by genetic and epigenetic factors. Though largely unexplored, we hypothesize that lincRNAs play a significant role in the inner ear as well. As a result, our goal is to build a complete inventory of the lincRNAs that are expressed throughout mouse inner ear development and to understand their impact on gene regulation. We generated RNA-seq data by high-throughput sequencing of sensory epithelia containing hair and supporting cells at two developmental stages. The reads were subjected to a pipeline for lincRNA profiling. We identified 1920 lncRNA genes, of these 403 are novel unannotated lincRNAs. Based on a number of criteria we chose several lincRNAs for further investigation, which included profiling in the whole animal and tissue and developmental stage expression. Next we focused on studying lincRNAs that are in close proximity to genes important for hearing and deafness in order to identify whether there are any cis-regulatory mechanisms operating between these gene pairs. This characterization of lincRNAs will aid us to comprehend the auditory and vestibular neurosensory systems, towards understanding mechanisms of hearing and balance disorders. .

2223F

Pax2/Pax2a **regulates** *Cyp1b1/cyp1b1* **expression in mice and zebrafish.** *L.A. Schimmenti[1], P.R. Preorius[2], M. McGrail[3], S.L. Solin[3], M.D. Gearhart[4], S.L. Lerach[5].* 1) Otorhinolaryngology, Mayo Clinic, Rochester, MN; 2) Department of Biology, Hanover College, Hanover, IN; 3) Department of Genetics, Cell Biology and Development, Iowa State University, Ames, IA; 4) Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN; 5) Pediatrics, University of Minnesota, Minneapolis, MN.

Purpose: Autosomal dominant pathogenic variants in *PAX2* cause Renal Coloboma syndrome (OMIM 120330), a condition characterized by optic nerve dysplasia/coloboma and renal hypodysplasia. To date, transcriptional targets of *PAX2* remain largely unidentified. We hypothesize that *PAX2* regulated genes may explain phenotypic overlap between optic nerve findings of Renal Coloboma Syndrome and other conditions, specifically glaucoma. The purpose of this study is to identify genes regulated by *Pax2/pax2a* during mouse and zebrafish eye development and to generate alleles of zebrafish *pax2a* to evaluate target gene expression. **Methods**: RNAseq was performed from biological replicates of E11.5 wildtype and *Pax2-/-*mouse eye tissue. qPCR confirmed differential target expression in embryonic mice and zebrafish. TALEN generated variants within the paired domain of zebrafish *pax2a* and were used for *in situ* hybridization to evaluate *cyp1b1* expression patterns. Bioinformatic analysis of *Cyp1b1* genomic sequence was performed to identify putative Pax2 binding sites. **Results**: RNAseq identified fifty differentially expressed eye genes in E11.5 *Pax2-/-* mouse eye tissue compared to wildtype; *Cyp1b1* expression was significantly reduced. *Cyp1b1/cyp1b1* differential expression was confirmed by qPCR in both mouse and zebrafish embryos. *cyp1b1* expression by *in situ* hybridization was present in the wildtype zebrafish eye at 30hpf. A restricted expression pattern was observed in *pax2a-/-* hypomorphs with greater restriction in *pax2a-/-* null alleles. Identification of Pax2/pax2a binding sites upstream of the *Cyp1b1/cyp1b1* ATG start site supports possible direct transcriptional regulation by Pax2/pax2a. **Conclusions**: We determined that *Cyp1b1* expression by RNAseq is reduced in mouse and confirmed by qPCR in both *Pax2a/pax2a-/-* mouse and zebrafish. *cyp1b1* expression in *pax2a-/-* zebrafish eyes is restricted in an allele specific manner with the greatest restriction occurring in null alleles compared to hypomorphic and wildtype alleles. It is important to note that in humans, autosomal recessive *CYP1B1* pathogenic variants cause congenital glaucoma as well as chorioretinal colobomas suggesting that phenotypic overlap in human patients with Renal Coloboma Syndrome may have a functional basis. Our results support that Pax2/pax2a regulates *Cyp1b1/cyp1b1* expression during vertebrate eye development and provides mechanistic insight into the optic nerve phenotypes observed in Renal Coloboma Syndrome.

2224W

Identification of *SPECC1L* mutations in patients with non-syndromic cleft lip and palate. *I. Saadi[1], E.G. Hall[1], N.R. Wilson[1], S. Undurty[2], J. Standley[2], E.A. Augustine-Akpan[3], Y. Kousa[4], D.S. Acevedo[1], L. Pitstick[5], N. Natsume[6], S. Yousaf[1], T. Busch[2], E. Kosa[1], C. Nguyen[7], M. Nguyen[7], W.L. Adeyemo[8], M.A. Eshete[9], A. Czirok[1], B.C. Bjork[5], B.C. Schutte[4], S. Suzuki[2,6], J.C. Murray[2], A. Butali[3].* 1) Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS; 2) Department of Pediatrics, Craniofacial Anomalies Research Center, University of Iowa, Iowa City, IA; 3) Department of Oral Pathology, Radiology and Medicine/Dow Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA; 4) Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI; 5) Department of Biochemistry, Midwestern University, Downers Grove, IL; 6) Division of Research and Treatment for Oral and Maxillofacial Congenital Anomalies, School of Dentistry, Aichi-Gakuin University, Nagoya, Japan; 7) Odonto Maxillo Facial Hospital, 263-265 Tran Hung Dao Street, District 1, Ho Chi Minh City, Vietnam; 8) Department of Oral and Maxillofacial Surgery, College of Medicine, University of Lagos, Lagos, P.M.B. 12003, Nigeria; 9) Department of Plastic and Reconstructive Surgery, Addis Ababa University, Addis Ababa, P.O. Box: 26493, Ethiopia.

SPECC1L mutations have been identified in patients with rare atypical clefts, and in patients with syndromic cleft lip and palate. These mutations cluster in the second coiled coil and calponin homology domains of SPECC1L, and severely affect the ability of SPECC1L to associate with acetylated microtubules. We wanted to determine if *SPECC1L* mutations also contributed to non-syndromic orofacial clefting. We sequenced 62 Caucasian, 89 Filipino, 90 Ethiopian, 90 Nigerian and 95 Japanese patients with nsCL/P. We identified four rare coding variants (A86T, M91I, T299A, R546Q). Interestingly, all of these variants reside outside of SPECC1L coiled coil domains, and appear to mildly affect the ability of SPECC1L to associate with acetylated microtubules, which is evident upon quantitative fluorescent activated cell-sorting analysis following transfection into mammalian cells of GFP-fusion constructs carrying these variants. Thus, these *SPECC1L* variants associated with nsCLP result in milder functional defects than variants associated with atypical or syndromic clefting. We previously showed that severe deficiency of *Specc1l* in a mouse model results in early embryonic lethality. To validate that moderate *Specc1l* deficiency in mouse affects palatogenesis, we generated a hypomorphic allele, *Specc1l[del200]*, which shows perinatal lethality. Indeed, *Specc1l[del200/del200]* homozygous mutants show rare occurrence of cleft palate. Consistently, *Specc1l[severe/del200]* compound mutants show higher penetrance of cleft palate phenotype. Live-imaging analyses indicate that primary mouse embryonic palatal mesenchyme cells from *Specc1l[severe/del200]* mutants are impaired in collective cell migration. We further show that SPECC1L is broadly expressed in the palate epithelium and mesenchyme, and that this SPECC1L expression is drastically reduced in *Irf6* mutant palates. Together, our data indicate that palatogenesis is specifically sensitive to deficiency of SPECC1L dosage and function in both mouse and human. *SPECC1L* represents a novel gene that functions downstream of *IRF6* in the etiology of nsCL/P.

2225T

Loss-of-function mutations in the X linked gene *STAG2* cause a novel cohesinopathy. *M.W. Wessels[1], W.K. Chung[2], J. Demmers[3], I.M. Frohn-Mulder[4], R. Willemsen[1], J. Saris[1], P.J. Willems[5], S. Sharaeian[6], K.S. Wendt[6].* 1) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) Departments of Pediatrics and Medicine, Columbia University Medical Center, New York, New York, 10032; 3) Proteomics Center, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) Department of Pediatric Cardiology, Erasmus Medical Center, Rotterdam, The Netherlands; 5) GENDIA, GENetic DIAgnostic Network, Antwerp, Belgium; 6) Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands.

Background Defects of the cohesin complex or cohesin regulators, commonly referred to as "Cohesinopathies", are associated with developmental syndromes. The most prominent cohesinopathy is Cornelia de Lange Syndrome (CdLS). Cohesin is important for genome stability, sister chromatid cohesion, DNA damage repair, chromatin looping as well as overall transcriptional regulation, and the complex consists of the core subunits SMC1A, SMC3, RAD21 and the orthologs STAG1 or STAG2. The *STAG2* gene localizes on the X chromosome and undergoes X inactivation. So far, only patients with duplications of the *STAG2* gene have been reported. **Methods and Results** Here we describe two female patients, one with a *de novo* heterozygous 160 kb deletion in Xq25, encompassing the 5' and a large part of the *STAG2* gene, and one with a *de novo* truncating *STAG2* mutation that presumably leads to loss of function. Clinical features in these patients include intellectual disability, microcephaly, mild dysmorphic features, short stature, scoliosis, hypertrophic cardiomyopathy, coarctation of the aorta and congenital diaphragmatic hernia. The dysmorphic features were not suggestive of Cornelia de Lange syndrome. X inactivation studies showed different patterns of X inactivation in different tissues, with skewed X inactivation in blood and reversal skewed X inactivation in fibroblasts in one of the two girls. Sister-chromatid cohesion and cell susceptibility to DNA damage and radiation was normal. Patient-derived cells without expression of STAG2 and cell-line based model systems were used to study the effect of the mutation for the different cohesin functions. **Conclusion** Heterozygous inactivating mutations of *STAG2* lead to a new human cohesinopathy and further confirms a role for the involvement of the Scc3/STAG subunits of cohesin in human disease. .

2226F

Identification of novel candidate genes for Disorders/Differences of Sex Development (DSD) by investigation of variants of uncertain significance identified in 46,XY DSD using the C57BL/6J-Y[POS] mouse model.
H. Barseghyan[1], M. Zadikyan[1], S. Kamsut[1], M. Almalvez[1], M. Bramble[1], M. Larson[1], R. Baxter[1], V. Arboleda[1], A. Eskin[1], S. Nelson[1], E. Delot[1,2], E. Vilain[1,2]. 1) Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California; 2) Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, California.

Disorders of Sex Development (DSD) are defined as "congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical." These conditions have an approximate frequency of 0.5-1% of live births and encompass a wide variety of urogenital abnormalities ranging from mild hypospadias to sex reversal. We have performed exome sequencing (ES) to identify the underlying genetic cause in patients diagnosed with 46,XY DSD. In the majority of cases a specific genetic diagnosis was not found, nevertheless ES yielded large numbers of variants of uncertain significance (VUS). To investigate the relevance of these VUS in regards to the patient's phenotype, we utilized a powerful mouse model for studying undervirilization in 46,XY individuals, in which the presence of a Y chromosome originating from a *M. domesticus poschiavinus* strain (Y[POS]) on C57BL/6J (B6) background results in XY undervirilization and sex reversal. We hypothesized that abnormal gonadal expression of specific genes in B6-Y[POS] males during gonadal development would correlate with VUS in genes of 46,XY DSD patients identified through ES. We isolated gonadal tissue from wild type (WT) B6 and undervirilized B6-Y[POS] males at embryonic day 11.5 (21ts) a time point when the surge of *Sry* gene was compete and performed RNA sequencing in order to assess differential gene expression. We identified 310 genes that were differentially expressed in B6-Y[POS] gonads compared to WT gonads with a fold change (FC) greater than 1.5. A missense variant with a minor allele frequency of less than 1% was present in 187 differentially expressed genes from mouse model and were classified as VUS in our unexplained 46,XY DSD cases. We performed further filtering of these variants by restricting the MAF to 0.1%, increasing expression difference to 2 fold and relying on *in silico* tools for prediction of damaging variants. Using these filtering criteria we identified 19 novel candidate genes involved in 46,XY DSD pathogenesis. Among these genes, several were chosen for further follow up validation using *in vitro* tools such as cell lines and assays. This method allowed identification of novel candidate genes, mutations in which were associated with 46,XY DSD.

2227W

Development of antisense oligonucleotide therapy for Fibrodysplasia Ossificans Progressiva. *R. Maruyama[1], J. Elliott[1], Y. Echigoya[1], T. Yokota[1,2], O. Caluseriu[1].* 1) Department of Medical Genetics, University of Alberta, Edmonton, AB, Canada; 2) Muscular Dystrophy Canada Research Chair, University of Alberta, Edmonton, AB, Canada.

Fibrodysplasia Ossificans Progressiva (FOP) is a rare autosomal-dominant (AD) disorder characterized by progressive heterotopic ossification. The worldwide prevalence is approximately 1 case in 2 million individuals. The median lifespan of the patients is approximately 40 years of age. Currently, there is no effective treatment available. More than 95% of cases are caused by a recurrent mutation of (617G>A; R206H) in *ACVR1*, a bone morphogenetic protein (BMP) type I receptor. The mutation renders ACVR1 responsive to Activin A, by which WT ACVR1 is not activated. The ectopic activation of ACVR1[R206H] by Activin A induces heterotopic ossification. Because ACVR1[R206H] is a hyperactive receptor, a promising therapeutic strategy is to decrease the activity of ACVR1 in patients. To accomplish this goal, we developed new therapies utilizing antisense Locked Nucleic Acid Gapmers (LNA Gapmers). LNA Gapmers are DNA oligonucleotides with fully modified phosphorothioated backbones and have locked nucleic acid (LNA) modification at both ends; they hybridize to a target sequence of mRNA, which induces the mRNA degradation by RNase H. We designed and screened LNA Gapmers specific to *ACVR1* to identify the ones that efficiently reduce the expression of *ACVR1* in FOP patient fibroblasts. Several of our custom designed LNA Gapmers successfully reduced the expression of ACVR1 both in RNA and protein level in the FOP patient fibroblasts. The *in vivo* efficacies of the LNA Gapmers will be tested by FOP mouse and xenograft models. Successful completion of this study will provide a proof of principle for applying antisense oligonucleotide therapy for FOP, and possibly for the treatment of other AD disorders with toxic gain of function mutations.

2228T

Aberrant splicing induced by the most common EPG5 mutation in an individual with Vici syndrome. *M.S. Kane[1], T. Vilboux[1], L.A. Wolfe[2,3], P.R. Lee[2], Y. Wang[1], K.C. Huddleston[1], J.G. Vockley[1,4], J.E. Niederhuber[1,5], B.D. Solomon[1,4,6].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Department of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, VA; 5) Adjunct Professor Oncology, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Inova Children's Hospital, Inova Health System, Falls Church, VA.

Vici syndrome is a severe autosomal recessive condition recently found to be caused by mutations in the *EPG5* gene (NG_042838.1) (Cullup et al, 2013.). The most common disease causing variant identified in Vici syndrome patients is *EPG5* NM_020964.2: c.1007A>G p.Gln336Arg. It has been suggested that this variant may affect splicing via *in silico* prediction models. However, multiple prediction tools yield discordant results regarding the impact on both mRNA splicing and pathogenicity of the amino acid change caused by this variant. Therefore, we have sought to characterize *EPG5* mRNA (expression level and splice isoforms) in fibroblasts from an individual with prototypical features of Vici syndrome who is homozygous for the *EPG5* c.1007A>G variant. We find that *EPG5* transcripts are reduced by 50% in the affected cells compared to control cells. By targeted amplification, cloning and sequencing of the residual transcripts, we observe both normal and abnormal splice products in the context of this Vici syndrome variant. While most of the abnormal splice isoforms encode a premature stop codon due to inclusion of *EPG5* IVS2, approximately 25% of the residual mRNA encodes the missense mutation. These results show the impact of this variant on the mRNA and may reveal that Vici syndrome may only be due to potential null allele variants in *EPG5*. Our results suggest the importance of performance of mRNA analyses in other patients carrying other rare missense variants associated with this disorder. Utilization of RNA sequencing for future analyses will yield a broader picture of all possible abnormal splice isoforms occurring in the context of *EPG5* deleterious variants.

2229F

Exome sequencing in patients with Acrocallosal syndrome like features and functional studies suggest disturbed ciliar Sonic Hedgehog signalling caused by mutations in either *KIF7*, *C5orf42* or *SHH* as underlying cause. *A. Rauch[1], R. Asadollahi[1], J.E. Strauss[2], M. Zenker[3], O. Beuing[4], S. Edvardson[5], O. Elpeleg[6], T.M. Strom[7], P. Joset[1], D. Niedrist[1], C. Otte[1], B. Oneda[1], P. Boonsawat[1], S. Azzarello-Burri[1], D. Bartholdi[1], M. Papik[1], M. Zweier[1], C. Haas[8], A.B. Ekici[9], A. Baumer[1], E. Boltshauser[10], K. Steindl[1], M. Nothnagel[11], A. Schinzel[1], E. Stoeckli[2].* 1) Inst Medizinische Genetik, Univ Zurich, Schlieren-Zurich, Switzerland; 2) Institute of Molecular Life Sciences and Neuroscience Center Zurich, University of Zurich, Zurich, Switzerland; 3) Institute of Human Genetics, University Hospital Magdeburg, Magdeburg, Germany; 4) Institute of Neuroradiology, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany; 5) Pediatric Neurology Unit, Hadassah University Hospital, Mount Scopus, Jerusalem, Israel; 6) Monique and Jacques Roboh Department of Genetic Research, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 7) Institute of Human Genetics, Technische Universität München, Munich, Germany; 8) Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland; 9) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 10) Department of Pediatric Neurology, University Children's Hospital, Zurich, Switzerland; 11) Cologne Center for Genomics, Department of Statistical Genetics and Bioinformatics, University of Cologne, Cologne, Germany.

Acrocallosal syndrome (ACLS) is an autosomal recessive neurodevelopmental ciliopathy which is known to be caused by mutations in *KIF7*. It is recognizable by macrocephaly and facial dysmorphism amongst other overlapping ciliopathies summarized as Joubert syndrome related disorders, for which 30 genes have been identified so far. In order to evaluate the specificity and aetiology of ACLS-like features, we performed genetic and functional analyses in patients with consistent craniofacial appearance and variable other features out of the Joubert syndrome related disorders spectrum. We analysed 13 patients from 9 families including the original two patients by whole-exome or targeted Sanger-sequencing of *KIF7*. Surprisingly, while 7 patients including the original ACLS cases harboured biallelic *KIF7* mutations, 4 patients showed biallelic mutations in the known Jourbert syndrome gene *C5orf42,* and one patient showed a novel de novo truncating mutation in *SHH,* which was known to cause holoprosencephaly, only. While all our patients shared the common minimal diagnostic criteria of cognitive deficits, ACLS-like craniofacial features and abnormal corpus callosum or MTS or vermis hypoplasia, *KIF7* mutated patients had more severe facial dysmorphism and in general had moderate to severe intellectual disability. Long-term follow-up of the patients allowed us to delineate that patients with *C5orf42* mutation usually caught up and remained only mildly to moderately retarded after initial severe developmental delay. However, one patient with homozygosity for a *KIF7* splice site mutation at position -3 also caught up and remained only mildly retarded, which may indicate some leakiness of this mutation. We also estimated the most recent common ancestors for two Swiss *KIF7* founder mutations to 1000-4000 years ago. While KIF7 and SHH were known proteins in ciliary Sonic Hedgehog (Shh) signaling, we showed for the first time abnormal primary cilia in fibroblasts of *C5orf42*-mutated patients. Additionally, *C5orf42* knock-down in chicken embryos revealed axonal pathfinding errors similar to those observed after perturbation of Shh signalling, abnormal neural circuit formation and aberrant craniofacial features consistent with facial widening in the patients. Our findings therefore suggest that ACLS-like craniofacial features are a more general sign of disturbed Shh signaling and can be caused by mutations in *KIF7*, *C5orf42* or *SHH*.

2230W

***IRF6* missense mutations decrease the transactivation ability without affecting the homodimer forming ability.** *K. Imaki, T. Sae, H. Soushi, A. Itsuro, T. Inamine, K. Tsukamoto, K. Shinji.* Dpt of Pharmacotherapeutics Nagasaki Univ Grad Sch, Nagasaki, Japan.

I. Kouta, T. Sae, H. soushi, A. Itsuro, T. Inamine, K. Tsukamoto, K. Shinji Dpt of Pharmacotherapeutics, Nagasaki Univ Grad Sch, Nagasaki, Japan [;Background and Objectives]; *Interferon regulatory factor 6 (IRF6)* mutations are involved in the development of Van der Woude syndrome (VWS) and popliteal pterygium syndrome (PPS). We selected seven VWS mutations, such as V321M, G325E, L345P, C347F, F369S, C374F, and K388E, and one PPS mutation S424L. To elucidate the pathogenic mechanism of VWS and PPS, we investigated whether these mutations that cause VWS or PPS affect the transactivation ability and the homodimer forming ability of *IRF6* in cell culture assays. [;Method]; Eight variant types of *IRF6* expression vectors we generated were transfected into HEK293T, MCF-7, and MDA-MB-231 cells and we compared the transactivation activity by means of a dual-luciferase assay. In addition, after a plasmid expressing AcGFP-mutant IRF6 fusion proteins and a plasmid expressing ProLabel-wild type IRF6 fusion protein were co-transfected into HEK293T cells, we performed a co-immunoprecipitation assay with anti-GFP antibody and compared the homodimer forming ability by measuring the ProLabel activity. Statistical analyses were performed with Dunnett's test after the ANOVA using the SPSS software package. A P value of less than 0.05 was considered to be statistically significant. [;Results]; The transactivation activities of V321M, G325E, L345P, C347F, C374W, K388E, and S424L were significantly reduced, compared with those of the wild-type. In contrast, the ProLabel activities showed no significant differences in the homodimer forming ability between the wild-type and the mutants. In addition, the stimulation with WNT-5A significantly increased the ProLabel activity in wild-type IRF6. [;Discussion □Conclusion]; In this study, we demonstrated that missense mutations of *IRF6* found among the patients with either VWS or PPS reduce the transactivation level statistically without affecting the IRF6 homodimer formation in cell culture assays. Moreover, we suggest that WNT-5A is involved in the IRF6 homodimer formation. Missense mutations may inhibit IRF6 from either translocating into the nucleus or binding to the elements of target genes. We propose that the decrease of transactivation and the change of the expression levels in the target genes by *IRF6* mutations may cause VWS and PPS, through the IRF6-mediated signaling pathway.

2231T

Altered ribosomes in the recessive ribosomopathy, Shwachman-Diamond syndrome. *M.E. Tourlakis[1], H. Liu[2,3], R. Gandhi[2], P. Hu[4], A.D. Paterson[2,5], J.M. Rommens[2,3].* 1) Life Sciences Division, Quest University Canada, Squamish, British Columbia, Canada; 2) Program in Genetics & Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 4) Department of Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; 5) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada.

Shwachman-Diamond syndrome (SDS) is an autosomal recessive disease characterized by growth retardation, exocrine pancreatic dysfunction, skeletal dysplasia, cognitive impairment and bone marrow failure. SDS is caused by mutations in *SBDS*. Recent studies indicate that Sbds functions together with Efl1to release Eif6 from the pre-60S complex, enabling ribosomal subunit joining for translation initiation. To assess the protein synthesis deficiencies that have been detected in culture and *in vivo* models of SDS, ribosomal profiles of SDS murine fetal organs were examined and found to lead to reduced 80S monosome peaks and polysomes with increased number and amplitude of sub-peaks. Ribosomal subunit level ratios (60S/40S) were not altered. To understand these observations and to learn how the SDS translatome is altered, total and polysomal mRNAs of mutant and control fetal liver samples were studied using cDNA microarray analyses. By comparing individual polysomal-bound probe level to respective steady state probe level, 1,089 probe sets (of 25,914 analyzed) exhibited altered polysome loading in mutant fetal livers, with 817 being increased (fold change ≥ 2; FDR adjusted P-value ≤ 0.05). These groups corresponded to a total of 799 RefSeq genes, but did not reveal specific biological pathways or point to alerted stress response pathways detected in other organs exhibiting SDS phenotypes. Further, the changes in polysome loading did not correspond to changes in steady state protein levels, as indicated by label-free mass spectrometry or immunoblotting. Rather, these changes reflected physical features of the transcripts, including 5' untranslated region composition as well as the nucleotide composition of the open reading frame and the 3' untranslated regions. Increased binding was consistent with low GC content and longer untranslated regions, features suggestive of bound ribosomes that are not functioning. In additional investigations, Eif6 was found to be present in profile fractions containing either the 60S subunit and 80S monoribosome complexes in mutants, in contrast to only 60S subunits in controls. Together, we conclude that the untimely release of Eif6 due to Sbds deficiency results in ribosomes with compromised translation and that SDS phenotypes reflect protein synthesis insufficiency due to the loss of translation initiation and the formation of sub-populations of 'SDS ribosomes' with non- or poor- translating capability.

2232F

OFD1 cooperates with Bicc1 to regulate protein synthesis at the centrosome. *B. Franco[1,6], D. Iaconis[1], M. Monti[2], M. Renda[1], A. van Koppen[3], M. Chiaravalli[4], F. Cozzolino[2], P. Pignata[1], C. Crina[1], P. Pucci[2], A. Boletta[4], V. Belcastro[1], R.H. Giles[3], E.M. Surace[1], M. Pende[5], R. Tammaro[1].* 1) 1.Telethon Institute of Genetics and Medicine (TIGEM), Via Campi Flegrei, 34, 80078, Pozzuoli, Naples, Italy; 2) 2.Dipartimento di Scienze Chimiche and CEINGE Biotecnologie Avanzate, Università di Napoli Federico II, Via Gaetano Salvatore 482, 80145 Napoli, Italy; 3) 3.Department of Nephrology and Hypertension, University Medical Center Utrecht, Heidelberglaan 100, 3584CX Utrecht, the Netherlands; 4) 4.Division of Genetics and Cell Biology, Dibit, San Raffaele Scientific Institute, Via Olgettina, 58 – 20132, Milan, Italy; 5) 5.Institut National de la Santé et de la Recherche Médicale (INSERM) U1151, Institut Necker Enfants Malades, Université Paris Descartes, Sorbonne Paris Cité, Paris, France; 6) 6.Department of Translational Medicine, University of Naples "Federico II", Via Sergio Pansini, 80131, Naples, Italy.

Protein synthesis is traditionally associated with specific cytoplasmic compartments. We now show that OFD1, a centrosomal/basal body protein, interacts with components of the Preinitiation complex of translation (PIC) and of the Eukaryotic Initiation Factor (eIF)4F complex and modulates the translation of specific mRNA targets in the kidney. We demonstrate that OFD1 cooperates with the mRNA binding protein Bicc1 to functionally control the protein synthesis machinery at the centrosome where also the PIC and eIF4F components were shown to localize in mammalian cells. Interestingly, Ofd1 and Bicc1 are both involved in renal cystogenesis and selected targets were shown to accumulate in two models of inherited renal cystic disease (i.e. a murine model of Oral-facial-digital (OFD) type I syndrome and a murine model of autosomal dominant polycistic kidney disease (PKD1). Our results highlight a possible role for the centrosome as a specialized station to locally modulate translation for specific functions of the nearby ciliary structures and may provide functional clues for the understanding of renal cystic disease.

2233W

Mutations in *DHH*, *WWOX* and *FLNB* cause 46,XY DSD through disruption of the MAP kinase pathway. *J. Loke, M. Syeda, K. Upadhyay, A. Pearlman, H. Ostrer.* Pathology, Albert Einstein Sch Med, Bronx, NY.

Introduction. Disorders of sex development (DSDs) affect up to 1:1000 individuals and may account for 7.5% of all birth defects. 46,XY DSD may occur when testes determination has failed or is incomplete. Previously, we showed that mutations in *MAP3K1* are a common, if not the most common, cause of 46,XY DSD (Loke et al. Hum Mol Genet 23:1073-83, 2014). These mutations decrease SOX9 expression and increase β-catenin activity, tilting the balance away from testis determination in the interacting male and female sex-determining pathways. Here, these genetic pathways have been extended to individuals with 46,XY DSD who have mutations in other genes. **Materials and methods.** Mutations in *DHH, WWOX* and *FLNB* were observed in individuals with 46,XY DSD. Notably, the individual with the biallelic mutations in *FLNB* also had Larsen syndrome, as did her father. These mutations were recreated in NT2/D1 cells, then assessed for expression of SRY, SOX9, FOXL2, WNT4, phospho p38, phospho ERK1/2 by qPCR and/or digital cell Western. The activity of β-catenin was assessed by the TOPFLASH luciferase assay. **Results.** The mutations in *DHH, WWOX* and *FLNB* caused deceased expression of SRY and SOX9, increased expression of FOXL2 and WNT4 and increased phosphorylation of p38 and ERK1/2, thus accounting for the phenotypes. **Conclusion.** Mutations in genes for signaling molecules (DHH), structural molecules (FLNB) and transcription factors (WWOX) can all disrupt sex determination by modifying the activities of genes in the MAP kinase-transcription factor signal transduction pathway. The NT2/D1 functional assay system can be adapted readily to understanding the effects of mutation in any candidate sex determining gene and to screen for new sex determining genes. Overlap in developmental pathways between Sertoli cells and chrondrocytes can cause co-occurring DSDs and skeletal dysplasias, as has been observed for campomelic dysplasia and now for Larsen syndrome.

2234T

SOX11 is necessary for development of the bony labyrinth. *E.D. Sperry[1,3], R. Hojjati[2], D.L. Swiderski[4], G. Wan[4], S. Brewer[4], V. Popov[4], W.M. King[4], G. Corfas[4], Y. Raphael[4], D.M. Martin[1,3,5].* 1) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 2) College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Department of Otolaryngology - Head and Neck Surgery, University of Michigan, Ann Arbor, MI; 5) Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI.

The transcription factor SOX11 is of increasing interest as a mediator of inner ear development. Heterozygous mutations in *SOX11* cause Coffin-Siris (CS) syndrome, a heterogeneous disorder resulting in craniofacial anomalies and sensory impairments, including sensorineural hearing loss. Interestingly, *Sox11* expression is regulated by the chromatin remodeler CHD7, altered in CHARGE syndrome, which presents with complex inner ear malformations. A recent report suggests that SOX11 is necessary for proper development of utricular hair cells. Here, we sought to further explore roles for SOX11 in common genetic pathways with CHD7 during inner ear development. By paintfill analysis, we observed lateral semicircular canal abnormalities in one-third of *Sox11*[+/-] (*N*=10/26) and all *Sox11*[-/-] (*N*=12/12) E14.5 ears. Examination of *Sox11*[-/-] E11.5 and E12.5 embryos revealed a failure of lateral semicircular canal fusion plate formation. *Sox11*[+/-] mice exhibited mild high-frequency hearing loss by auditory brainstem response (ABR) testing and a significant reduction in P1 latencies by vestibular sensory evoked potential (VsEP) testing. Quantitative RT-PCR, *in situ* hybridization, and immunofluorescence on E10.5 and E12.5 *Sox11* and *Chd7* mutant ears showed spatiotemporal misregulation of several genes critical for ear morphogenesis, including *Bmp4*, *Otx1*, and *Shh*. Further, dosage-dependent reductions in *Sox11* mRNA were detected at E10.5 and E12.5 in *Chd7*[Gt/+] and *Chd7*[Gt/Gt] mutant ears relative to *Chd7*[+/+] controls. SOX11 immunofluorescence was reduced in *Chd7*[Gt/+] and absent in *Chd7*[Gt/Gt] inner ears. *Chd7;Sox11* double heterozygous mutant inner ears exhibited the same morphologic changes in semicircular canal structure as *Chd7*[Gt/+] mice. Taken together, our studies demonstrate that loss of *Sox11* in mice results in lateral and posterior semicircular canal dysplasia, concurrent with spatiotemporal misregulation of genes critical for development of the bony labyrinth. Further, we observe a possible role for SOX11 in adult mouse auditory and vestibular function. Finally, *Sox11* may mediate some of the effects of CHD7 during mouse inner ear development. These studies help clarify molecular relationships between *Sox11* and *Chd7* in inner ear development and provide insights about the pathogenic mechanisms involved in CS and CHARGE syndromes.

2235F

Fetal neuronal migration defects and severe cardiac abnormalities associated with a *de novo* **pathogenic mutation in the** *PIK3R2* **gene.** *G. Nimmo[1,2], A. Staines[2], P. Shannon[3], S. Keating[3], G. Mirzaa[4], D. Chitayat[1,2].* 1) Department of Paediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, The University of Toronto, Toronto, Ontario, Canada; 2) The Prenatal Diagnosis and Medical Genetics Program, Dept. of Obstetrics and Gynaecology, Mount Sinai Hospital, Toronto, Ontario, Canada; 3) The Prenatal Diagnosis and Medical Genetics Program, Dept. of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Department of Pediatrics and Center for Integrative Brain Research , Seattle Children's Research Institute, The University of Washington, Seattle, WA, USA.

The Megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome is known to be associated with mutations in three PI3K-AKT-MTOR pathway genes: *PIK3R2, AKT3* and *CCND2*. So far all reported human cases were in liveborn individuals and the fetal findings have not well been delineated. Further, none of the mutation-positive MPPH individuals had non-CNS abnormalities, with the exception of postaxial polydactyly. Here, we report the first fetus with a *de novo* mutation in the most commonly seen MPPH gene, *PIK3R2*, who had severe CNS and cardiac defects, further expanding the clinical spectrum of *PIK3R2*-related disorders. The mother was a 32 year old G1P0 woman of Anglo-British descent and the father was 38 years old and of Anglo-Saxon descent. Both parents were healthy and non-consanguineous. Initial prenatal ultrasound done at 13 weeks showed NT of 1.6 mm and a detailed fetal ultrasound at 20 weeks gestation showed hypoplastic left heart with coarctation of the aorta. The couple was counseled and decided to interrupt the pregnancy. A fetal autopsy confirmed the cardiac abnormalities and also showed cerebral abnormalities including aberrant cortical lamination predominantly involving the temporal lobes, as well as numerous leptomeningeal and subcortical heterotopias. Whole-exome sequencing revealed a previously unreported *de novo* apparently constitutional mutation in the *PIK3R2* gene. Although only a handful of variants have been reported in patients with MPPH, all but one describe variants affecting the SH2 domain of PIK3R2 (p.G373R). Herein, we documented a non-synonymous amino acid substitution affecting a highly conserved residue outside the SH2 domain and in a region proposed to interact with the catalytic subunit of PI3K and expand the role of this gene to include cardiogenesis.

2236W

Gene discovery in Mabry syndrome (OMIM 239300): Disruption of the gly-cophosphatydylinositol (GPI) anchor biosynthesis pathway associated with a metabolic disorder causing profound mental retardation, seizures and other neurologic deficits. *M. Thompson[1], C.C. Mabry[2].* 1) Laboratory Medicine and Pathobiology, Banting Inst, Univ Toronto, Toronto, ON, Canada; 2) Pediatrics, University of Kentucky, Lexington, KY, USA.

Mabry syndrome (OMIM 239300), an autosomal recessives syndrome of hyperphosphatasia (elevated circulating alkaline phosphatase (AP), seizures and neurologic deficits) was first described in 1970. Over the ensuing four decades, few cases were reported. In 2010, however, new families were identified and the syndromic nature of the disorder confirmed. This enabled the use of exome sequencing to identify the causative defects in the glycosyl phosphatidylinositol (GPI) biosynthetic pathway. Genetic variants of the phosphatidylinositol glycan (PIG) biosynthesis gene products (active in the endoplasmic reticulum) and the Post-GPI Attachment to Proteins (PGAP) gene products (active in the Golgi) have been identified. We continue our work of gene discovery in Mabry syndrome using exome sequencing of original index cases. This work will identify novel genes and/or mutations associated with the form of this autosomal recessive disorder as it was first described in 1970. Since these patients exhibit lysosomal storage, we will have the opportunity to identify the role this material may play in the etiology of seizures and cognitive defects seen in Mabry syndrome. While the contribution of hyperphosphata-sia to the pathogenesis of the syndrome is not known, since there are more than 250 other proteins dependent on GPI for their anchoring to the plasma membrane, the present study of Mabry syndrome may provide a new window on AP function in growth and development.

2237T

Cross-species comprehensive phenotyping to reveal the genetic basis of disease: Informatics for the International Mouse Phenotyping Consor-tium. *T. Meehan[1], A. Mallon[2], H. Parkinson[1], D. Smedley[3], MPI2 Consortium.* 1) EMBL-EBI, Hinxton, Cambridge, United Kingdom; 2) MRC Harwell, Oxford, UK; 3) Genomics England, Queen Mary University of London, London, UK.

Correlating phenotypic aspects of disease with causal genetic variants are often confounded by a lack of functional gene knowledge. The Interna-tional Mouse Phenotyping Consortium (IMPC) is addressing this challenge by building a functional catalog of a mammalian genome by producing and characterizing a knockout mouse strain for every protein-coding gene. This G7 recognized global research infrastructure carefully coordinates activities across a dozen international partners to generate strains for the biomedical re-search community. Data from a standardized, broad-based phenotyping pipe-line is collected, quality controlled, archived and analyzed by the IMPC-Data Coordinating Center. Annotation with biomedical ontologies allows biologists and clinicians to easily find mouse strains with phenotypic traits relevant to their research and facilitates integration with other resources. With phenotype data available for over 3000 knockout strains at mousephenotype.org, new gene-phenotype correlations are providing novel insights into development, fertility, disease and sexual dimorphism. We will present our informatics ap-proach for identifying and characterizing hundreds of new rare disease models and demonstrate how clinical researchers can benefit from IMPC resources. .

2238F

IGF2 and its network partners: Key factors in the etiology of Russell-Silver syndrome. *T. Eggermann, L. Soellner, M. Begemann.* Dept Human Gen, Univ Hosp, Aachen, Germany.

With the identification of 11p15.5 disturbances in Russell-Silver syndrome (RSS/Silver-Russell syndrome) ten years ago, the central role of at least the telomeric imprinted region ICR1 in 11p15.5 for the etiology of the disease became obvious. Two of the genes regulated by the ICR1 are IGF2 and H19. Due to its physiological function and its imprinting status, IGF2 had been suggested as the causative gene for RSS, but the final clue was still missing. With the recent identification of a paternally inherited IGF2 mutation in a family with growth retardation and RSS, we could confirm the hypothesis that IGF2 is the key factor for the growth retardation and other features in RSS. In the last decade it has emerged that IGF2 and other imprinted genes belong to common functional networks, either because of their physiological function or because they belong to the imprinted genes network (IGN), and molecular alterations of genes of IGF2 interaction partners also contribute to growth retardation. Comprehensive studies in our patient cohort (n>500) referred for RSS diagnostic testing helped us to identify further molecular alterations in the 11p15.5 region itself, but also in other genes and chromosomes (e.g. GRB10 in 7p13, MEST in 7q31, IGF1R in 15q26). Our data and those from animal and cellular models indicate that IGF2 and the coregulated H19 gene as well as further imprinted and/or functionally related factors belong to the IGN and the IGFII-IGF1R signaling cascade by cis and trans acting mechanisms. These findings contribute to a better understanding of human growth and its disturbances. The translational use of this knowledge has significantly improved molecular diagnostics of RSS and will have an impact on personalized patient management.

2239W

Identification of the first causative missense mutation of the *GPSM2*-gene in a patient with Chudley-McCullough syndrome (CMS). *P.M. Kroisel[1], G. Schaller[2], T. Kau[3], R. Birnbacher[4], M.R. Speicher[1], C. Windpassinger[1], K. Wagner[1], T. Schwarzbraun[1].* 1) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 2) Pediatrician, Villach, Austria; 3) Department of Pediatrics, Pediatric Radiology, Hospital Klagenfurt, Klagenfurt, Austria; 4) Department of Pediatrics, Regional Hospital Villach, Villach, Austria.

In the second child of a consanguineous couple of Turkish origin typical symptoms of Chudley-McCullough syndrome (CMS [MIM 604213]) are present. Both parents are healthy first degree cousins. The patient is now 4 years old and shows characteristic MRI findings of his brain including dysgenesis of corpus callosum and asymmetry of ventricles and complex microgyria although brain anomalies in CMS can show a considerable variability. Congenital deafness and a developmental delay of his expressive speech were also diagnosed. Initially he did also show a delay of his motor development but this is not present anymore. By autozygosity mapping a homozygosity of several larger genomic segments also including the *GPSM2*-gene was confirmed and subsequent TruSight One gene panel analysis revealed the homozygous missense mutation g.128G>A, p.Gly43Asp of the *GPSM2*-gene. Interestingly up to now according to the literature and HGMD professional just 10 causative mutations of the *GPSM2*-gene in CMS and non syndromic hearing loss have been reported and all of these mutations are loss of function mutations. These are either homozygous or compound heterozygous nonsense mutations, one splicing mutation or two one base pair deletions. Our CMS patient is therefore the first reported with a causative missense mutation of the *GPSM2*-gene. Analysis of the identified missense mutation on function of the G protein signaling modulator 2 could provide a better understanding of the relevant domains of this gene product. It will be interesting to follow up the further development of our patient and compare it with other described cases to learn more on the impact of this new mutation of the *GPSM2*-gene on prognosis with detailed characterization of the genotype-phenotype relationship in CMS patients.

2240T

In vitro modelling Richieri-Costa-Pereira syndrome with iPSC-derived neural crest cells. *C.M. Musso[1], G.S. Kobayashi[1], F.A.A Ishiy[1], L.C.J. Caires[1], E. Goulart[1], L. Alvizi[1], A.G. Morales[1], K. Grisei-Oliveira[2], M.R. Passos-Bueno[1].* 1) Institute of Biosciences, University of São Paulo, São Paulo, Brazil; 2) Hospital Israelita Albert Einstein, São Paulo, Brazil.

Richieri-Costa-Pereira syndrome (RCPS) is a rare acrofacial dysostosis likely caused by disturbances in cranial neural crest development. RCPS is characterised by Robin sequence, cleft mandible and limb defects and is caused by non-coding expansions in the 5'UTR of *EIF4A3,* which are involved in many aspects of RNA metabolism (mRNA splicing and nonsense-mediated decay); nevertheless, the pathogenetic mechanism needs to be elucidated. The 5'UTR of the EIF4A3 is characterized by multiple allelic patterns, varying in size and organization of repeats, and RCPS patients present an increased number of motifs in this region. Here, we have demonstrated by Luciferase assay, *in vitro,* that the increased alleles are directly associated to a decreased gene expression; however, the impact of the *EIF4A3* deficiency in human development remains unknown. These results indicated a regulatory role for this region, which may be due a differences in the transcription factor binding sites between control and mutant alleles, evidenced by *in silico* analysis. Further, in order to investigate the eatiology of this syndrome, we have successfully established iPSC-derived neural crest cells (iNCCs) and iNCC-derived mesenchymal cells (nMSCs) from RCPS subjects and controls. All the cells were characterised by RT-qPCR, flow cytometry and immunofluorescence analysis and we have evaluated the migration potential, apoptosis and mesenchymal differentiation of these cells. We showed that RCPS cells possess increased osteogenic potential in comparison to controls, evidenced by increased expression of *RUNX2*, *ALP* and *BGLAP,* increased alkaline phosphatase activity and mineralized matrix deposition. We have demonstrated that this altered osteogenic differentiation is due to transcriptional dysregulation of osteogenesis-associated genes and such dysregulation was also observed in RCPS knockdown cells, supporting the relationship between reduced *EIF4A3* expression and osteogenic potential. This work indicates a possible pathogenetic mechanism for RCPS and shows the relevance of iPSCs to study craniofacial disorders. CAPES, FAPESP, CNPq.

2241F

Regulation of colonic proinflammatory cytokines and chemokines by the Ankylosing spondylitis-associated, GWAS-identified ERAP1 gene. *Y.A. Aldhamen[1], F.S. Alyaqoub[1], Y. Pepelyayeva[1], D.P Rastall[1], A. Raedy[1], P. O'Connell[1], C. Pereira-Hicks[1], S. Godbehere[1], L.R. McCabe[1,2], A. Amalfitano[3].* 1) Microbiology and Molecular Genetics, Michigan State University , East Lansing , MI; 2) Department of Pediatrics, College of Osteopathic Medicine, Michigan State University, East Lansing, MI 48824; 3) Department of Physiology, Biomedical Imaging Research Center, Michigan State University, East Lansing, Michigan, USA.

Endoplasmic reticulum aminopeptidase 1 (*ERAP1*) gene polymorphisms have been linked to several autoimmune diseases that are characterized by chronic inflammation such as Ankylosing spondylitis (AS), type I diabetes, multiple sclerosis, and Behcet's disease; however, a role for *ERAP1* in colitis has not been described. Notably, gut inflammation was found in 49% of AS patients and AS patients are also known to develop colitis. However, the exact molecular mechanism that predisposes AS patients to intestinal inflammation has not been defined. We previously demonstrated that *Erap1* deficiency alters innate and adaptive immune responses during pathogen recognition. In addition, we also demonstrated that autoimmune disease-linked variants of *ERAP1* activate the NLRP3/caspase-1 inflammasome pathway. Previous reports showed that mice deficient for NLRP3 inflammasome components are highly susceptible to dextran sodium sulfate (DSS)-induced colitis. Here, we hypothesized that *Erap1* may modulate the susceptibility of animals to DSS challenge, due to *Erap1*'s impact on innate immune responses. We show that mice deficient for *Erap1* (*Erap1*-/-) are highly susceptible to DSS-induced colitis, as compared to WT mice. We also identified exaggerated colon proinflammatory cytokine and chemokine responses in the DSS treated *Erap1*-/- mice, as compared to WT mice, resulting in increased mortality rates and severe intestinal inflammation. In addition, decreased production of the protective cytokine, IL-18, was also noted in *Erap1*-/- mice. Assessment of the intestinal inflammation marker lipocalin-2 (Lcn-2), reveled potent increases in Lcn-2 in DSS-treated *Erap1*-/- mice. Moreover, utilization of *Erap1*/β2m double knockout mice suggested that these responses are mediated in part by a mechanism that was independent of antigen presentation functions, that is also mediated by *Erap1*. These results demonstrate that *Erap1* mediates a protective function during colitis, a function that we correlated with its ability to regulate the production of intestinal proinflammatory cytokines and chemokines. These results are highly relevant to humans as *ERAP1* polymorphisms are associated with autoimmune diseases that are also associated with higher risks for development of colitis.

2242W

Pleiotropic effects of a missense variant in *BMP4* in dental phenotypes.
F. Geller[1], X. Liu[1], E.A. Nøhr[2], D.M. Hougaard[3], B. Feenstra[1], M. Melbye[1,4,5].
1) Dept. of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Research Unit of Gynaecology & Obstetrics, Institute of Clinical Research, University of Southern Denmark, Odense; 3) Danish Centre for Neonatal Screening, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, Denmark; 4) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 5) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

Dental arch dimensions and tooth eruption are under genetic influence. Since 1972, the dentist visits of Danish school children are recorded in a central database. Apart from the eruption status of primary/permanent teeth and their caries status, additional traits of malocclusion and crowding are reported. Analyzing genome-wide data of more than 13,000 individuals, we identified a genetic variant, rs17563, associated with mandibular crowding (OR 1.19, $P=4\times10^{-11}$). The risk allele frequency was significantly decreased in individuals with excessive spacing and also in cases with diastema. The variant in bone morphogenetic protein 4 (*BMP4*) has previously been described in the context of primary and permanent tooth eruption, with the risk allele for crowding associated with later eruption, which could be a consequence of the limited space in the jaw. Furthermore, *BMP4* plays an important role in embryonic development and can cause eye, brain, and digit anomalies. Previous studies showed that *BMP4* is expressed early in tooth development and our findings underline a central role of *BMP4* in multiple dental traits.

2243T

Analysis of matrix metalloproteinase expression in pressure ulcer healing.
A. Bhagat[1], M. Dwivedi[1], R. Srivastava[1], A. Chandra[2], S. Raj[1]. 1) Orthopaedic Surgery, King George's Medical University, Lucknow, India; 2) Surgical Gastroenterology, King George's Medical University, Lucknow, India.

Introduction: Matrix metalloproteinases (MMPs) play a major role in wound healing: they can degrade all components of the extracellular matrix. In pressure ulcers there is an excess of MMPs and a decrease of the tissue inhibitors of MMPs (TIMPs). This imbalance is probably one cause of impaired healing. However, little is known about changes in MMPs during wound healing. This study was planned to evaluate the levels of MMPs at different follow-up.**Material & Methods:** 22 subjects with pressure ulcers were enrolled in the study. Wound tissue was collected regularly during the 9 week follow-up period, for measurement of MMP-8, MMP-9 and TIMP-1. Results were analyzed by the degree of wound healing: good healers (defined by a reduction of 65% wound surface area at 5 weeks) and bad healers (reduction of less than 65% in wound surface area at 5 weeks). Levels of MMPs were analyzed by ELISA.**Results:** Levels of MMP-8 and 9 decreased earlier in good healer group. The initial levels of MMP-8 were similar in good and poor healers ($P = 0.1$) but the level increased significantly at week 2 in good healers ($P = 0.03$). This was continued in successive weeks. There was a significant correlation between a high ratio of MMP-8/TIMP-1 and good healing ($r = 0.65$, $P = 0.008$). Pierson correlation analysis showed that an MMP-8/TIMP-1 ratio of 0.39 best predicted wound healing (sensitivity = 71%, specificity = 87.5%).**Conclusion:** A low level of MMP-8 seems essential to wound healing, while an excess of MMP-8 and -9 is deleterious, and could be a target for new topical treatments. The MMP-8/TIMP-1 ratio is a predictor of healing in pressure ulcers.

2244F

An ENU mutagenesis screen using inbred mice identifies mutations in the collagen-modifying enzymes *Colgalt1* **and** *Plod3***, which result in nearly identical phenotypes.** *K.A. Geister, A.E. Timms, S. Ha, D.R. Beier.* Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA.

The discovery of modifier genes in human and mouse has seen slow progress. In mice, this is largely due to the reliance on inherent variability among inbred strains, which makes identifying the causal variant, even in highly resolved mapped loci, quite challenging. We have developed a fast and cost-effective method for genetic mapping of causal ENU-induced mutations using Next Generation Sequencing that combines single nucleotide polymorphism (SNP) discovery, mutation localization, and identification of causal sequence variants. This approach precludes the need for an outcross to facilitate mapping, as we use the ENU-induced variants as SNP markers. This strategy allows us to perform both primary screens and modifier screens on fully inbred lines. We have successfully mapped causal mutations using this approach, and we are currently testing the efficacy of this strategy with regard to modifier discovery using two C57BL/6 congenic strains as our sensitized lines. We anticipate that we may ascertain more complex genetic interactions as well as new ENU-induced mutant phenotypes that model human birth defects. In our first screen we discovered two ENU-induced mutations in genes that encode collagen-modifying enzymes that result in a nearly identical phenotype. One mutation is in *Colgalt1* (collagen (beta 1-O) galactosyltransferase type 1), which is required for proper galactosylation of hydroxylysine residues in a number of collagens. The other mutation is in *Plod3*, which can catalyze the formation of hydroxylysine, and results in a number of disruptions in the vasculature, skin, and other organs when mutated in human patients. *COLGALT1* mutations, in contrast, have yet to be documented in humans. *Colgalt1* mutant embryos exhibit skeletal, skin, and muscular defects, and we hypothesize that the *Colgalt1* mutant could serve as a model of a human connective tissue disorder and/or congenital muscular dystrophy or myopathy. The mutants also show variable expressivity of cleft palate and exencephaly. Examination of the sequencing data appears to exclude the likelihood that these are due to an unlinked ENU-induced variant, illustrating that stochastic effects remain an important source of variation in developmental phenotypes. Funding: NIH R01 HD36404.

2245W

The effect of *VDR* **knockdown on T Cell Apoptosis.** *Z. Shirvani-Farsani[1], M. Behmanesh[2].* 1) Department of Genetics, Faculty of Biological Sciences, Shahid Beheshti University, Tehran, Iran, PhD; 2) Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran, PhD.

BACKGROUND: Vitamin D receptor (VDR) is a member of the nuclear receptor family of transcription factors. VDR play an essential role in immun cells apoptosis. However, the molecular mechanisms underneath this effect remain unclear. Therefore, We evaluated the effect of vitamin D receptor (VDR) knockdown on T cell apoptosis. **METHODS:** Jurkat T cells lines were cultured under controlled conditions. Then we used shRNAs targeting the human VDR and a scrambled non-targeting control. The efficiency of VDR down-regulation in the cells was determined using real-time RT-PCR. Next, apoptosis was assessed using the Flow Cytometery and Annexin V FITC apoptosis detection kit. Statistical analysis of the data was performed using two-way ANOVA test. **RESULTS:** These results demonestrated that VDR knockdown statistically significantly increased apoptosis in Jurkat T up to ~13%. **CONCLUSIONS:** In conclusion, shRNA study suggests that the VDR may be a novel target in apoptosis investigations. Therefore, identifiable markers related to VDR may allow for a use of those in different diseases. **Keywords:** VDR, T cells, Apoptosis, Flowcytometery. .

2246T

Global loss of BBS proteins is associated with increased levels of reactive astrocytes and pro-inflammatory cytokines, IL1b, IL6, IL15 and TNFa in brains of mouse models of Bardet Biedl Syndrome. *M. Singh, V. Sheffield.* University of Iowa, Iowa City, IA.

Bardet Biedel Syndrome (BBS) is a rare human cilioopathic genetic disorder. Principally it is characterized by obesity, retinitis pigmentosa, polydactyly, hypogonadism, and renal failure. Proteins in primary cilia and basal bodies mediate numerous signaling pathways including Notch signaling and inflammation. It has been shown that loss of BBS1 and BBS4 lead to increased notch signaling due to defective Notch receptor trafficking with notch receptor accumulating in late endosomes. The Notch-STAT3 signaling axis is reported to control reactive astrocyte proliferation and the JAK/STAT3 signaling pathway is a common inducer of reactive astrocytes. Increased reactive astrocytes secrete molecules that promote neuro-inflammation leading to neurodegeneration. Reactive astrocytes are associated with patients with brain injury and diverse central nervous tissue disorders. Therefore, understanding the signaling network that regulates reactive astrocytes has the potential to provide therapeutic targets to treat many central nervous disorders. In this study it was hypothesized that loss of BBS proteins in mouse models of BBS syndrome promotes astrocyte reactivity in brains of BBS mice. We examined RNA and protein samples by RT-PCR and western blot analysis and performed confocal immunofluorescence imaging from brains of wild type littermates and BBS null mice for molecular markers of reactive astrocytes including GFAP, Vimentin, and Serpina3n. We find that expression of GFAP, Vimentin, and snerpinA-3n are significantly increased in brains of BBS mice when compared to their wild type littermates. In addition, we also find several pro-inflammatory molecular markers of IL15, IL1b, TNFa, and IL6 are significantly increased without any alteration in microglia markers of reactivity in brains of BBS mice when compared to their wild type littermates. All together these results suggest that BBS proteins plays an important role in regulating astrocyte function and that increased pro-inflammatory molecules indicate the presence of neuro-inflammation in brains of BBS mice. .

2247F

Gene coexpression network analysis reveals thymic sexual dimorphism in human infants. *C.A. Moreira-Filho[1], S.Y. Bando[1], F.B. Bertonha[1], L.R. Ferreira[1], A. Coutinho[2], M. Carneiro-Sampaio[1].* 1) Pediatrics, University of Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Instituto Gulbenkian de Ciência, Oeiras, Portugal.

It is widely known that in human populations females are more susceptible to autoimmune diseases than males. Recently, Dragin et al. [J Clin Invest. 2016 126:1525] shown that estrogen-mediated downregulation of AIRE influences sexual dimorphism in autoimmune diseases. Here we studied whole mRNA and microRNA expression in human thymus explants obtained from infants aged 3-17 months (11 males and 11 females) undergoing corrective cardiovascular surgery. In the male (M) versus female (F) comparison, 775 GO annotated differentially expressed (DE) genes (out of 12,469 valid transcripts) were found to be upregulated and two were downregulated in the M group. The *AIRE* gene has essentially the same expression level in M and F groups. Gene coexpression networks (GCNs) were inferred for M and F groups considering the DE genes. High hierarchy genes were identified for M and F groups revealing noteworthy features: genes related to autoimmune diseases, as *STAT3* and *TYK2*, are high-hubs respectively in F and M networks. Community detection analysis was accomplished and coarse-grained community structure revealed the interconnections between gene communities, showing that F communities are, in general, more interconnected. MicroRNA target analysis indicated a minor role for epigenetic mechanisms in the maintenance of GCNs topological dimorphism. These data suggest that genetic mechanisms and prenatal hormonal influences probably act synergistically in establishing thymic sexual dimorphism.

2248W

Investigating the roles of miRNAs in the differentiation of human induced pluripotent stem cells into kidney. *L. Li, H.H. Cheung, K.K. Miu, D.D. Cao, J.J. Tu, W.N. Law, W.Y. Chan.* School of Biomedical Sciences, Chinese University of Hong Kong, Hong Kong, Hong Kong.

Studying human embryogenesis *in vivo* is still impeded by the impracticality of obtaining embryo samples. Therefore, human induced pluripotent stem cells (hiPSCs), which are capable to different into any somatic cell types, offer an ideal model system to study human embryogenesis *in vitro*. Previous studies suggested that microRNA (miRNAs) are the key regulators of lineage specification. However, whether miRNAs are key regulators in renal formation and associated diseases remains obscure. To investigate the roles of miRNAs in renal differentiation, we recapitulate the early stage of embryonic kidney development (KD) *in vitro* by synchronously inducing both nephron progenitors and ureteric bud from hiPSCs. During the first 10 days of induction, we chose four time-points to perform miRNA microarray analyses. By integrating the array data and bioinformatics analysis, we found one novel downregualted miRNA – *miR-105-5p* whichwas predicted to target *PKD1* and *PKD2*. Normal expression of polycystin-1 and polycystin-2, encoded by *PKD1* and *PKD2*, contribute to, while their dysregulation perturb, the maintenance of normal epithelial structure in developing kidney. The luciferase reporter assay confirmed the interaction between *miR-105-5p* and *PKD1*/*PKD2*. Interestingly, both the expression of *PKD2* mRNA and Polycytin-1/2 were reversely correlated with that of *miR-105-5p*, whereas the expression of *PKD1* did not change during KD. Therefore, it is possible that *miR-105-5-5p* inhibit *PKD1* via translational repression while suppress *PKD2* via transcript degradation. Furthermore, the gain- and loss- of function assay from KD 0 to KD 3 revealed that *miR-105-5p* might contribute to mesoderm formation by regulating Wnt signaling pathway. Thus, we provided the evidence that *miR-105-5p* is an important regulator in the early stage of embryonic kidney development.

2249T

Early embryo stride over the mountain in regulatory gene networks to achieve differentiation - The morula is the turning point of the embryonic development. *X. Kong, Y. Zhang, M. Jiang, Y. Zhu.* Molec Gen, Inst Hlth Sci, Shanghai, Shanghai, China.

Background/Objective: Early embryonic development is a complex process that involves a cascade of morphologic changes that transform an oocyte into a zygote, morula, and blastocyst and finally into an embryo implanted in the uterus. Changes in genes promote the development from one stage to the next. But how these genes are activated and how the gene regulatory relationship changes are still unclear. It is worthy to study the changes in gene regulatory network during early embryonic development. **Method:** We examined single-cell RNA sequencing data of preimplantation embryos to define gene regulatory network for each stage. K-core scoring was used to simply depict the structure of network and identify the core regulatory genes. Genes were grouped into co-expression modules to examine the preservation of gene relationships between stages. The critical transition of gene relationships were identified by the dynamical network biomarker (DNB) analysis. The changes of connection number and connecting targets were considered to identify the different connection genes. **Results:** We found that the extent of regulatory connection rapidly increased from the 4-cell to the morula, after which the number of connections suddenly decreased. We identified the core regulatory genes in the each stage and found gene networks were different between stages. We further confirmed the morula is a critical transition stage of gene connections. The morula can be compared to the mountain where one side is non-differentiated cells and the other side is differentiated cells. We discovered that the genes showing significant connection changes between before and after morula became the core regulatory genes after morula and we found these genes were prone to be up-regulated in the morula. The core regulatory genes of morula seemed to be the main triggers to actuate striding across the morula mountain. These results have potential ramifications for the mechanisms controlling early embryonic development, for central regulatory factors that might be good candidates for further analysis of cell differentiation.

2250F

Opposing effects of FGF2 and TGF-β/Activin/Nodal signalling inhibition on BMP4-mediated trophoblast differentiation of hESCs. *M. Koel[1,2], U. Võsa[3], K. Krjutškov[1,4,5], E. Einarsdottir[4,5], J. Kere[4,5], J. Tapanainen[6], S. Katayama[4], S. Ingerpuu[2], V. Jaks[2,7], U.-H. Stenman[8], K. Lundin[6], T. Tuuri[6], A. Salumets[1,9,10].* 1) Competence Centre on Health Technologies, Tartu, Estonia; 2) Department of Cell Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 4) Department of Biosciences and Nutrition, and Center for Innovative Medicine, Karolinska Institutet, Huddinge, Sweden; 5) Molecular Neurology Research Program, University of Helsinki and Folkhälsan Institute of Genetics, Helsinki, Finland; 6) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 7) Department of Biosciences, Karolinska Institutet, Huddinge, Sweden; 8) Department of Clinical Chemistry, University of Helsinki, Helsinki, Finland; 9) Department of Obstetrics and Gynaecology, University of Tartu, Tartu, Estonia; 10) Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia.

Several studies have demonstrated that human embryonic stem cells (hESC) can be differentiated into trophoblast-like cells if exposed to morphogenic protein 4 (BMP4) and/or inhibitors of fibroblast growth factor 2 (FGF2) and the transforming growth factor beta (TGF-b)/Activin/Nodal signalling pathways. The goal of this study was to investigate how much the inhibitors of these pathways can individually improve the efficiency of hESC differentiation when compared to basic BMP4 treatment. RNA sequencing was used to analyse the effects of all possible inhibitor combinations on the differentiation of the hESCs into trophoblast-like cells over 12 days. Genes differentially expressed compared to untreated cells were identified in seven time points. Additionally, expression of total chorionic gonadotropin (hCG) and its hyperglycosylated form (hCG-H) were determined by immunoassay from cell culture media. We showed that sole FGF2 inhibition up-regulates syncytiotrophoblast-specific genes (*CGA*, *CGB* and *LGALS16*), induces several molecular pathways involved in embryo implantation and triggers hCG-H production. In contrast, inhibition of the TGF-β/Activin/Nodal pathway decreases the ability of hESCs to form trophoblast-like cells. This information about the conditions needed for hESCs differentiation toward trophoblast-like cells helps us to find an optimal model for studying the early development of human trophoblasts in normal and complicated pregnancy.

2251W

Partial characterization of the content of extracellular microvesicles released by cultured murine chondrocytes. *L.R. Aguiar, R. Pogue.* Universidade Católica de Brasília, Brazil.

Chondrocytes are the constituent cells of cartilage, residing within an extracellular matrix made up principally of collagens and sulfated proteoglycans. These cells constantly send and receive signals, employing various signaling pathways such as TGF-b, BMP, hedgehog, WNT and FGF. These signals result in various cellular activities such as alterations in gene expression, differentiation and apoptosis. Recently, a new player has been identified in the paracrine signaling processes occurring in all tissues. Extracellular microvesicles are membrane-encapsulated structures that are released by all cell types so far tested, and are present in all bodily fluids. They are commonly classified based on their mechanism of biogenesis (exosomes, microvesicles, apoptotic bodies), and have been widely shown to contain lipids, proteins and different nucleic acids, which may reflect their cell-type of origin. The objective of the current study was to collect the microvesicular fraction released into culture media by murine chondrocytes, and to test for the presence of well-known chondrocyte-associated transcripts within that fraction. We detected the presence of mRNA representing Col2a1, Aggrecan, WDR5, FGFR2, FGFR3, IHH, BMP4 and GAPDH, as well as miR16, miR100, miR125, miR140 and miR328. The RNA content of the microvesicles reflects their chondrocytic origin and gives insight into their possible roles in inter-cellular communication within cartilage.

2252T

Role of the apico-basal polarity gene *SCRIBBLE1* in the pathogenesis of neural tube defects. *F. Kharfallah[1], A. El Hassan[1], M.C.G. Guyot[1], E. Merello[2], P. De Marco[2], G. Di Cristo[1], V. Capra[2], Z. Kibar[1].* 1) Chu Sainte-Justine. Montreal University, Montreal, Quebec, Canada; 2) U.O. Neurochirurgia, Instituto G. Gaslini, Genova, Italy.

Scribble (*Scrib*) is a key regulator of apicobasal (AB) polarity of epithelial cells and has long been studied as a tumor suppressor gene in mammals. *Scrib1* was also implicated in another form of polarity called planar cell polarity (PCP) where epithelial cells become polarized along the plane of the epithelium. In vertebrates, PCP, also called the non-canonical Frizzled/Dishevelled (Fz/Dvl) pathway, mediates a morphogenetic process called convergent extension (CE) during gastrulation and neural tube formation. *SCRIB1* encodes a large, 195-kDa cytoplasmic scaffolding protein that comprises a large leucine-rich repeat (LRR) region and 4 PDZ domains that regulate protein-protein interactions, including the PCP protein *VANGL2*. *Scrib1* physically and genetically interacts with *Vangl2* and plays a direct role in its asymmetric targeting during PCP signaling in the cochlea. In mouse Circletail (Crc), homozygous mutations at *Scrib1* cause a lethal form of neural tube defects (NTDs) called craniorachischisis whereby the neural tube remains open throughout the spinal cord. In humans, rare variants in SCRIB1 were associated with NTDs in a small fraction of patients. This project aims at investigating the role of *SCRIB1* as a gene involved in both ABP and PCP in a large cohort of human NTD patients and in the mouse mutant Crc. We identified 5 rare missense mutations in 6 NTD patients that were absent in 400 controls analyzed and predicted to be pathogenic using bioinformatics. Using the yeast two hybrid system, we identified two mutations that affect the well- established interaction between SCRIB1 and VANGL2. We also identified two NTD-associated mutations that affect protein localization of SCRIB1 in MDCK cells. In parallel, we used the Crc mouse to look at the expression pattern of Scrib1, Vangl2 and other polarity markers. Our study will help us better understand the role of *SCRIB1* as a molecular link between ABP and PCP during normal and abnormal neural tube development.

2253F

Investigating the role of SOX6 during the development of the peripheral nervous system. *C. Gopinath[1], W.D. Law[2], J.F. Rodríguez-Molina[3], A.B. Prasad[4], L. Song[5], G.E. Crawford[5,6], J.C. Mullikin[4], J. Svaren[7,8], A. Antonellis[1,2].* 1) Cellular and Molecular Biology Program, University of Michigan Medical School, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 3) Cellular and Molecular Pathology Program,University of Wisconsin-Madison, Madison, WI; 4) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Center for Genomic and Computational Biology,Duke University Medical Center, Durham, NC; 6) Department of Pediatrics, Duke University Medical Center, Durham, NC; 7) Waisman Center,University of Wisconsin-Madison, Madison, WI; 8) Department of Comparative Biosciences, University of Wisconsin-Madison, Madison, WI.

Myelinating Schwann cells insulate motor and sensory axons in the peripheral nervous system (PNS) to allow rapid saltatory conduction along peripheral nerves. The transcription factor SOX10 is essential for all stages of Schwann cell development and activates the expression of key myelin genes in the PNS. Thus, identifying and characterizing SOX10 target genes will improve our understanding of Schwann cell function. Toward this, we performed computational and functional analyses that revealed highly conserved, intronic SOX10 response elements including one at the *SOX6* locus. Interestingly, SOX6 competes with SOX10 for DNA binding at myelin-related loci in oligodendrocytes, the glial cells of the central nervous system. This results in inhibition of premature myelination during oligodendrocyte development. It is possible that a similar mechanism occurs in the PNS; however, the function of SOX6 has not been studied in the PNS and SOX10 has not been shown to regulate *SOX6* transcription in any cells. Our hypothesis is that SOX10 activates *SOX6* expression during Schwann cell development. Subsequently, SOX6 inhibits the expression of a subset of SOX10 target genes to maintain a premyelinating state. Our data demonstrate that *SOX6* is expressed in Schwann cells and that the enhancer identified in the initial screen acts as an alternative promoter. To determine if SOX10 regulates *SOX6* via this alternative promoter, we deleted the SOX10 consensus sequences, which resulted in reduced promoter activity in cultured Schwann cells. Additionally, expressing SOX10 in SOX10-negative motor neurons induced the activity of this promoter and enacts transcription of endogenous *SOX6* mRNA expressed from this promoter. Combined, these data indicate that SOX10 regulates the expression of *SOX6* via an alternative promoter. The SOX10 site within this promoter is conserved to zebrafish, therefore we are deleting this promoter and the associated first exon in zebrafish using CRISPR-Cas9 technology. Future work will focus on determining if *SOX6* represses SOX10 function in myelinating Schwann cells and to test this we will overexpress SOX6 in myelinating Schwann cells and perform RNA-seq to identify target loci that may be regulated by SOX6. In sum, we have identified SOX6 as a target gene of SOX10 in Schwann cells. Our future efforts will reveal the role of SOX6 in the peripheral nerve and improve our understanding of the regulatory pathways important for Schwann cell development.

2254W

Expression quantitative trait and allele specific analyses identify tissue-specific associations in a non-human primate model. *I. Zelaya[1,2], S. Service[1], A. Jasinska[1], M.J. Jorgensen[4], J. Kaplan[4], W. Warren[5], R. Woods[3], G. Coppola[1,3], N. Freimer[1].* 1) Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA; 2) Bioinformatics Interdepartmental PhD Program, University of California, Los Angeles, Los Angeles, CA; 3) Department of Neurology, University of California, Los Angeles, Los Angeles, CA; 4) Department of Pathology, Section on Comparative of Medicine, Winston-Salem, NC; 5) McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Avenue, St Louis, MO 63108, USA.

Genome-wide association studies over the past decade have identified numerous variants associated to disease, most of them located in non-coding regions of the genome. Expression quantitative trait locus (eQTL) analysis is a powerful approach to identify associations between genotypes and gene expression, however systematic eQTL studies relevant to humans are limited by either sample availability (for human tissues) or evolutionary distance (rodent and invertebrate tissues). The vervet monkey (*Chlorocebus aethiops sabaeus*) constitutes an attractive model for these studies, as considerable tissue resources have already been collected and the genome sequenced. We obtained RNA-seq data from six different tissue types from 58 animals at developmental stages ranging between birth and adulthood, and performed an eQTL analysis. We used 500k SNPs with a MAF >= 0.10 as mapping set. To test for association of genotypes to gene expression we used the Efficient Mixed-Model Association eXpedited (EMMAX) statistical test, including age and sex as covariates. We identified hundreds of genes with local tissue-specific associations and a handful of genes with distant tissue-specific associations. The number of associations shared among all six tissues, both local and distant, was much smaller, reinforcing the idea that genetic variation plays different roles across tissues. We also completed allele specific expression analyses. The identified loci significantly overlapped with our local eQTL results (p= 6.7e-09), and a handful of genes were identified as having only allele-specific expression differences, occurring possibly at the isoform level, and not detectable by our eQTL analysis. This work was supported by the NIH grants R01RR016300/R01OD010980, R37MH060233-09S1, U54HG003079 and P40OD010965.

2255T

The use of large neutral amino acids (LNAAs) as a dietary supplement is not effective in reducing plasma Phe levels but improves neuropsychological performances in Phenylketonuria. *I. Scala[1], M. Marino[2], M.P. Riccio[2], C. Bravaccio[2], G. Parenti[1], P. Strisciuglio[1].* 1) Department of Pediatrics, Clinical Genetic Unit, Federico II University Hospital, Naples, Italy; 2) Department of Pediatrics, Child Neuropsychiatric Unit, Federico II University Hospital, Naples, Italy.

Introduction: Phenylketonuria (PKU) is a metabolic disorder caused by phenylalanine hydroxylase deficiency. The treatment is based on a low-Phe diet. The diet begins at birth and continues for a lifetime, with regular monitoring of phenylalanine (Phe) concentrations. Large neutral amino acids (LNAA) therapy has been developed as an alternative to dietary restriction: LNAAs compete with Phe at the level of the brain and intestinal transporters, and may be useful in reducing Phe blood concentrations. Studies showed a possible 20 to 50% reduction of plasma Phe levels in subsets of PKU subjects, while others showed improvements of neurocognitive performances. However, data are still inconclusive. Objectives: To evaluate the effectiveness of LNAAs in reducing plasma Phe levels and improve neuropsychological performances in adult PKU subjects. Patients and methods: 9 subjects with PKU (4F/5M; 7 cPKU, 2 mPKU; mean age: 24 +/- 4,4 years) were supplemented for 9 months with LNAA (Moviscom, 0.5-1 g/kg) divided in 3 daily doses at main meals, with complete replacement of the previous non-LNAA formula. Each patient underwent to Phe, Tyr, Phe/Tyr ratio monitoring; neuropsychological assessment at T0 and T + 3 months by using the American Psychological General Well-Being Index (PGWBI), the Wisconsin Card Sorting Test (WCST), TAP tests and HPG test. Results: LNAA supplementation was ineffective in reducing plasma Phe levels (p>0.05), while tyr levels increased in all patients (mean±DS: 49.1 ± 15.27 *vs* 104.02 ± 23.41; p <0.01). Post-treatment Phe/Tyr ratio decreased up to 59% in all patients. Neuropsychological assessment showed: significant increase in all areas of the WCST test in 9/9 patients (45-65° centile at T0 vs 75-85° centile at T+3); improvements of anxiety, positivity, wellness, vitality and self-control areas at the PGWBI tests; improved performance in the dominant hand in 2/9 patients at the 9-HPG; decreased reaction times of alertness in all the analysed patients; improvement of the reaction time in 6/9 subjects at the sustained attention test. Conclusions: LNAA therapy was not effective in reducing plasma Phe levels, but was effective in increasing plasma levels of Tyr. The neuropsychological evaluation showed a global improvement of sustained attention, alertness, planning, problem solving, an improvement of anxiety, self-control and vitality and this improvement was independent of Phe metabolic control.

2256F

Stable isotope breath testing as a measure of methylmalonyl-coa mutase (MUT) activity in methylmalonic acidemia. *J. Gagné[1], I. Manoli[1], E. Harrington[1], S. Smyth[2], J. Hattenbach[2], J.L. Sloan[1], K.Y. Chen[2], C.P. Venditti[1].* 1) National Human Genome Research Institute, Bethesda, MD; 2) National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

Isolated methylmalonic acidemia (MMA) is a grave autosomal recessive inborn error of metabolism, caused by a defect in methylmalonyl-CoA mutase (MUT). In some severe patients, liver (LT), kidney (KT), or combined liver and kidney (LKT) transplantation has been used as a surgical treatment for MMA. Previous studies have shown that $1\text{-}^{13}C$-propionate oxidation can be used to measure whole body MUT activity in gene therapy treated and transgenic MMA mice and have encouraged translation to patients, with the aim of using this approach as an outcome measure for therapeutic interventions such as gene therapy. We therefore investigated $1\text{-}^{13}C$-propionate oxidation in 31 patients with isolated MMA (*mut[0]* N=19, *mut[-]* N=6, *cblA* N=6, age range: 4-41 years), including 3 KT, 1 LT, and 4 LKT recipients, as well as one patient who received a heterotopic liver allograft post KT (hLKT). In addition, 7 healthy volunteers (3 males, 4 females, age range: 21-44 years) and 8 heterozygote controls (age range: 35-62 years) were tested to establish test reproducibility. Baseline CO_2 production of the participants was measured using an indirect calorimetry cart before sodium $1\text{-}^{13}C$-propionate was administered as an oral or G-tube bolus, and breath samples were collected serially over 2 hours via disposable breath collection kits to measure $^{13}CO_2$ enrichment. Propionate oxidation was decreased in all non-transplanted MMA patients compared with controls (*p*<0.0001). Both LT and LKT recipients demonstrated oxidation rates similar to control levels (*p*=NS), as did KT recipients responsive to vitamin B-12 supplementation (N=2) with *cblA* MMA. However, the remaining KT recipient, a severely affected mut[0] patient carrying two stop mutations, showed almost no metabolism of label. Repeat testing conducted on healthy volunteers and 9 MMA patients (7 *mut*, 2 *cblA*) supported excellent reproducibility. Notably, similar oxidation rates were observed in 2 patients despite markedly different plasma methylmalonic acid concentrations between one assessment and the next. This suggests that measured $1\text{-}^{13}C$-propionate oxidation may be a more consistent and reproducible clinical biomarker than plasma metabolites that depend, in part, on protein intake and kidney function. This safe and minimally invasive method therefore has great potential as an outcome measure for enzyme replacement strategies in MMA, such as transplantation, and gene or cell therapy.

2257W

Disease symptoms and impacts as reported by patients with acid sphingomyelinase deficiency (ASMD) and caregivers. *R. Avetisyan[1], A. Hareendran[2], S. Stringer[3], S. Tan[1], B.J. Sanson[1], S. Hass[1].* 1) Sanofi Genzyme, Cambridge, MA, United States; 2) Evidera, London, United Kingdom; 3) Evidera, Bethesda, MD, United States.

Acid sphingomyelinase deficiency (ASMD), also called Niemann-Pick Disease Types A and B (NPD A and B), is a rare, progressive, life-threatening genetic disorder with an estimated incidence of 0.4-0.6 per 100,000. The metabolic defect, deficiency of the enzyme acid sphingomyelinase, leads to abnormal accumulation of sphingomyelin resulting in multiple organ damage and dysfunction. ASMD represents a spectrum of phenotypes; the most severe form (infantile neurovisceral ASMD, NPD A) is characterized by early onset of disease and rapid progression of psychomotor degeneration, leading to death, usually by 3 years of age from respiratory failure. An intermediate, chronic neurovisceral form has later symptom onset and slower disease progression. Patients with chronic visceral ASMD (NPD B), lack neurological symptoms but have significant hepatosplenomegaly, respiratory, gastrointestinal, musculoskeletal, and hematological manifestations. The primary objective of this qualitative research study was to obtain information directly from patients and caregivers about the disease symptoms and impact on patient's lives. Interviews were conducted with 11 adult and 9 pediatric patients with chronic ASMD and/or their caregivers. The interviews included both open-ended and structured questions; structured questions were developed based on literature review and clinical expert interviews. Questionnaires were also administered to evaluate demographic and clinical characteristics of the patients. The mean age of patients was 37 (SD 15.5) years and 10 (SD 4.7) years for adults and children, respectively, and 64% of adult and 45% of pediatric patients were female. The most commonly reported symptoms were enlarged abdomen (due to hepatosplenomegaly), breathing difficulties, frequent respiratory infections (including pneumonia), abdominal pain, bleeding/bruising, back and limb pain, and fatigue. Specific symptoms reported by patients to be most bothersome/ important to them included: fatigue, shortness of breath, bruising/ bleeding, lower limb/ joint pain, sleep disturbances, headaches, and diarrhea. ASMD was found to substantially impact physical, social, and emotional functioning in both adults and children. This study describes symptoms experienced by patients with chronic ASMD and the significant impact of the disease on patients' lives. The findings also offer a framework for the development of an ASMD disease-specific patient-reported outcome measure. Supported by Sanofi Genzyme.

2258T

Diverse clinical manifestations and heterogeneous natural history of acid sphingomyelinase deficiency. *M. McGovern[1], R. Avetisyan[2], B.J. Sanson[2], O. Lidove[3].* 1) Dept Pediatrics, Stony Brook Univ, Stony Brook, NY; 2) Sanofi Genzyme, Cambridge, MA; 3) Hôpital de la Croix Saint Simon, Paris, France.

Acid sphingomyelinase (ASM) deficiency (ASMD), historically known as Niemann-Pick disease (NPD) type A and B, is a rare autosomal recessive lysosomal storage disorder caused by different *SMPD1* mutations. ASMD pathogenesis is characterized by the abnormal accumulation of the primary ASM substrate, sphingomyelin, in cells of the monocyte-macrophage system in various organs. Current treatment options are limited to symptom management and supportive care. An enzyme replacement therapy is in clinical development. Early disease recognition and timely diagnosis of ASMD is important, but often hampered by the rarity and phenotypical heterogeneity of the disease. The following review of the published literature and the synthesis of the evidence is intended to facilitate a better understanding of ASMD manifestations and natural history. The natural history of patients with NPD A is almost uniformly characterized by onset in early infancy and rapidly progressing psychomotor degeneration, which typically results in death by age 3 years from respiratory failure. In contrast, patients with NPD B have a widely heterogeneous natural history. NPD B is a chronic progressive disease that may affect multiple organs, with typical manifestations including hepatosplenomegaly, progressive infiltrative lung disease, liver dysfunction, heart disease, growth delay, osteopenia/osteoporosis, atherogenic lipid profile, and cytopenias. Bleeding, shortness of breath, joint/limb pain, abdominal pain, and bruising are common symptoms and may result in chronic fatigue, limited physical activity, inability to work, and social isolation. Patients with a variant form of NPD B may also have neurologic symptoms, which range from mild hypotonia to cognitive disability or progressive motor degeneration. Most patients with NPD B succumb to disease complications before or in early adulthood; common causes include pneumonia, respiratory failure, bleeding, and liver failure. ASMD is a rare, multi-systemic, and heterogeneous disease, and these features contribute to low disease awareness and delays in diagnosis. Despite recent progress in the characterization of ASMD, more information on disease progression, subpopulations, genotype-phenotype relationships, prognostic factors, and impact on quality of life is needed. A better characterization of ASMD subpopulations and their natural histories may facilitate earlier disease recognition in this rare patient population. Supported by Sanofi Genzyme.

2259F

Two sibling cases of aspartate-glutamate carrier 2 (Citrin) deficiency: Does diet affect prognosis? *Y. Watanabe[1,2], K. Fukui[1], N. Harada[1], K. Tashiro[2], T. Inokuchi[2], S. Yano[1,3], Y. Yamashita[1], Watanabe.* 1) Dept Pediatrics, Kurume Univ, Kurume, Japan; 2) Research Inst GC/MS, Kurume Univ, Kurume, Japan; 3) Genetics/Pediatrics, USC, Los Angeles, CA, United States.

   Background: Citrullinemia Type 2 (CTLN2) is due to the defective mitochondrial aspartate-glutamate carrier isoform 2 (AGC2 or citrin). Citrin is a component of the malate-aspartate shuttle, which transfers reducing equivalents from NADH from the cytosol to the mitochondria and is expressed in the liver. Patients with CTLN2 develop sudden onset hyperammonemia with neuropsychiatric symptoms in adulthood. Mutations in citrin are also known to cause neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD). Patients with citrin deficiency often have characteristic food preference for high protein and high fat diet, and aversion to carbohydrate rich diet. Treatment with sodium pyruvate has been shown to have beneficial effects in lowering lactate/pyruvate ratio and ureogenesis in the liver-perfusion system in citrin-deficient mice, possibly by lowering cytosolic NADH/NAD ratio. **Case Study**: Two Japanese full siblings (62y male and 59y female) with citrin deficiency showed significantly different clinical courses, possibly due to their diet. The elder brother is an owner chef of a "Tempura" restaurant for more than 30 years. He had been well until age 19y when he developed recurrent pancreatitis. Liver biopsy at age 56y due to chronic liver disease, showed the findings suggestive of Hemochromatosis and fatty liver. At age 62y, he developed hepatic coma with hyperammonemia (312uM). Plasma amino acid analysis showed high citrulline, suggestive of CTLN2. Plasma lipid and lipoprotein analysis showed normal total cholesterol, low HDL, and low apoA1 and apoB concentrations. Lipoprotein fractionation showed a midband between the LDL and VLDL bands. He has been on protein rich diet; however, he cannot tolerate high fat diet due to chronic pancreatic dysfunction. His sister has been well without any specific citrin deficiency related symptoms. She likes high protein food and "Tenkasu", crunchy surface of Tempura, and avoids high carbohydrate diet. **Results:** Compound heterozygous mutations in *SLC25A13*[g.IVS13+1G>A/g.IVS16ins3kb] were identified, confirming the diagnosis of citrin deficiency in the two siblings. Implementation of MCT oil in the brother improved the liver function including total bile acid metabolism. **Discussion**: Sodium pyruvate has been used to treat Citrin deficiency. Supplementing fatty acids including MCT oil may also have beneficial effects, possibly due to increasing mitochondrial matrix NADH/NAD ratio leading to ATP synthesis.

2260W

C26-ceramide is a new and sensitive biomarker for Farber disease. *P. Bauer[1], C. Cozma[1], M.I. Iurascu[1], S. Zielke[1,2], A. Rolfs[1,2].* 1) Centogene AG, Rostock, Germany; 2) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Germany.

   Background: Farber disease (Farber's lipogranulomatosis, ceramidase deficiency), is an autosomal recessive, extremely rare disease caused and characterized by a deficient acid ceramidase activity encoded by ASAH1 gene. Low ceramidase activity is resulting in accumulation of fatty substances, mainly ceramides. At clinical level, Farber disease is manifesting through hallmark symptoms such as: periarticular nodules, lipogranulomas, swollen and painful joints and a hoarse voice or a weak cry; in addition to these, also hepatosplenomegaly, rapid neurological deterioration or developmental delay are reported. Seven different Farber types were described, with phenotypes varying from mild cases with a longer life expectancy to very severe cases, where the patients do not survive past their first year of life. Methods: The screening through over 40 different ceramide-like molecule show that only C26 is specifically increased in samples from Farber patients. Results: We present here a new method of diagnosis of Farber disease by determining the concentration of C26 ceramide isoforms using LC/MRM-MS and C25 ceramide as internal standard. Moreover, we found that cis-isomer of the C26 ceramide is a specific biomarker for Farber disease, with pathological values in a range of 39.2-150.0 nmol/L blood (normal range 13.6-23.4 nmol/L blood, N=192, healthy individuals). The current study based on dried blood pots samples from 11 Farber patients and 11 Farber carriers – the highest number of Farber patients enrolled in one study to present. The new biomarker can be determined directly in the dried blood spot extract with low sample consumption, easy sample preparation, high reproducibility and it presents the possibility of to be used in high throughput screenings. Conclusion: Cis-C26 ceramide is a specific biomarker for Farber Disease.

2261T

Lyso-sulfatide is a specific and sensitive biomarker for the early diagnosis of metachromatic leukodystrophy. *S. Eichler[1], C. Cozma[1], G. Kramp[1], M.I. Iurascu[1], L. Demuth[1], S. Oppermann[1], S. Zielke[2], A. Rolfs[1,2].* 1) Centogene AG, Rostock, Germany; 2) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Germany.

   Background: Metachromatic leukodystrophy (MLD) is a lysosomal storage disease caused by mutations in ARSA gene, inherited in autosomal recessive manner. MLD is characterized by a low aryl sulfatase A (ARSA) activity which leads to the accumulation of sulfatides (sulfated glycocerebrosides) in tissues, destroying the myelin sheath of the nervous system with serious consequences manifested in clinical symptoms such as: weakness, muscle rigidity, developmental delay, blindness, convulsions, paralysis, and dementia. A recent studies suggested that lyso-sultatides (sulfatide without an acyl group or sphingosine-glycoside-sulfate) are the toxic species responsible for the development of MLD. Recently developed enzyme replacement therapy and the steep progression of the disease in ARSA patients make the development of high throughput screening diagnosis of high importance. **Methods:** We report here the development of biochemical screening method for ARSA patients using a combination of two techniques: (i.) arylsulphatase enzymatic activity in leucocytes pellet by fluorimetry and (ii.) quantification of ARSA specific biomarker lyso-sulfatide by multiple reaction monitoring mass spectrometry (LC/MRM-MS). **Results:** 29 subjects with reduced ARSA enzymatic activity enrolled in the study (27 ARSA clinically and genetically confirmed ARSA patients; 2 patients with ARSA pseudodeficiency clinically confirmed) were involved in the current study. ARSA gene was sequenced for all patients enrolled in the study and MLD diagnosis was genetically confirmed. Arylsulfatase A activity was investigated in leukocytes extract and lyso-sulfatide was investigated in plasma of MLD patients, pseudo-ARSA patients and normal controls. Lyso-sulfatide was found to be 0.54 ± 0.29 ng/mL for MLD patients and below limit of quantification for normal controls and pseudo-arylsulfatase A deficiency patients. The results were subjected to ROC analysis and show a specificity and sensitivity of over 99% for the analyzed samples. **Conclusion:** Lyso-sulfatide is a sensitive biomarker for MLD **Keywords**: ARSA deficiency, AS enzymatic activity, ARSA biomarkers .

2262F

Lyso-SM-509 is an easy-measurable and sensitive biomarker for Niemann-Pick disease: A one-year study. *A. Rolfs[1,2], C. Cozma[1], G. Kramp[1], G. Oprea[1], S. Eichler[1], A. Giese[2].* 1) Centogene AG , Rostock, Germany; 2) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Germany.

   Niemann Pick Type C (NPC) disease is an autosomal recessive disease caused by mutations in NPC1 or NPC2 genes translated in defects of the lysosomes transport system (namely cholesterol transporters) leading to abnormal accumulation of cholesterol and glycolipids in the lysosome. Although other organs may be affected (e.g. hepatosplenomegaly), NPC is characterized mainly by progressive neurological deterioration manifested through symptoms e.g. cerebellar ataxia, epilepsy, cataplexy, dystonia, spasticity, hypotonia, ptosis, microcephaly, psychosis, progressive dementia. Recently developed treatment renders the NPC diagnosis of high importance. We present here our experience regarding the NPC diagnosis using lyso-SM-509 biomarker concentration determination, followed by genetic analysis of NPC1 and NPC2 genes. The levels of lyso-SM-509 in blood reflect the burden of the NP disease and it can be used for the easy diagnosis of NP patients and for the monitoring of the disease progression. Determination of lyso-SM-509 is performed by multiple reaction monitoring mass spectrometry (LC/MRM-MS), an analytical method proven to be reliable and reproducible in plasma, serum, EDTA blood and dried blood spots (DBS). Moreover, by combining the results with the levels of lyso-SM-465 the type A/B and type C patients can clearly be distinguished. We validated the lyso-SM-509 biomarker for clinical diagnostic in samples of patients suspected of NP in our laboratories sent by customers from all over the world. We identified 186 NPC patients and 110 NP A/B patients using lyso-SM-509 and the diagnosis was directly confirmed by sequencing of the NPC1/NPC2 or SMPD1 genes. Lyso-SM-509 has a sensitivity of 100 % and specificity of 99.15 % for the diagnosis of NP. Most of the NPC cases were diagnosed at an age of 3 to 10 years (30.65 %). The gender distribution was even between the two sexes. We have identified a total of 426 pathological alleles in the different NPC cases. 75 unique variants (44.6%) were previously published, and 94 (56.4%) are described in CentoMD® exclusively. From the 169 unique NPC1 and NPC2 variants, 87% are located within the coding genetic regions, and 13% have non-coding location. All cases identified lyso-SM-509 biomarker were genetically confirmed proving its specificity. The ease of DBS based lyso-SM-509 makes the marker an ideal parameter for the simple and confident diagnosis of NP as well as the monitoring of the diagnosed NP patients.

2263W

Recessively inherited *LRBA* mutations cause neonatal diabetes. *M.B. Johnson[1], E. De Franco[1], A. Al Senani[2], N. Elbarbary[3], Z. Siklar[4], M. Berberoglu[4], Z. Imane[5], M. Blud[6], A. Haghighi[7], I. Ullah[8], D. Gardner[9], A.T. Hattersley[1], S. Ellard[1,10], S.E. Flanagan[1].* 1) Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, England, United Kingdom; 2) The Royal Hospital, Muscat, Oman; 3) Department of Paediatrics, Ain Shams University, Cairo, Egypt; 4) Ankara University School of Medicine, Turkey; 5) Childrens Hospital, Rabat, Morocco; 6) Beasat Hospital, Hamedan, Iran; 7) Harvard Medical School, Boston, USA; 8) Sultan Qaboos University Hospital, Sultanate of Oman; 9) Academia Endocrinology Department, Singapore General Hospital, Singapore; 10) Royal Devon and Exeter Hospital, Exeter, UK.

   Background and Aims Trio analysis of exome sequencing data is an effective approach for detecting causative *de novo* mutations. Compound heterozygous mutations are the next most likely aetiological cause in outbred pedigrees and their identification may lead to the characterisation of novel recessive disorders or expand the phenotype associated with mutations in known genes. We sought to identify the genetic cause of neonatal diabetes (NDM) in a patient in whom no mutation had been detected in the known aetiological genes.   **Materials and Methods** We performed exome sequencing in a patient with NDM diagnosed at 7 weeks and autoimmune lymphoproliferative disease (diagnosed at 3 years) and his unrelated, unaffected parents. A targeted next generation sequencing (NGS) panel of all known/candidate NDM genes, including *LRBA*, was used to screen 66 probands with infancy onset diabetes (diagnosed before 1 year) who were either the result of a consanguineous union and SNP typing had identified homozygosity for the *LRBA* genomic region or who had features suggestive of an immunodysregulatory disorder. Mutations identified were confirmed by Sanger sequencing or droplet digital PCR.   **Results** No pathogenic *de novo* mutation was identified in the index patient with NDM and autoimmune lymphoproliferative disease. Further analysis identified compound heterozygous *LRBA* mutations (p.S2659X/c.3156del). *LRBA* mutations were previously described in patients with chronic variable immunodysregulation 8 (CVID-8), but NDM was not reported as a feature of this condition.   Targeted next generation sequencing of *LRBA* identified homozygous *LRBA* mutations in an additional 7 patients from 6 families. All mutations are predicted to be functionally null. Diabetes was diagnosed between 6 weeks and 15 months. Other common features in the patients included autoimmune haematological disorders (5/8 patients) and enteropathy (4/8 patients). Effective personalised therapy is available for these patients in the form of Abatacept, a mimetic for the suppressive receptor that is post-translationally regulated by *LRBA*.   **Conclusion** Our study confirms the role of *LRBA* as an NDM gene and expands the disease phenotype associated with *LRBA* mutations. A combined strategy based on both genetic prior probability and likely phenotype is an effective method of identifying patients with CVID-8. A genetic diagnosis is clinically important for these patients as it predicts prognosis and will guide medical management.

2264T

Reclassifying the most common *HNF4A* MODY mutation as an example of a rare low penetrance variant. *T.W. Laver[1], K. Colclough[1,2], M. Shepherd[1], K. Patel[1], J.A.L. Houghton[1,2], P. Dusatkova[3], S. Pruhova[3], A.D. Morris[4], C.N. Palmer[4], M.I. McCarthy[5,6,7], S. Ellard[1,2], A.T. Hattersley[1], M.N. Weedon[1].* 1) Institute of Biomedical and Clinical Science, University of Exeter, Exeter, UK; 2) Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK; 3) Department of Paediatrics, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic; 4) Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, U.K; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K; 6) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, U.K; 7) National Institute for Health Research Oxford Biomedical Research Centre, Churchill Hospital, Oxford, U.K.

   Genetic variants encompass rare, pathogenic variants causing monogenic disease and common variants that predispose to polygenic conditions. However there is third category: rare, low penetrance variants which cause disease in only a subset of individuals. Dominantly acting *HNF4A* mutations cause maturity onset diabetes of the young (MODY), a young-onset (<25yrs), sulfonylurea-sensitive form of diabetes and are associated with increased birth weight and transient neonatal hypoglycaemia. Here, we reclassify the most common *HNF4A* MODY variant, p.R114W, as an example of a rare variant which causes monogenic disease with low penetrance. Despite being the most frequently reported *HNF4A* mutation there were question marks over the pathogenicity of p.R114W: functional studies reported inconsistent results, there was a lack of co-segregation in some pedigrees, and an unexpectedly high frequency in ExAC. We sequenced *HNF4A* in 2289 MODY patients and compared the frequency of p.R114W to that in 12890 controls and 9185 patients with type 2 diabetes. We confirm that p.R114W is a pathogenic variant with an odds ratio of 30.4 (95% CI: 9.79 – 125, $P=2\times10^{-21}$) for diabetes in our MODY cohort compared to controls. However p.R114W has reduced penetrance; only 54% of heterozygotes developed diabetes by age 30 compared to 71% for other *HNF4A* variants. p.R114W heterozygotes do not have the increased birth weight characteristic of patients with other *HNF4A* variants (3476g vs. 4147g, $P=0.0004$). Additionally fewer patients responded to sulfonylurea treatment (48% vs. 73%, $P=0.038$). In conclusion, we reclassify the most common *HNF4A* MODY variant, p.R114W, as an example of a rare variant which causes monogenic disease with low penetrance. It causes a distinct clinical subtype of *HNF4A* MODY with reduced sensitivity to sulfonylurea treatment and no effect on birth weight. This has implications for diabetes treatment, management of pregnancy and predictive testing of at-risk relatives and highlights the increasing difficulties of assigning pathogenicity to low penetrance variants in monogenic disease.

2265F

A case of non-autoimmune normal weight diabetes in association with multiple congenital anomalies overlapping with CHARGE syndrome. *M. Sanyoura[1], A. Bindal[1], D. Del Guadio[2], D. Waggnner[2], R. Naylor[1], S. Greeley[1], L. Philipson[1]. 1) Section of Adult and Pediatric Endocrinology, Diabetes & Metabolism, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL.*

　CHARGE syndrome is a rare multisystem autosomal dominant or sporadic disorder consisting of eye coloboma, heart defects, choanal atresia, retardation of growth and development, genito-urinary anomalies, and ear abnormalities. Many of these features, including genital hypoplasia, cleft palate, and heart defects, overlap with other multiple anomaly syndromes such as 22q11.2 deletion, Kallmann, and Kabuki Syndrome.　Here we describe an atypical case who presented with characteristic features of CHARGE syndrome and developed adult-onset diabetes.　EH is a 22 year old female with a history of multiple congenital anomalies that include right iris coloboma, choanal atresia, congenital deafness, absent ribs, dental hypoplasia, bilateral 2nd branchial cleft sinus tracts, aplasia of the posterior semicircular canals, and a history of chronic otitis. She also has amastia and athelia with hypogonadotropic hypogonadism and primary amenorrhea, but is able to have menses with estradiol replacement. In addition, she has been noted to have osteopenia and hyperlipidemia with mildly high total LDL, HDL, and triglyceride levels. Family history is only significant for one brother who has ventricular septal defect and developmental issues. Diabetes was diagnosed at age 18 with blood sugars as high as 800 mg/dL with HbA1c at 14%. BMI is 17.5 kg/m2. Antibody testing was negative with easily detectable C-peptide levels that have persisted for more than 4 years after diagnosis. She has had stable good glycemic control on low dose insulin therapy (0.2U/kg/day of mostly glargine). Recent addition of a DPPiv-inhibitor resulted in improvement of HbA1c from 7.5% to 6.7%.　Sequence analysis of all 37 coding exons and flanking intronic sequences of *CHD7* revealed no causative mutation. Large deletions or duplications were excluded by CGH. Next generation targeted sequencing panels for hypogonadotropic hypogonadism and diabetes were also negative. Trio exome sequencing revealed a de-novo mutation in *KMT2D* responsible for Kabuki Syndrome. This mutation, c.10784A>G p.Tyr3595Cys, has never been reported and is predicted to be damaging.　Our study expands on the clinical phenotype of Kabuki Syndrome and proposes an association between it and antibody negative insulin dependent diabetes. Advances in molecular genetic testing can help establish the correct diagnosis and Kabuki Syndrome should be considered in future patients with features of CHARGE syndrome without a *CHD7* mutation.

2266W

Genotype-first approach for identifying maturity-onset diabetes of the young in a clinical population. *A.H. Wardeh[1], U.L. Mirshahi[1], J. Goehringer[1], J.L. Williams[1], C.B. Manney[1], J.B. Leader[1], T.I. Pollin[3], A.R. Shuldiner[2], D.J. Carey[1], Geisinger-Regeneron DiscovEHR Collaboration. 1) Geisinger Health System, Danville, PA, USA; 2) Regeneron Genetics Center, Tarrytown, NY, USA; 3) University of Maryland School of Medicine, Baltimore, MD, USA.*

　Maturity Onset Diabetes of the Young (MODY) is a monogenic (autosomal dominant) form of diabetes that is estimated to account for 1-2% of all diabetes cases. It is usually misdiagnosed as type-1 or type-2 diabetes due to phenotypic overlap. Traditionally, MODY is defined as non-insulin dependent diabetes before the age of 25 in lean individuals, but many affected individuals do not fit this definition. Accurate diagnosis of MODY can alter treatment recommendations, often enabling the replacement of insulin with oral medication or no treatment, and better risk assessment for related individuals. There are fourteen known MODY types defined by genetic etiology. Mutations in *HNF4A*, *GCK*, and *HNF1A* account for ~90% of known MODY cases.　We used a genotype-first approach to identify individuals with potential MODY variants. Within the DiscovEHR Cohort exome sequence database we identified rare variants (<0.05% MAF) in *HNF4A*, *GCK*, and *HNF1A*. These were further classified as loss-of-function (LOF)—frameshift, stop-gain, indel, splicing, and initiation codon variants—and missense variants. The latter were classified as potential MODY variants based on data in the Leiden Open Variant Database. Diabetes phenotype was determined by the presence in the electronic health record (EHR) of diabetes ICD9, medications, and/or lab values.　In 50,791 whole exome sequences, 60 variants matched the criteria for potential MODY variants listed above (20 variants in *HNF4A*, 12 in *GCK*, and 28 in *HNF4A*). There were 180 heterozygous carriers for these variants, of which 75 were excluded based on insufficient EHR data. The remaining individuals included 34 *HNF4A*, 19 *GCK*, and 52 *HNF1A* carriers.　Seventy-eight percent of putative MODY variant carriers met criteria for diabetes (p = 0.0001, OR = 3.20, 95% CI [1.75, 5.84] compared to age, sex, and BMI matched non-carriers). These results strongly support a pathogenic role for the majority of the MODY variants identified. Based on these findings, we estimate the prevalence of MODY in this clinical population to be approximately 1 in 500.　Current efforts are focused on refining the diabetes phenotype assessment and analysis of pedigrees to further improve the identification of pathogenic variants as implemented in the ACMG/AMP standards. Identified variants will be confirmed in a CLIA/CAP-accredited laboratory, enabling disclosure of variants to patients and providers and potentially more individualized and effective and less invasive treatment.

2267T

Single *NPHS2* mutation identifies 27% of steroid resistance in Black South African Children with nephrotic syndrome: Implications for precision medicine. *C.A Winkler[1], K. Asharam[2], R. Bhimma[2], H.M. Coovadia[3], W.P. Wenkosi[2], T. Naicker[2], J.B. Kopp[4], R. Johnson[1], S. Limou[1], G.W. Nelson[1].* 1) Leidos Biomedical Research Inc./National Cancer Institute, Frederick, MD; 2) Nelson R. Mendela School of Medicine, University of KwaZulu-Natal, Durban, SA; 3) School of Public Health, University of Witwatersrand, Johannesburg, SA; 4) 6Kidney Disease Branch, National Institute of Diabetes and Digestive Diseases, Bethesda, MD, USA.

   NPHS2 mutations are a common cause of childhood steroid resistant nephrotic syndrome (SRNS), which frequently progresses to end stage kidney disease. Among children in South Africa with nephrotic syndrome, Black children are more likely to have SRNS than Indian children. We investigated whether *NPHS2* variants contributed to this disparity. Unrelated Indian and Black children in ZuluNatal Province with nephrotic syndrome (n=78) were enrolled; children with SRNS underwent kidney biopsy. *NPHS2* was sequenced for 64 NS cases and 107 ethnicity-matched controls. A replication group comprised 20 Black SR-FSGS cases and 19 controls. Children were analyzed for consanguineous inheritance and for response to immunosuppressive therapy. 55% of Indian and 97% of Black children with NS were steroid resistant ($p<10^{-5}$); 8/30 (27%) of SRNS Black children with FSGS were homozygous for *NPHS2* V260E. V260E homozygosity was replicated in 6/20 (30%) Black children with SR-focal and segmental glomeroloslcerosis (FSGS). The mutation was not observed in Indian cases or controls although one Black blood donor was heterozygous for V260E. Black children homozygous for V260E developed SRNS at an earlier age than non-carriers (34 vs 78 months, p=0.009) and none achieved either partial or complete proteinuric remission (0% vs. 47%, p=0.002). SRNS among unrelated Black children is partially explained by homozygosity for the V260E pathogenic mutation, present in 32% of SR-FSGS cases, which is not due to cryptic consanguinity among affected children. Genotyping a single *NPHS2* mutation in Black African children with NS will identify many with SR-FSGS making it possible to avoid kidney biopsy and ineffective steroid treatment. This study highlights the need for genetic studies in diverse populations. Identification of this mutation as a part of differential diagnosis would be a cost-effective alternative to kidney biopsy in homozygous carriers, and identify a large subset of patients who are unresponsive to immunosuppressive agents (specifically, oral steroid treatment and cyclophosphamide). This would spare these children the potential severe and life threatening adverse effects of these agents while at the same time reducing health care costs for an overburdened health care system.

2268F

Severe adverse reactions in 3 Mexican brothers with MPS-II to Idursulfase. *F. Campos - Romero[1], J.I. Navarrete -Martinez[2], I. Godoy - Escobar[3], G.H. Wakida -Kusonoki[3], A.E. Limon - Rojas[4], D.E. Cervantes - Barragan[2].* 1) Allergy and Clinical Immunology, Hospital Central Sur Alta Especialidad Pemex, Mexico, Mexico, Mexico; 2) Genetics, Hospital Central Sur Alta Especialidad Pemex, Mexico, Mexico, Mexico; 3) Pediatrics Hospital Central Sur Alta Especialidad Pemex, Mexico, Mexico, Mexico; 4) General Dictatorate , Hospital Central Sur Alta Especialidad Pemex, Mexico, Mexico, Mexico.

   MPS-II is a X linked lysosomal storage disease, caused by *I2S* gene mutations leading to deficit of the iduronate-2-sulfatase enzyme that results in the degradation of glycosaminoglycans pathway blockage. The disease has an extensive multiorgan involvement. Enzyme replacement therapy (ERT), is one of the few therapeutic options for the disease. The infusion is applied IV 0.5 mg/kg each week. Common adverse reaction (AR) include headache, itching, fever, hypertension, urticaria in 75% approximately. We report a series of 3 Mexican brothers with MPS-II with severe adverse reactions in the HSCAE PEMEX. Case 1:The propositus is a 3y male, product of a monocigotic twin pregnancy, from Veracruz. Repeat upper way infections and developmental delay were noticed, the diagnosis of MPS-II was stablished; with IDS/IDSP1 inversion mutation genotype. ERT began in 2008. At the 3rd, and 39th infusion presents fever 39°C, general malaise and headache. In 2011 ERT was suspended by thrombocytopenia. After 15 days he developed severe sepsis and died in 2011. Case 2: In the second twin the IDS/IDSP1 inversion genotype was confirmed. Sinubronquial chronic syndrome and allergy to rice was present. ERT began 2008, with erythema, urticaria, angioedema, respiratory distress in replacements 2,3,4 and 5. Before the 6th infusion wheezing and fever starts leading to ERT withdrawal. Two weeks after developed infusion, bleeding gums, plaquetopenia and haemolytic crises leading to treatment with gamma globulin and steroids, platelet apheresis and splenectomy. He dies in 2010.Case 3:The youngest brother has the same genotype and whole egg and rice allergy.ERT began at 2008. Adverse reactions include fever and erythema at 3rd infusion. Before 5 years of ERT he starts with fever, diaphoresis, tachycardia,erythema, respiratory distress and wheezing.We performed in February 2016 SPT with idursulfase (2mg/ml), and ID skin test (0.02,0.2 and 2 mg/ml) all results were negative, ruled out type I hypersensitivity.We re-start ERT with a 1/10 (0.0012mg/ml) dilution at 4 ml/hr, increasing every 15 minutes until tolerate 32 ml/hr with total time of 6 hours without adverse event. So far continues his ERT at doses 1:1 without AR. Idursulfase adverse reactions are common, but thrombocytopenia is a rare one. It is important in suspected AR perform SPT and IDT to make a timely diagnosis, and prevent fatal outcomes.

2269W

The eliglustat ENCORE trial: Outcomes after 4 years of treatment in adults with Gaucher disease type 1 who were previously stabilized on enzyme replacement therapy. *T.A. Burrow[1], T.M. Cox[2], G. Drelichman[3], R. Cravo[4], M. Balwani[5], A.M. Martins[6], E. Lukina[7], B. Rosenbloom[8], O. Goker-Alpan[9], N. Watman[10], A. El-Beshlawy[11], P.S. Kishnani[12], M.L. Pedroso[13], S.J.M. Gaemers[14], R. Tayag[15], M.J. Peterschmitt[14].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 2) University of Cambridge, Addenbrooke's Hospital, Cambridge, UK; 3) Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina; 4) State Institute of Haematology 'Arthur de Siqueira Cavalcanti', Rio de Janeiro, Brazil; 5) Mount Sinai Hospital, New York, NY, USA; 6) Universidade Federal de São Paulo, São Paulo, SP, Brazil; 7) National Research Center for Hematology, Moscow, Russia; 8) Cedars-Sinai/Tower Hematology Oncology, Beverly Hills, CA, USA; 9) Lysosomal Disorders Unit, O&O Alpan, LLC, Fairfax, VA, USA; 10) Hospital Ramos Mejia, Buenos Aires, Argentina; 11) Pediatric Hematology, Cairo University, Egypt; 12) Duke University School of Medicine, Department of Pediatrics, Durham, NC, USA; 13) Hospital de Clinicas, Universidade Federal do Paraná, Curitiba, PR, Brazil; 14) Sanofi Genzyme, Cambridge, MA, USA; 15) Prometrika, Cambridge, MA, USA.

Background: Gaucher disease type 1 (GD1) is caused by acid β-glucosidase deficiency leading to accumulation of glucosylceramide in lysosomes. Enzyme replacement therapy (ERT) augments a patient's own residual enzyme activity and breaks down accumulated glucosylceramide. Substrate reduction therapy (SRT) partially inhibits glucosylceramide synthase and slows down production of glucosylceramide. Eliglustat, an oral SRT, is approved in the US and Europe as first-line treatment for adults with GD1 with compatible CYP2D6 metabolizer phenotypes (>90% of patients). In the ENCORE (NCT00943111, Sanofi Genzyme) primary analysis of GD1 adults (N=159) previously stabilized on ERT, eliglustat was non-inferior to imiglucerase in maintaining stability after 12 months (Cox. *Lancet.* 2015). **Methods:** During the ENCORE trial, patients could receive eliglustat for 2 to 4+ years, depending on time of enrollment (which spanned 2 years), trial randomization, and geographic location (51 US patients were switched to commercial therapy when it became available as per the trial protocol). We report on long-term safety and efficacy with respect to years on eliglustat therapy for all 157 eliglustat-treated patients in ENCORE, 46 of whom were followed for ≥4 years. Repeated Measures ANOVA methods were used to analyze hemoglobin, platelet count, and spleen and liver volume data. **Results:** Mean values for hemoglobin, platelets, and spleen and liver volumes remained stable over 4 years. Year to year, all 4 measures also remained stable collectively (composite endpoint relative to baseline values) in ≥85% of patients, and individually in ≥91%. Year to year, all 4 therapeutic goals for hemoglobin concentration, platelet count, spleen and liver volumes for patients on ERT (Pastores. *Semin Hematol.* 2004) were maintained collectively in ≥92% of patients; each individual goal was maintained in ≥94% of patients. Mean bone mineral density Z-scores (lumbar spine and femur) remained stable and within healthy reference ranges throughout the trial. Eliglustat was well-tolerated over 4 years during the primary and extension phases; 4 patients (2.5%) withdrew due to adverse events considered related to eliglustat. No new or long-term safety concerns were identified. **Conclusions:** Adults with GD1 treated with eliglustat for up to 4 years in the ENCORE trial maintained clinical stability by both composite and individual measures.

2270T

A natural history study of patients with acid sphingomyelinase deficiency (ASMD). *L. Clarke[1], G.F. Cox[2], R. Giugliani[3], M.M. McGovern[4].* 1) Dept Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Sanofi Genzyme, Cambridge, MA, USA; 3) Department of Genetics, UFRGS, Medical Genetics Service, HCPA, Porto Alegre, Brazil; 4) Department of Pediatrics, Stony Brook University School of Medicine, Stony Brook, New York, USA.

Acid sphingomyelinase deficiency (ASMD), historically referred to as Niemann-Pick disease type A (NPD A) and B (NPD B), is a rare, autosomal recessive, lysosomal storage disorder. ASMD presents as a spectrum of severity, and is fatal by 3 years of age in its most severe, infantile neurovisceral form (NPD A). Two chronic forms of the disease also exist, an intermediate form with neurovisceral involvement (NPD B variant) and a less severe, chronic visceral form (NPD B) with no neurologic involvement but considerable disease-related morbidity and mortality. To assess the burden of disease in patients with ASMD, clinical and healthcare use data were abstracted from the medical records of 100 patients with infantile neurovisceral (13%), chronic neurovisceral (6%), and chronic visceral (81%) disease. Most patients were female (62%) and/or Caucasian (68%). Nearly half of the patients were from either the Brazil (46%) or US (41%) site. The majority of patients (98%) were diagnosed with ASMD via enzyme assay at a median age of 0.6 years (range 0.6-1.7), 2.9 years (range 0.4-42.8), and 6.0 years (range 0.5-40.1) for infantile neurovisceral, chronic neurovisceral, and chronic visceral ASMD, respectively. Height/length and weight Z-scores decreased markedly over the first 5 years of life for all three forms. After age 5, mean Z-scores continued to decrease but remained within normal limits in patients with chronic forms of the disease. Developmental delay (92%, 33%, 15%) and regression (54%, 33%, 1%) were more common in the infantile and chronic neurovisceral forms than in the chronic visceral form, respectively. Overall, the majority of patients (54%) underwent at least one surgical procedure. Annual hospitalization rates (2.23, 0.42, 0.41/yr) and emergency room visits (0.72, 0.07, 0/yr) were higher for patients with infantile neurovisceral disease than those with chronic neurovisceral or visceral ASMD. Platelet and WBC counts decreased over time in patients with chronic ASMD. Hepatic and respiratory disorders were the most common morbidities across patients. For patients with chronic visceral ASMD, earlier age at diagnosis was associated with greater morbidity. Analgesics, antibiotics, and respiratory drugs were the most commonly used medications. This chart review demonstrates significant disease burden and healthcare use in patients with ASMD. Diagnosis is often significantly delayed in patients with chronic ASMD. Funded by Sanofi Genzyme.

2271F

Generating a Drosophila model of Mucopolysaccharidosis IIIB as a testbed for chaperone therapy. *F.T.C. Silva[1,2], S.T. Sweeney[2].* 1) University of Brasília, Brasília, Brazil; 2) University of York, York, United Kingdom.

Mucopolysaccharidosis (MPS) IIIB is a lysosomal storage disorder caused by N-Acetyl-a-glucosaminidase (NAGLU) enzyme deficiency resulting in accumulation of glycosaminoglycan heparan sulfate. Patients present in childhood with neurodegeneration, aberrant behaviour, skeletal problems and reduced life span. We developed a Drosophila model for MPSIIIB with a view to identifying therapies. We generated an allelic series of mutants in NAGLU and assessed their phenotypes, totalizing 4 mutant genotypes and 1 control wild type (n=100 for each genotype). Longevity and behavioural responses (climbing and vortex recovery) were assessed. Wild type life span averaged 27.5 ± 5.2 days, while mild mutant alleles had an average of 20.6 ± 4.6 and 16.5 ± 4.8 days. Null alleles with larger genomic deletions averaged 11.7 ± 3.4 days and 12.8 ± 4.1 days. Behavioural responses for climbing showed hyperactivity in young mutants while vortex recovery failed to differentiate between genotypes. These results recapitulate longevity and hyperactivity behavior that is characteristic of affected children with MPSIIIB. We initiated experiments to test potential 'chaperone therapy'. Mutants in NAGLU patients are often mis-sense and resulting proteins could potentially produce function if the protein quality surveillance machinery did not degrade them first. We wish to rescue function by the feeding substrate mimics (chaperones) to encourage proper protein folding of NAGLU. We therefore designed transgenes encoding *NAGLU* with clinical mutations to be expressed in the *NAGLU* null background. Using a new NAGLU substrate for enzyme visualization in stained tissues of 3rd instar larvae, we can then assess protein activity and test for rescue phenotypes. *NAGLU* and mutant *NAGLU* were cloned into pUAST-AttB vector, allowing insertion in a known genomic site to ensure equal protein expression. We are now in a position to rescue *NAGLU* mutants to assess rescue by physiology, histology and enzyme activity. Developing animal models is an essential tool to test and generate new therapies for genetic disorders like MPS IIIB, and open new perspectives to improve patients' life quality and longevity.

2272W

Lyso-Gb1 correlates with Gaucher disease parameters in clinical trials using velaglucerase alfa and is a Gaucher-specific biomarker. *A. Zimran[1], D. Elstein[2], B. Mellgard[2], Q. Dinh[3], P. Chang[4], Y. Qiu[5], Y. Qin[4], C. Cozma[6], S. Eichler[6], A. Rolfs[6,7].* 1) Gaucher Clinic, Shaare Zedek Medical Center, affiliated with the Hebrew University-Hadassah Medical School, Jerusalem, Israel; 2) Shire Global Medical Affairs, Shire, Zug, Switzerland; 3) Shire Global Medical Affairs, Shire, Lexington, MA, USA; 4) Biostatistics and Statistical Programming Department, Shire, Lexington, MA, USA; 5) Bioanalytical and Biomarker Development, Shire, Lexington, MA, USA; 6) Centogene AG, Rostock, Germany; 7) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Rostock, Germany.

BACKGROUND Gaucher disease (GD) is an autosomal recessive disorder, in which glucosylceramide accumulates in macrophages, due to deficient lysosomal β-glucosidase activity. Although biomarkers have been described, none are specific to GD. Based on studies in heterogeneous populations, deacylated glucosylceramide (glucosylsphingosine; lyso-Gb1) has been identified as a potential GD-specific biomarker. METHODS Lyso-Gb1, chitotriosidase and CCL18 were quantified in patients (pts) enrolled in velaglucerase alfa (vela) clinical trials, and correlations with clinical GD parameters investigated. Pts naïve to enzyme replacement therapy (ERT-naïve) and those treated previously with imiglucerase (imi; Switch) were included. Age at baseline and genotype were considered as independent variables with respect to response. RESULTS 164 blood samples from 22 ERT-naïve pts (age 6–62 y; 7 N370S/N370S genotype) and 141 samples from 22 switch pts (6–69 y; 6 N370S/N370S) were analyzed.

Lyso-Gb1 concentrations (ng/mL)			
Visit during vela treatment	ERT-naïve (45/60 U/kg vela EoW)	Switch (15–60 U/kg vela EoW)	P value#
Week 0 (baseline)	323.2±140.5 [n=22]	91.8±84.8 [n=22]	p<0.0001
Week 53 (ERT-naïve) or 51 (switch)	117.8±78.1* [22]	82.5±85.9 [21]	p=0.0005
Week 209	60.4±39.1*† [12]	42.3±37.9 [4]	

*p<0.0001 for the change from baseline within the group; †p=0.0012 for the change from Week 53 within the group; # between groups. EoW – every other week In all ERT-naïve pts, baseline lyso-Gb1 concentrations (range 146–563 ng/mL) were higher than reported previously in untreated N370S/N370S pts (Rolfs et al. Plos One 2013). In 21/22 switch pts, baseline lyso-Gb1 concentrations were elevated (16–302 ng/mL), despite ≥22 months of prior imi. Lyso-Gb1 concentrations correlated with platelet counts (p<0.0001 for ERT-naïve pts; p=0.0109 for switch pts) and normalized spleen volumes (p<0.0001 and p=0.0008, respectively) over time in a repeated measures linear mixed-effect model, despite interpatient variation in these clinical parameters. Between weeks 25–209, lyso-Gb1 correlated with chitotriosidase and CCL18 in ERT-naïve (*r*≥0.38 and *r*≥0.44) and switch pts (*r*≥0.89 and *r*≥0.86). CONCLUSIONS For the first time, we show correlations between a GD-specific biomarker, lyso-Gb1, and clinical GD parameters and established non-specific biomarkers. Treatment with vela resulted in significant decreases in lyso-Gb1 in ERT-naïve pts and in those switched from imi, among whom only 1/22 had near-normal baseline lyso-Gb1 concentrations.

2273T

Argininosuccinate lyase deficiency: A hepatic glycogen storage disorder? *L. Burrage[1], S. Madan[1], X. Li[1], M. Jiang[1], R. Cela[1], M. Finegold[2], S. Nagamani[1], D. Bali[3], B. Lee[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pathology, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Duke University Medical Center, Durham, NC.

Urea cycle disorders (UCDs) are among the most common inborn errors of metabolism. Since the introduction of expanded newborn screening, every infant born in the United States is screened for several UCDs at birth allowing for early diagnosis and initiation of treatment to prevent hyperammonemia. However, natural history studies have demonstrated that the pathophysiology of UCDs cannot be completely explained by hyperammonemia. Thus, other metabolic mechanisms must explain many UCD complications. Liver dysfunction is one complication that may occur independently of hyperammonemia. Although various forms of liver dysfunction may occur in UCDs, patients with argininosuccinate lyase deficiency (ASLD) have higher plasma AST and ALT as compared to patients with more proximal disorders. Liver disease in ASLD ranges from evidence of hepatocellular injury to fibrosis and cirrhosis requiring liver transplantation. By light microscopy, one unique feature of the hepatic pathology of UCDs is increased hepatic glycogen deposition which is similar to that observed in glycogen storage disorders. The ASL-deficient ($Asl^{neo/neo}$) mouse which has hepatomegaly, evidence of hepatocellular injury, and increased hepatic glycogen deposition, serves as a model for studying the hepatocellular injury and hepatic glycogen accumulation observed in patients with UCDs. Quantification of hepatic glycogen after 3-hour, 12-hour, and 24-hour fasts in $Asl^{neo/neo}$ mice revealed significantly increased hepatic glycogen content compared to wild type (WT) mice at all time points with larger differences observed at the 12-hour and 24-hour time points. These results suggest that hepatic glycogen catabolism is altered in $Asl^{neo/neo}$ mice. Investigations of the activity of hepatic glycogenolytic enzymes revealed that the activity of glycogen phosphorylase is significantly reduced in $Asl^{neo/neo}$ mice compared to WT mice. The hepatocellular injury and hepatic glycogen accumulation are rescued by a helper dependent adenovirus expressing ASL using a liver specific (ApoE) promoter suggesting that the phenotype is intrinsic to hepatocytes. Studies investigating the mechanism underlying this reduction in glycogen phosphorylase activity are ongoing. Overall, our studies of hepatic glycogen storage in ASLD will provide novel insights into the relationship between the urea cycle and glycogen metabolism and may provide new avenues for research aimed at understanding long-term hepatic complications of UCDs.

2274F

Next-generation sequencing analysis of 53 patients diagnosed with *mut* methylmalonic aciduria. *J. Chu[1], M. Pupavac[1], D. Watkins[1], X. Tian[2], Y. Feng[2], S. Chen[2], R. Fenter[2], V.W. Zhang[2], J. Wang[2], L.J. Wong[2], D.S. Rosenblatt[1].* 1) Department of Human Genetics, McGill University, Montreal, Canada; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States.

Isolated methylmalonic aciduria (MMA-uria) is caused by a deficiency in either the mitochondrial methylmalonyl-CoA mutase (MCM) enzyme or the synthesis of its cofactor, adenosylcobalamin. Biallelic mutations in the *MUT* gene, which encodes MCM, are responsible for the *mut* form of MMA-uria. Patients with this disorder present symptoms of metabolic acidosis, failure to thrive, recurrent vomiting, hypotonia, lethargy, and dehydration in the newborn period. Nearly 250 different disease-causing mutations in the *MUT* gene are known. In this study, a next-generation sequencing (NGS) based gene panel, recently developed at Baylor Miraca Genetics Laboratories, was used to analyze 53 patients that had been diagnosed with *mut* MMA-uria by somatic cell complementation analysis. Two mutations in the *MUT* gene and a total of 53 different mutations were identified in 48 of 53 patients (91%). As the clinical and biochemical findings of these patients are consistent with the *mut* phenotype, it is predicted that all of the identified mutations are pathogenic. Of these, 16 mutations were novel, including the largest insertion mutation in the *MUT* gene to date. Phenotypic rescue and analysis of cDNA were used to confirm that the major insertion was responsible for the MCM deficiency observed in the patient's fibroblasts. No *MUT* gene mutations were detected in 5 of 53 (9%) patients. Review of the cellular complementation data used to initially diagnosis these 5 patients showed the results to be equivocal, putting the initial diagnosis into question. No diagnosis was possible in four of the patients. One patient was found to carry two novel mutations in the *SUCLG1* gene which are likely causal for the patient's phenotype. Combination of the NGS panel and functional analyses provided optimal diagnoses for this cohort of patients.

2275W

Tackling the Duarte galactosemia problem. *J.L. Fridovich-Keil[1], M.E. Lynch[2], N. Potter[3], C. Coles[2], J.G. Mulle[4], G. Howard[1], C.E. Barrett[2], S. Edwards[2], E.H. Wilson[3], K. Johnstone[2], J. Kable[2], M.P. Epstein[1].* 1) Dept Human Gen, Emory Univ Sch Med, Atlanta, GA; 2) Psychiatry and Behavioral Sciences, Emory Univ Sch Med, Atlanta, GA; 3) Speech and Hearing Sciences, Washington State Univ, Spokane, WA; 4) Epidemiology, Rollins Sch of Public Health, Emory Univ, Atlanta, GA.

Duarte galactosemia (DG) is one of the most common metabolic disorders identified by newborn screening in the US; it results from partial deficiency of galactose-1P uridylyltransferase. Newborns with DG display a compromised ability to metabolize galactose, a sugar abundant in milk, and manifest elevated galactose metabolites in blood and urine while drinking milk. Reports about long-term developmental outcomes in DG are mixed and healthcare providers disagree about whether intervention is warranted. Specifically, some healthcare providers advise dietary restriction of milk for their patients with DG, at least for the first year of life, while others do not. Those who recommend intervention argue that the galactose metabolites that otherwise accumulate might be deleterious. Those who do not recommend intervention argue that the documented benefits of breastfeeding outweigh any possible negative consequences of the accumulated metabolites. The ongoing disagreement leaves parents in a quandary and infants at potential risk. We are conducting a 3-year multi-state study to clarify: (1) whether children with DG, ages 6-12 years old, are at increased risk for developmental problems, and if so, (2) whether patient exposure to milk in early childhood shows an association with developmental outcomes. To conduct this study we are recruiting nearly 300 children, half with DG (cases) representing a spectrum of milk exposures, and half unaffected controls. To assess child developmental outcomes we are using validated tests of cognitive skills (especially memory, executive function, and auditory processing), communication (speech and language), physical development (including motor skills, coordination, and occurrence of tremors), social-emotional development, and auditory sensation all administered by trained professionals. We are collecting information about child diet and possible covariates using a parent/guardian response survey. We will analyze the final data set using stepwise linear regression with generalized estimating equations to determine significant predictors of outcome while controlling for relevant covariates (e.g. child age, gender, SES, sib relationships). The results will provide a foundation of shared knowledge enabling healthcare providers and the families they serve to make evidence-based decisions about what to feed their baby with DG. Depending on the outcome of the study, there also may be implications for newborn screening.

2276T

Triple therapy (arginine fortification + lysine restricted diet + pyridoxine) for pyridoxine dependent epilepsy. *S. Jaggumantri[1], C.R. Coughlin[2], W. Al-Hertani[3], A. Shuen[3], R.M. Jack[4], C. Burns[2], D. Mirsky[5], R. Gallagher[2], C. van Karnebeek[1], J. Van Hove[2].* 1) Paediatrics, University of British Columbia, Vancouver, BC, Canada; 2) Section of Clinical Genetics and Metabolism, Department of Pediatrics, University of Colorado, Aurora, CO, United States; 3) Department of Medical Genetics, Montreal Children's Hospital, McGill University of Health Centre, Montreal, QC, Canada; 4) Department of Laboratory Medicine, Seattle Children's Hospital Laboratory, Seattle, WA, United States; 5) Department of Radiology, University of Colorado, Aurora, CO, United States.

Pyridoxine dependent epilepsy (PDE) is an epileptic encephalopathy caused by mutation in *ALDH7A1* gene resulting in impaired cerebral lysine catabolism and accumulation of neurotoxic metabolites. Despite adequate seizure control with pyridoxine, 75% of individuals with PDE have significant developmental delay and intellectual disability. We describe a new combined therapeutic approach to reduce putative toxic metabolites from impaired lysine metabolism. This approach utilizes pyridoxine, a lysine-restricted diet to limit the substrate that leads to neurotoxic metabolite accumulation and l-arginine to compete for brain lysine influx and liver mitochondrial import. We conducted an open label observational cohort study in 6 PDE patients on triple therapy (Arginine fortification -150mg/kg/day, Lysine restricted diet – as per PDE consortium guidelines and pyridoxine -15-30mg/kg/day) for 6-12 months. Neurodevelopmental outcome was measured using, standard scales along with level of plasma, urine and cerebrospinal fluid (CSF) biomarkers as biochemical outcomes. Triple therapy was well tolerated and safe. Patients on triple therapy showed near normalization of urine α-aminoadiphic semialdehyde and plasma pipecolic acid. Level of biomarkers in CSF further decreased after arginine fortification in patients already on lysine restricted diet. All patients showed improvement in developmental and cognitive domains, specifically in motor skills and speech; younger patients seemed to benefit most. Furthermore, a decrease in calculated brain lysine influx was noted. Decreasing brain lysine flux via dietary restriction and arginine supplementation (competitive inhibition over the blood-brain-barrier) is effective in decreasing neurotoxic metabolite production and improving cognitive function. Residual disease symptoms could be related to early injury suggested by initial MR imaging prior to initiation of treatment or from severe epilepsy prior to diagnosis. Further studies are needed to generate more robust evidence and elucidate optimal treatment. Early diagnosis via newborn and neonatal screening allow for early intervention and potentially improved cognitive outcomes.

2277F

IEMBASE: Community-based expert knowledgebase and mini-expert platform for inborn errors of metabolism. *J. Lee[1,2], C. van Karnebeek[1,3], W. Wasserman[1,4], G. Hoffmann[5], N. Blau[6,7].* 1) Centre for Molecular Medicine and Therapeutics, BC Children's Hospital, Vancouver, BC, Canada; 2) Genome Science and Technology Graduate Program, University of British Columbia, Vancouver, BC, Canada; 3) Division of Biochemical Diseases, Department of Pediatrics, BC Children's Hospital, Vancouver, BC, Canada; 4) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 5) Department of Pediatrics, University Children's Hospital, Heidelberg, Germany; 6) Dietmar-Hopp Metabolic Center, University Children's Hospital, Heidelberg, Germany; 7) Division of Metabolism, University Children's Hospital, Zürich, Switzerland.

Objectives: Inborn errors of metabolism (IEMs) represent a large class of 500+ rare genetic disorders. The flourishing number of identified disorders can be attributed to a century-long history of research and diverse community of investigators and clinicians. Given the treatments and therapies which have become available for many IEMs, timely and accurate diagnosis is essential. Yet, it is often challenging due to their clinical heterogeneity and required knowledge on hundreds of IEMs. Such a challenge can be relieved by the rich community knowledgebase that the IEM community has established. However, access to this knowledgebase has been limited as much of the knowledge expanded before modern data systems. To address this gap, we have now developed an online platform, which combines a comprehensive community knowledgebase with a basic diagnosis support system. **Design and Methods:** We extracted disease-characterizing profiles of clinical symptoms and biochemical markers (n=2324) from an expert-generated database of 530 IEMs. The extracted IEM profiles were transferred to the nascent knowledgebase to populate. The knowledgebase was then mapped to the Human Phenotype Ontology in order to ensure interoperability with external rare disease databases, as well as exploitability by our diagnosis support system. Using this knowledgebase, we developed an algorithm which combines semantic similarity and cosine similarity in order to provide a tiered list of possible IEMs matching the user-provided disease profile. The disease-matching algorithm has undergone preliminary evaluation using 88 retrospective cases contributed by different metabolic centers. **Results:** IEMBASE (www.iembase.org) accepts an array of biochemical markers and clinical symptoms from a user and returns a ranked list of possible IEM disorders that match the input profile. In addition, the system curates expert-contributed information on IEMs, provides differential diagnoses, lists possible treatment options, and provides access to external resources. The disease-matching algorithm was able to retrieve the exact diagnosis (100% fit) in 66% of retrospective cases using a combination of clinical and biochemical information. 88% of cases had a correct diagnosis within the top 5 suggested diagnoses. **Conclusions:** We expect that this unique system dedicated to IEMs will significantly improve practice to benefit patients and families suffering these rare diseases.

2278W

Prospective treatment of argininosuccinate lyase deficiency with sodium phenylbutyrate does not prevent development of hepatic disease or recurrent hyperammonemia. *A. Yu[1], P. Chakraborty[2], M.T. Geraghty[2].* 1) Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Pediatrics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada.

Argininosuccinate lyase (ASL) deficiency is a urea cycle defect with clinical features that include hyperammonemia, neurocognitive deficiencies, hepatic disease, trichorrhexis nodosa and hypertension. Long-term management typically involves a protein-restricted diet, arginine supplementation, oral nitrogen scavenging therapy and possibly orthotopic liver transplantation. While arginine supplementation alone has been used, the addition of sodium phenylbutyrate may prevent recurrent hyperammonemia and improve hepatitis. There are 2 reported cases of prospective arginine therapy for ASL deficiency but there have been no reported cases which include the prospective use of sodium phenylbutyrate. We report a 4 year old boy who was ascertained on the basis of an affected older sister. His sister was rescued from a neonatal hyperammonemic coma and had significant global developmental delay, seizures, and liver disease. She died during an episode of hyperammonemia and status epilepticus at 4 years old. Our patient was treated prospectively with low dose arginine, sodium phenylbutyrate, and dietary restriction of protein. He did not develop neonatal hyperammonemia but despite therapy, he had ongoing hepatic disease and had 3 episodes of hyperammonemia (maximum = 435umol/L). At 2 years old, he was treated with a liver transplant. Post-transplant, he has not had any subsequent hyperammonemic events and his argininosuccinic acid level has declined. Prospective treatment of ASL deficiency with low dose arginine and sodium phenylbutyrate does not prevent hepatic disease or recurrent hyperammonemia and a liver transplant may still be required.

2279T

The natural history of HARS deficiency: Precipitous clinical deteriorations represent a critical feature. *L. Badalato[1,2], C.A. Rupar[3,4], V.M. Siu[3,4].* 1) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) University of Ottawa, Ottawa, Ontario Canada; 3) London Health Sciences Centre, London, Ontario, Canada; 4) University of Western Ontario, London, Ontario, Canada.

Histidyl-tRNA synthetase (HARS) deficiency, previously described as Usher Syndrome type 3B (USH3B MIM [614504]), is a rare autosomal recessive condition that has been exclusively observed in the Old Order Amish populations of Pennsylvania and Ontario. The causative founder mutation in *HARS* (c.1361A>C) was initially reported in 2012, and the condition was primarily characterized by progressive vision and hearing loss, ataxia, and visual hallucinations during infectious illnesses. Since then, we have identified the largest known population of patients with HARS deficiency, including 15 molecularly confirmed and 4 suspected cases in Southwestern Ontario. A striking distinction arose between our patients and the two reported in Pennsylvania: we observed a high frequency of catastrophic clinical deteriorations in response to febrile illness, which were previously described as rare. 8 of our 19 patients have died secondary to complications from febrile illnesses between the ages of 0 to 11. Three of our patients have survived similar life threatening clinical deteriorations, and none of our cohort has yet survived to adulthood. Autopsy reports from 2 of our patients suggested acute respiratory distress syndrome (ARDS), which is consistent with the clinical presentation we observed in others. Hypotension, encephalopathy, visual hallucinations, elevated CK, and gastrointestinal bleeding have also been observed with febrile illnesses. Acute vision/hearing loss and ataxia can also occur, typically with partial or full recovery once the illness resolves. Aside from these episodes, most of our population shared similar features with the Pennsylvania patients. Some were too young to manifest certain features or undergo complete ophthalmologic and audiologic exams, but where data was available, 10/11 patients had progressive vision loss (average age of onset 4.3y), and 9/11 had ataxia. 12/14 had progressive hearing loss; age of onset was <3 years in all cases, and 6 required cochlear implants. Variable developmental delays were seen in 4/12, and leukodystrophy was noted in 4/5 where brain imaging or pathology was available. In summary, here we provide the first comprehensive report of the natural history of HARS deficiency, based on the largest population of such patients reported to date. We highlight the precipitous nature of the acute clinical deteriorations that can occur in these patients, an understanding of which is crucial to optimize their care.

2280F

Rescue of neonatal lethality in a *cblC* mouse model by AAV gene therapy and hydroxocobalamin treatment enables examination of disease-associated retinal degeneration. *M.L. Arnold[1], J.L. Sloan[1], N.P Achilly[1], G. Elliot[2], I.F. Onojafe[3], B.P. Brooks[3], H. Qian[4], C.P Venditti[1].* 1) Organic Acid Research Section, Genetics and Molecular Biology Branch, NHGRI, Bethesda, MD; 2) Genomics Core, NHGRI, Bethesda, MD; 3) Pediatric, Developmental, and Genetic Ophthalmology Section, NEI, Bethesda, MD; 4) Visual Function Core, NEI, Bethesda, MD.

Combined methylmalonic acidemia and homocysteinemia, *cblC* type (*cblC*), is the most common inborn error of cobalamin (vitamin B12) metabolism and is caused by mutations in the *MMACHC* gene. MMACHC transports and processes intracellular cobalamin into its two active cofactors, 5'-adenosyl-cobalamin and methylcobalamin, necessary for the enzymatic reactions of methylmalonyl-CoA mutase and methionine synthase, respectively. Disease manifestations include growth failure, anemia, heart defects, developmental delay, neurocognitive impairment and a progressive retinopathy that causes blindness, usually by the end of the first decade. Treatment with cobalamin and other supplements improves the metabolic abnormalities but fails to improve the retinal disease. To explore pathophysiology and develop more effective therapies for *cblC*, we created a mouse model using TALENs and focused our studies on one allele: *Mmachc* c.162_164delCAC p.S54_T55delinsR (Δ3). *Mmachc*[Δ3/Δ3] homozygous mutant mice displayed early juvenile lethality (n=42, median=6 days, p<0.0001) and body weight (n=15, p<0.0001), and massive metabolic perturbations (elevated plasma methylmalonic acid, homocysteine, cystathionine and decreased methionine, p<0.05 for all metabolites). To assess the potential for gene therapy as a treatment for *cblC*, we generated two AAV vectors: rAAVrh10-CBA-m*Mmachc* and rAAV9-CBA-h*MMACHC* and compared a single intrahepatic dose (1 x 10[11] GC) of AAV in neonatal mice with biweekly injections of OH-cobalamin, the standard therapy. *Mmachc*[Δ3/Δ3] mice in all treatment groups (AAVrh10 n=11, AAV9 n=5, OH-Cbl n=8) displayed improved growth and increased survival with the majority surviving greater than 6 months (p<0.0005 for all treatments). We examined the ocular pathology in AAV and OH-Cbl treated *Mmachc*[Δ3/Δ3] mice which showed thinning of the outer nuclear layer and shortening of photoreceptor outer segments, consistent with the pathology described in treated patients. Furthermore, electroretinography (ERG) confirmed impaired retinal function *in vivo*: the a and b waves of dark-adapted *Mmachc*[Δ3/Δ3] mice were significantly reduced, indicating rod photoreceptor involvement (p<0.02). Successful treatment of *Mmachc*[Δ3/Δ3] mice enabled us to model, for the first time, *cblC* associated ocular pathology and will provide a translational platform for the development of new therapies, including gene therapy.

2281W

Lipoprotein lipase deficiency in a Mexican child homozygote for the Gly-188Glu mutation. *A. Colima[1,2], M.T. Magañn[2], T. Hernández[1,2], A. Vázquez[3], J.R González[2], N. Solís[4].* 1) CENTRO UNIVERSITARIO DE CIENCIAS DE LA SALUD, UNIVERSIDAD DE GUADALAJARA,GUADALAJARA, JALISCO,MÉXICO; 2) CENTRO DE INVESTIGACIÓN BIOMEDICA DE OCCIDENTE,IMSS, GUADALAJARA, JALISCO, MÉXICO; 3) UNIVERSIDAD AUTÓNOMA DE GUADALAJARA, GUADALAJARA, JALISCO, MÉXICO; 4) HOSPITAL GENERAL DE LEÓN, LEÓN, GUANAJUATO, MÉXICO.

Lipoprotein lipase (LPL) deficiency (MIM 238600) is an autosomal recessive inherited disorder characterized by a severe hypertriglyceridemia due to an impaired catabolism of triglycerides and an accumulation of triglyceride-rich lipoproteins. LPL deficiency is caused by mutations in *LPL* gene; to date more than 180 LPL-deficiency mutations have been reported. We studied a 6-years old girl from Guanajuato, México with biochemical and clinical characteristics of severe hypertriglyceridemia. The patient has handled very high triglyceride levels which range from 1937mg/dl to 5987mg/dl, biochemical analyses also showed low levels of LDL and HDL (30mg/dl and 10mg/dl respectively). Physical examination of the patient showed the presence of eruptive xanthomas on dorsal hands, she also referred to had anemia, nosebleeds, abdominal pain, and pancreatitis which are characteristic of LPL deficiency. The patient is under pharmacological treatment with 200 mg of bezafibrato per day and a fat-restricted diet. Her mother showed slightly elevated triglycerides levels (209 mg/dl) while her father showed high triglycerides levels (356mg/dl) and low HDL levels (31mg/dl). DNA was obtained with the CTAB-DTAB method for a molecular analysis. The primer pairs to PCR were design using the Oligo 6 software. Mutational screening of *LPL* gene was performed by direct sequencing using an ABI PRISM310 genetic analyzer. The patient was homozygous for the missense mutation in exon 5 which produces an amino acid substitution of glycine for glutamic acid in the codon 188 (c.563G>A, P.Gly188Glu, rs118204057) which has been previously reported to cause LPL deficiency. Her parents, who are second cousins, were carriers of this mutation. No other variants were identified in the other exons. Molecular analyses was performed in other 16 family members, of which eleven were adults and five were minors, nine of them were carriers of the mutation (seven adults and two minors). Biochemical analyses showed that only one of them had elevated triglyceride levels (259mg/dl), the rest of them were normal. To our knowledge this is the first case of LPL deficiency reported in Mexico.

2282T

A deletion mutation of the *POLD1* gene in a Japanese woman with congenital partial lipodystrophy, mandibular hypoplasia, sensorineural deafness and short stature identified by whole-exome sequencing analysis. *K. Yanagi[1], H. Sasaki[2], K. Ohokubo[3], K. Kobayashi[4], Y. Tajiri[5], M. Iso[1], Y. Matsui[1], Y. Kuroki[1], H. Ogata[6], K. Nakabayashi[8], K. Okamura[7], K. Hata[6], Y. Matsubara[8], T. Kaname[1], Initiative on Rare and Undiagnosed Diseases in Pediatrics (IRUD-P).* 1) Dept. of Genome Medicine, Center for National Research Institute for Child Health and Development; 2) Counseling and Health Center, Kyushu University; 3) Dept. of Laboratory Medicine, Fukuoka University School of Medicine; 4) Dept. of Endocrinology and Diabetes Mellitus, Fukuoka University Chikushi Hospital; 5) Dev. of Endocrinology and Metabolism, Kurume University School of Medicine; 6) Department of Maternal–Fetal Biology, Center for National Research Institute for Child Health and Development; 7) Department of Systems Biomedicine, Center for National Research Institute for Child Health and Development; 8) Center for National Research Institute for Child Health and Development.

A 44-year-old Japanese female has been described with congenital partial lipodystrophy (CPL), mandibular hypoplasia, sensorineural deafness, and short stature including the progeroid features of unknown genetic origin. She was born to non-consanguineous parents at full term gestation (43 weeks) with birth weight of 3650 g. Fat-loss on the limbs and growth retardation were noticed around the age of 12 month by her parents. At the age of 14 years, she was admitted to the Fukuoka University Hospital due to short stature and muscle cramp of her lower extremities several occasions. She had progeroid feature with symmetrical fat loss (predominantly her face and extremities, but excessive accumulation of visceral fat), skin atrophy, and short stature (between -2SD and -3DS). Sparse hair, mandibular hypoplasia with crowded teeth and bilateral hearing loss were presented. Tissue calcification of Achilles tendon and juvenile cataract, however, were not observed. She was also affected with impaired glucose tolerance (IGT) or diabetic pattern on O-GTT with hyperinsulinemia and severe insulin resistance. Low plasma leptin and adiponectin were observed. Pituitary-adrenal axis including GH and IGF-1 was within normal. Whole exome sequencing (WES) analysis was performed in herself and family members using Human All Exon V6 kit (Agilent) and HiSeq 2500 (illumina). Then, a de novo mutation in the *POLD1* gene (NM_001256849: c.1812_1814delCTC, p.Ser605del, affecting the polymerase active site) was found in this patient, which was also confirmed by Sanger sequencing. The mutation was not found in in-house data of exome sequencing in Japanese individuals. No pathogenic mutations were noted in this case as known to the lipodystorophy-related genes, such as *AGPAT2, BSCL2, CAV1, CIDEC, FBN1, LIPE, LMNA, NK, PIK3R1, PLIN1, PPARG, PSMB8, RTRF, ZMPSTE24* and *WRN*. Thus, according to the both unique clinical manifestations and the results of WES analysis, this case was diagnosed so-called "MDPL syndrome (MIM#15381)" associated with mandibular hypoplasia, deafness, progeroid features and lipodystrophy, which was recently described as a novel subtype or new causative syndrome of CPL or progeroid syndrome. MDPL syndrome has been reported as autosomal dominant disorder with 13 various ethnic patients to date. The mutation, p.Ser605del, found in our Japanese case is also found in 11 cases out of these patients. The mutation might be a hotspot of MDPL syndrome in worldwide without ethnicity.

2283F

Familial hypercholesterolemia-associated genetic variants among a cohort of 50,726 patients within a single healthcare provider organization. *N.S. Abul-Husn[1], K. Manickam[2], L.K. Jones[2], E.A. Wright[2], D.N. Hartzel[2], C. Gonzaga-Jauregui[1], C. O'Dushlaine[1], J.B. Leader[2], H.L. Kirchner[2], D.M. Lindbuchler[2], M.L. Barr[2], M.A. Giovanni[2], J.D. Overton[1], J.G. Reid[1], R.P.R. Metpally[2], A.H. Wardeh[2], I.B. Borecki[1], A. Baras[1], A.R. Shuldiner[1], O. Gottesman[1], D.H. Ledbetter[2], D.J. Carey[2], F.E. Dewey[1], M.F. Murray[2], Geisinger-Regeneron DiscovEHR Collaboration.* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health Center, Danville, PA.

Familial hypercholesterolemia (FH) is a public health genomics priority, but remains under-diagnosed and under-treated despite widespread cholesterol screening. This represents a missed opportunity to prevent FH-associated cardiovascular morbidity and mortality. Pathogenic variants in three genes (*LDLR*, *APOB*, and *PCSK9*) account for the majority of FH cases. We assessed the prevalence and clinical impact of FH variants in 50,726 individuals from the MyCode® Community Health Initiative at Geisinger Health System who underwent exome sequencing as part of the DiscovEHR human genetics collaboration with the Regeneron Genetics Center. This is the first large-scale study utilizing exome sequencing for the evaluation of FH in a U.S. Healthcare Provider Organization. Thirty-five known and predicted pathogenic variants in *LDLR*, *APOB*, and *PCSK9* were identified in 229 individuals, corresponding to a carrier frequency of 1:222. Surprisingly, FH variants were found to explain only 2.5% of severe hypercholesterolemia (LDL-C ≥ 190 mg/dl) in the cohort. Applying Dutch Lipid Clinic Network diagnostic criteria to electronic health record (EHR) data, we found criteria supporting a clinical diagnosis of FH in 62% of FH variant carriers. We estimated that 14 of the 35 FH variants might have reduced penetrance, as not all carriers of these variants met pre-sequencing diagnostic criteria for FH despite having LDL-C data available in their EHRs. Active statin use was identified in 57% of FH variant carriers; however, only 33% had a most recent LDL cholesterol level below 100 mg/dl. FH variant carriers had significantly increased odds of general and premature coronary artery disease (CAD; OR = 2.62, p = 4.3x10^{-11}, and OR = 3.72, p = 5.5x10^{-14}, respectively). The increased odds of premature CAD was greatest in carriers of *LDLR* predicted loss-of-function variants (OR = 10.29, p = 9.8x10^{-19}) and smallest in carriers of *APOB* variants (OR = 1.74, p = 0.12), demonstrating the potential value of molecular sub-classification of disease. In summary, we show that large-scale genomic screening in patients with longitudinal EHR data has the ability to detect FH, uncover and characterize novel pathogenic variants, determine disease prevalence, and enhance overall knowledge of clinical impact and outcomes. Our findings highlight the under-treatment of FH, and demonstrate a potential clinical benefit for the large-scale sequencing planned by the national Precision Medicine Initiative.

2284W

Expanding the spectrum of novel and founder mutation(s) in children affected with Tay-Sachs disease from India. *M.A. Mistri[1], P.M. Tamhankar[2], C. Datar[3], M. Kamate[4], S. Idicula-Thomas[5], S. Nampoothiri[6], M. Kabra[7], R. Puri[8], K.M. Girisha[9], S. Mehta[10], D. Solanki[11], F. Sheth[1], S. Gupta[12], J. Sheth[1].* 1) Biochemical and Molecular Genetics, FRIGE's Institute of Human Genetics, Ahmedabad, Gujarat-India, India; 2) ICMR Genetic Research Center, National Institute for Research in Reproductive Health, Mumbai, Maharashtra, India; 3) Department of Genetics, Clinical Geneticist, Sahyadri Medical Genetics and Tissue Engineering Facility (SMGTEF), Pune, India; 4) Department of Pediatric Neurology, KLES Prabhakar Kore Hospital, Belgaum, Karnataka, India; 5) Biomedical Informatics Center, National Institute for Research in Reproductive Health, Mumbai, Maharashtra, India; 6) Department of Pediatric Genetics, Amrita Institute of Medical Science & Research Centre, AIMS Ponekkara PO, Cochin, Kerala, India; 7) Genetics Unit, Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India; 8) Center of Medical Genetics, Sir Ganga Ram Hospital, New Delhi, India; 9) Department of Medical Genetics, Kasturba Medical College, Manipal University, Manipal, Karnataka, India; 10) Usha-deep Hospital, Naranpura, Ahmedabad, Gujarat, India; 11) Mantra Child Neurology & Epilepsy Clinic, Bhavnagar, Gujarat, India; 12) Department of Biochemistry, M.S. University, Vadodara, Gujarat, India.

Tay-Sachs disease (TSD) is an autosomal recessive neurodegenerative disorder caused by impaired activity of the lysosomal enzyme β-Hexosaminidase-A (EC 3.2.1.52) due to the mutation in *HEXA* gene. Although, the mutation spectrum for TSD in India has been partly addressed by us, present study is to further understand the molecular heterogeneity of the disease in Indian population. Present study report the molecular characterization performed on 21 unrelated affected patients with classical TSD. Complete sequencing followed by MLPA analysis revealed 17 variants, 9 of which were novel, including 3 missense variants [V206L, Y213H, G454R], 1 nonsense variant R420X, 1 splice site variant c.460-1G>A, 1 small deletion c.1349delC (A450VfsX3), 2 large deletion [Exon 2-3 deletion & Exon 1 deletion] and 1 heterozygous duplication of exon-1 with unknown second allele. Previously known 5 missense variants [M1T, R170Q, D322N, E462V, R499C], 2 splice site variants [c.459+4A>C, c.1073+1G>A] and 4bp insertion [c.1278insTATC (Y427IfsX5)] were also identified. In 2 cases with clinically suspected TSD; due to normal enzyme level activator protein deficiency was suspected. Molecular analysis of *GM2A* gene was also carried out in 2 patients and has revealed 2 novel homozygous variants including 1 nonsense E158X and 1 splice site c.243-2A>T variants. All novel missense variants of *HEXA* & *GM2A* gene were shown to be pathogenic by SIFT, Polyphen2 and Mutation taster algorithms while splice site variants were confirmed by Mutation taster and NNSplice 0.9 algorithms. Protein homology modeling study was carried out to further establish the effect of novel variants. The sequence alignment of *HEXA* & *GM2A* genes from various species reveals that all residues are highly conserved. The E462V variant was present either in homozygosity or as compound heterozygote, in 5 (~25%) unrelated patients further reconfirm this as a founder mutation in our population. We have also demonstrated that all *HEXA* mutations for Indian patients are associated with severe phenotypes and pseudo-deficient or adult onset alleles are highly unlikely in patients from India. Moreover, 2 cases have also shown deficiency in GM2 activator protein which is very rare and only 9 cases so far reported in the world literature. It can be concluded from these cases that children clinically suspected with TSD having normal β-Hexosaminidase (Hex-A & total-Hex) study should be further investigated for GM2 activator protein deficiency.

2285T

A very rare complication of sitosterolemia. *E.E. Conboy[1], F. Bazerbachi[2], K.D. Watt[3], P.S. Kamath[3], D. Babovic-Vuksanovic[1].* 1) Clinical Genomics, Mayo Clinic, Rochester, MN; 2) Gastroenterology, Mayo Clinic, Rochester, MN; 3) Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN.

Sitosterolemia is a rare autosomal recessive metabolic disease that was first described in 1974 by Bhattacharyya and Connor who reported on two intellectually normal sisters with tendinous and tuberous xanthomas and elevation of blood levels of β-sitosterol, campesterol and stigmasterol. The STSL locus was mapped to chromosome 2p21 in 1998. Soon afterward, mutations in *ABCG5* and *ABCG8* were found to be genetic cause of sitosterolemia. These two genes encode for (ATP)-binding cassette (ABC) transporters sterolin-1 and sterolin-2, which are sterol transporters that normally cooperate to limit intestinal absorption by resecreting plant sterols back into the glut lumen and to promote biliary excretion of plant sterols. In healthy individuals, almost all of the small percent of absorbed plant sterols are quickly secreted into the bile so that they are left with only trace amounts of these phytosterols in the blood. Altered forms of these transporters predispose to sterol accumulation, allowing patients to absorb a much larger percentage of ingested sterols, and the transporter dysfunction in the liver allows these patients to only secrete a fraction of these absorbed sterols back into the bile. This results in various complications of increased phytosterol levels, including premature coronary artery disease, tendon xanthomas, hematologic abnormalities including hemolytic anemia and macrothrombocytopenia, as well as a less well-known complication of liver cirrhosis. We review a case of a 21 year old man who presents with bicytopenia, liver failure and premature atherosclerosis and evaluation for liver and heart transplant. The underlying cause for his organ dysfunction was undiagnosed until futher workup revealed a sitosterol level of 174.5 (normal 0-5 mg/L). Genetic testing was sent and he was found to have a homozygous pathogenic mutation in ABCG5, c.1336C>T and p.Arg446*. Prior to this case, there has only been one other known case of a patient with sitosterolemia suffering from liver cirrhosis. It is estimated that there are around 100 reports of patients with sitosterolemia worldwide, however, given that the usual test for plasma cholesterol concentration does not measure phytosterols it is likely an under-recognized entity.

2286F

Molecular follow-up of a newborn screening (NBS)-positive case of medium-chain acyl-coenzyme A dehydrogenase deficiency (MCADD) identified two *ACADM* variants: Are they clinically pathogenic? *Y. Ho, C. Morgan, B. Johnson.* Clinical Genomics, Invitae Corp, San Francisco, CA.

MCADD is one of the most common recessively inherited metabolic diseases. It is caused by pathogenic mutations in the ACADM gene that affect the β-oxidation of fatty acids 6–12 carbons long. Affected individuals generate insufficient ketones for energy during periods of glucose and glycogen depletion and can experience extreme hypoglycemia. Untreated episodes can lead to coma and death. Adverse outcomes can be prevented by avoiding metabolic stress and following simple dietary interventions.We report an NBS-positive infant with confirmatory testing results revealing continued elevation of C6, C8, C10, C10:1, and elevated C8/C10 ratio. Next-generation sequencing identified two ACADM variants: c.985A>G (p.Lys329Glu) and c.797A>G (p.Asp266Gly). The pathogenic p.Lys329Glu mutation—most prevalent in patients with clinically symptomatic MCADD— is reportedly as high as 80% homozygous but is considerably less frequently observed in individuals identified through NBS (63%/47% homozygous). The p.Asp266Gly variant, currently classified as a variant of uncertain significance, has been previously reported in newborns with biochemical markers consistent with MCADD and shown to cause partial enzymatic deficiency; however, it has not been reported in clinically symptomatic patients.The highly prevalent pathogenic p.Lys329Glu variant confers high risk for metabolic decompensation, although unaffected homozygous adults have been reported. Conversely, whereas the p.Asp266Gly variant has been identified in MCADD-positive NBS, it causes only partial deficiency of the enzyme and, to the best of our knowledge, has not been reported in clinically symptomatic patients. In addition, although this genotype has been previously described in one NBS-positive MCADD case, the long-term outcome is unknown.Our patient showcases the uncertain reality patients and clinicians often face when dealing with a rare disease like MCADD for which NBS positivity with confirmatory results does not necessarily predict clinical consequences. Emerging evidence supports an MCADD genotype–phenotype correlation, although further substantiation of this relationship is needed. We share this report with the goal of contributing to MCADD genotype and phenotype data collection, which is essential for understanding the natural history and more accurately projecting the long-term clinical consequences of MCADD.

2287W

Regulation of HDL-c levels via ABCA1: A role for *LRP1.* *F. Oldoni[1], J. van Capelleveen[2], N. Dalila[3], M.M. Motazacker[2], J. Heeren[4], R.J. Sinke[5], G. Dall-inga-Thie[2], D. Hui[6], R. Frikke-Schmidt[3,7], A. Tybjærg-Hansen[3,7,8], K. Hovingh[2], J.A. Kuivenhoven[1].* 1) Department of Molecular Genetics, University Medical Centre Groningen, Groningen, The Netherlands; 2) Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands; 3) Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 4) Department of Biochemistry and Molecular Cell Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 5) Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands; 6) Department of Pathology and Laboratory Medicine, Metabolic Diseases Institute, University of Cincinnati College of Medicine, Cincinnati, Ohio; 7) Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 8) Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

Background: The genetic architecture of HDL-c deficiencies (HDLD) is well-acknowledged through rare mutations in *ABCA1* (MIM 600046)*, APOA1* (MIM 107680) and *LCAT* (MIM 606967). However, recent studies in liver-specific *Lrp1* (LDL receptor related protein 1) knock-out mice have indicated that Lrp1 also modulates plasma HDL-c (high-density lipoprotein cholesterol) levels. So far, a role for LRP1 in human HDL metabolism has not yet been elucidated. This study focuses on the molecular mechanism underlying a low HDL cholesterol phenotype in heterozygotes for *LRP1* mutations. **Methods:** Targeted-sequencing efforts in individuals with extremely low HDL-c levels (<0.6mmol/L or 23.2mg/dL) identified the first two missense mutations in *LRP1*, c.9730G>A (p.Val3244Ile) and c.11949G>T (p.Glu3983Asp). The *LRP1* locus was found to be strongly associated with HDL-c and ApoA1 levels in a large Danish population cohort (The Copenhagen City Heart Study) suggesting its potential involvement in HDL metabolism. Both *in-vitro* (overexpression and CRISPR-Cas9 deficient cell lines) and *ex-vivo* (patient-derived fibroblasts) approaches have been employed to evaluate the mutations functional consequences and study the molecular mechanisms regulating ABCA1 plasma membrane localization. **Results and Conclusions:** Mutant LRP1 proteins are significantly reduced by 40-60% and less stable when compared to wild-type (WT). Notably, ABCA1 cell surface localization has been shown to be significantly decreased by 60-70% in carrier fibroblasts and LRP1-deficient cells. Molecular evidence supports the notion that alteration of both lysosomal enzyme Cathepsin D and Prosaposin levels affect ABCA1 plasma membrane presence. In addition, phenotype rescue studies in knock-out cells enabled the full recovery of ABCA1 cell surface expression upon introduction of the WT form of LRP1 but not the mutant. The use of extreme genetics has led to the identification of the first LRP1 deficient individuals featuring with extremely low HDL-c levels. Our data suggest that in humans as in mice, LRP1 affects intracellular trafficking of ABCA1 to the plasma membrane, which could explain low HDL-c in our study subjects.

2288T

Niemann-Pick disease type C1: A case with an unreported mutation.
R.J. Sandoval-Muñiz[1], A. Cordero[2], R. Cruz[2], E. López[2], E. Rivera[2], A. Rea[3], J. Hernández[3], Y. Nakashima[3], R. Nieto[3], E. Zapata[3], G. González[2], J.R. Corona[1,3], L. Arnaud-López[1,3]. 1) Doctorado en Genética Humana, Universidad de Guadalajara, Genetics Service, Pediatric Division of Hospital de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, Mexico; 2) Hospital Civil de Guadalajara "Dr. Fray Antonio Alcalde", Hospital No. 278, Col. El Retiro, Guadalajara, Jalisco, México; 3) Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, Salvador de Quevedo y Zubieta No. 750, Col. La Perla, Zip Code 44340.

Niemann-Pick disease type C1 (NPDC1; MIM 257220) is an autosomal recessive lysosomal storage disease. According to its pathophysiology it is classified into types A, B, C and D. NPDC1 despite a highly variable clinical spectrum, spleen, liver and brain are affected. Mutations in gene *NPC1* mapped on chromosome 18q11.2, that predicts a protein of 1278 amino acids trigger the disease. This protein acts as a transmembrane efflux pump apparently causing inadequate exogenous cholesterol esterification. Case report:A 28 months-old boy, young parents with no consanguinity or know medical problems. He started his conditions at 5 months of age with delayed acquisition of skills and regression after 8 months, at 27 months presented spasticity and arthralgia. Physical examination: W: 11.4 kg (-1.65 SD), H: 90 cm (+0.28 SD), CC: 47.9 cm (-1.07 SD). Irritable, clinical dystonia, spasticity generalized, hyperreflexia, muscular atrophy, *pectus excavatum*, bloating by hepatosplenomegaly, no ascites. Index of diagnostic suspicion: 88. Hypochromic-microcytic anemia (HG 10.3 g/dL, HTO 35%), thrombocytopenia (95.1 miles/µL), impaired lipid metabolism (TGL 203, LDL 30.4 and VLDL 41 mg/dL) and increased liver enzymes (TGO 93.8 IU/L). Peripheral blood smear showed microcytosis, poikilocytosis, hypochromia, normal platelets. Bone marrow aspirate: increased red and megakaryocytic series, white series with large, round, eccentric nucleus and foam cells. Radiology: PA thoracolumbar presented hepatosplenomegaly. Abdominal ultrasonography and liver elastography exhibited hepatosplenomegaly that suggesting storage disease. Mechanics swallowing: absent suction and trachea-esophageal reflux. Cranium NMR: cortico-subcortical atrophy. Molecular study: homozygous mutation of the gene *NPC1*, c.3734C>T(p.P1245L). NPC-509 1.7 ng/mL and chitotriosidase 2500 nmol/h/mL. Conclusions: NPDC1 is a lysosomal hoarding, characterized by storage of sphingosine, glycosphingolipids and cholesterol with a decreased of lysosomal calcium wich leads to neurovisceral damage and innate inmune dysfunction. Index suspicion helps to identify patients with visceral, neurological and psychiatric clinical data. Presence of foam cells are no considered pathognomonic diagnosis and may be required skin biopsy as a complement diagnostic with inmunohistochemical staining. *NPC1* gene homozygous mutation c.3734C>T (p.P1245L) previously unreported and correspondsto the classification 1 of the American College of Medical Genetics.

2289F

Phenotypic characterization of cell lines from patients with spondylometaphyseal dysplasia with cone-rod dystrophy (SMD-CRD). *J. Jurgens, N. Sobreira, A. Franca, J. Hoover-Fong, D. Valle.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

   SMD-CRD is a rare, autosomal recessive disorder caused by partial or complete loss-of-function variants in *PCYT1A* [Hoover-Fong et al., 2014; Yamamoto et al, 2014]. Clinical characteristics include progressive, early-onset photoreceptor degeneration as well as short stature, bowing of the long bones, metaphyseal flaring, rhizomelic shortening, platyspondyly, and scoliosis. *PCYT1A* encodes CTP:phosphocholine cytidylyltransferase α (CCTα), an enzyme which catalyzes the rate-limiting step in *de novo* synthesis of phosphatidylcholine (PC), the predominant membrane phospholipid in mammalian cells. Here we describe the phenotypic characterization of cultured skin fibroblasts from SMD-CRD patients. Western blot analysis of CCTα expression in cultured fibroblasts of 5 SMD-CRD patients (4 probands and an affected sib) showed variable levels of CCTα. When compared to controls, Ala99Val homozygotes had a 30% reduction in protein levels; Glu129Lys homozygotes had a 70% reduction; and patients homozygous for a 1-bp frameshifting insertion (Ser323Argfs*38) had no detectable CCTα. To assess PC synthesis, we measured incorporation of [³H]-choline into PC in intact fibroblasts over time. As compared to controls, fibroblasts homozygous for the Ser323Argfs*38 allele were severely deficient in their ability to convert [³H]-choline into PC. This defect was exacerbated by pre-incubation (10 min at 400 µM) of cells with oleate, a known activator of CCTα. Oleate pre-incubation promoted an ~ 4-fold increase in PC synthesis in control cells with no effect on SMD-CRD patient cells (Ser323Argfs*38 homozygote). Using cells from Drosophila and mice, Guo et al. (2008) and Krahmer et al. (2011) demonstrated that CCTα-deficient cells formed fewer and larger lipid droplets than did control cells after oleate exposure. We hypothesized that SMD-CRD patient fibroblast cells would also exhibit abnormal lipid droplet accumulation in response to oleate. Interesting, we found that SMD-CRD fibroblasts (Ser323Argfs*38 homozygote) failed to show significant differences in lipid droplet numbers or sizes as compared to control fibroblasts. We conclude that SMD-CRD patient fibroblasts display variable levels of reduction in PC synthesis and CCTα protein but, surprisingly, have no accompanying lipid droplet abnormalities.

2290W

Mutations in the tyrosine-protein kinase *Lyn* cause an early-onset neutrophilic vasculitis syndrome: Report of two unrelated cases. *A. Almeida de Jesus[1], G. Montealegre[1], H. Freeman[2], N. Martin[3], E. Omoyinmi[4], B. Marrero[1], K. Calvo[5], R. Chyi-Chia Lee[6], A. Brundidge[7], D. Kleiner[8], S. Hewitt[6], D. Chapelle[7], Y. Huang[1], S. Brooks[7], S. Moir[8], E. Maffre[9], P. Brogan[4], M. Merchant[6], Z. Deng[7], R. Goldbach-Mansky[1].* 1) Translational Autoinflammatory Disease Section, NIAID, NIH, Bethesda, MD, USA; 2) Raigmore Hospital, Inverness, Scotland; 3) Royal Hospital for Children, Glasgow, Scotland; 4) University College London Institute of Child Health, London, UK; 5) Hematology Service, Department of Laboratory Medicine, NIH, Bethesda, MD, USA; 6) National Cancer Institute (NCI), NIH, Bethesda; 7) National Institute of Arthritis Musculoskeletal and Skin Diseases (NIAMS), Bethesda, MD, USA;; 8) Immunopathogenesis Section, NIAID, NIH; 9) Yale University, New Haven, CT.

   Statement of purpose: Tyrosine-protein kinase Lyn is a Src-family tyrosine kinase. Lyn[up/up] mice have a gain-of-function mutation at the tyrosine position 508 (p.Y508F) and present with severe anemia, autoimmune glomerulonephritis and positive ANA. We describe two patients with a *de novo* germline mutation in *LYN*, a gene not previously associated with a Mendelian human disease. **Methods** WES was performed using human exome capture Agilent V5 enrichment kit and Illumina HiSeq2000 platform. **Case Reports:** Patient 1 is a 5 year-old Caucasian male born at 31 weeks' gestation who presented with hydrops fetalis and post partum developed a purpuric skin rash, hepatosplenomegaly, periorbital erythema and testicular pain. Laboratory exams showed increased C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), anemia, thrombocytopenia and increased liver enzymes. At 9 months of age a splenectomy was performed and he developed leukocytosis and thrombocytosis while the anemia persisted. A liver biopsy showed a periportal lymphocytic infiltrate and vanishing bile duct disease. He had vasculitis on a skin biopsy and circulating autoantibodies. We identified a *de novo* mutation in a candidate gene (*LYN*) by WES that was confirmed by Sanger sequencing. The mutation *LYN* NM_002350 c.1524C>G leads to a premature stop-codon at p.Y508*. B-lymphocytes showed a constitutive phosphorylation of Lyn and a diminished frequency of immature B cell populations in peripheral blood and in bone marrow. The tyrosine kinase inhibitor dasatinib was initiated and patient responded with reduction of steroid dose, lower frequency of disease flares, resolution of anemia, reduction in total lymphocyte count and normalization in liver enzymes. Patient 2 is 14 year-old Scottish male who presented with an erythematous skin rash on the second day of life. At the age of 3 months, he developed a purpuric skin rash and fevers, abdominal and testicular pain, diarrhea, periorbital erythema and increased CRP and ESR. The recurrent symptoms have responded to oral steroids. He was recently started on etanercept with a good preliminary clinical response. Targeted gene-panel sequencing revealed a *LYN* NM_002350 c.1523A>T, p.Y508F mutation in *LYN*, which was confirmed to be de novo by Sanger sequencing. **Conclusion:** We have used WES and targeted deep sequencing to identify activating mutations in Lyn kinase gene (*LYN*) in two patients with neonatal-onset of fever, small vessel vasculitis and systemic inflammation.   .

2291T

Effect of alpha-lipoic acid and curcumin against tumor necrosis factor-alpha levels in lymphocytes cultured from Fanconi Anemia. *E. Torres-Anguiano[2,3], A. Corona-Rivera[1,2], R. Silva-Cruz[2], U. Santana-Bejarano[2,3], A. Vázquez-Reyes[2,3], J. Roman-Corona[4], L. Bobadilla-Morales[1,2].* 1) Unidad de Citogenética, Servicio de Hemato-Oncología Pediátrica, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jal., México; 2) Laboratorio. de Citogenética Genotoxicidad y Biomonitoreo, Instituto de Genética Humana "Dr. Enrique Corona Rivera" CUCS, Guadalajara, Jal., México; 3) Doctorado en Ciencias en Biología Molecular en Medicina, Universidad de Guadalajara, CUCS, Guadalajara, Jal., México; 4) Servicio de Genética, división de Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jal., México.

Fanconi anemia (FA) is an autosomal recessive genetic disorder characterized by chromosomal fragility, progressive bone marrow failure, somatic abnormalities and susceptibility to cancer. FA is caused by the disruption of FA-BRCA pathway, 19 human genes have been implicated in the causation of disease. The patients with FA are often found overproducing TNFα, which may directly affect hematopoietic stem cell (HSC) function by impairing HSC survival and proliferation, or indirectly change the bone marrow microenvironment critical for HSC homeostasis and function, therefore contributing to disease progression in FA. The aim of the present work is to evaluate the effect of curcumin and alpha lipoic acid on the TNFα levels in lymphocyte cultures from Fanconi anemia patient, with exons 12 to 18 deletions mutation of the *FANCA* gene. From FA patient 6 ml peripheral blood were collected, and then lymphocytes were isolated by Lymphoprep density gradients and cultured in RPMI 1640 medium containing 15% fetal calf serum, antibiotics (10,000 units/ml of penicillin and 10,000 µg/ml of streptomycin) and L-glutamine. Incubate at 37°C with 5% CO2 atmosphere, for 72 h. After 24 h incubation, the cells were treated with α-lipoic acid (80µM) and curcumin (25µM). TNFα levels were determined using Human TNFα ELISA Kit; all measurements were made in triplicate. Statistical comparison among groups was estimated using one-way ANOVA. Lymphocyte cultures without antioxidant treatment the TNFα level was 91.1pg / ml, cultures treatment with curcumin was 52.1 pg/ml TNFα, the cultures with treatment alpha lipoic acid was 81.4 pg / ml TNFα and cultures treatment with the combined α-lipoic acid plus curcumin was 70.5 pg/ml TNFα. The lymphocyte cultures treatment with curcumin and the combined α-lipoic acid plus curcumin showed a significant reduction in the TNFα levels compared to Lymphocyte cultures without antioxidant treatment (p<0.05). Globally, the percentage of reduction the TNFα in lymphocyte cultures treated with curcumin was 42% (p<0.05) and with the combined α-lipoic acid plus curcumin was 22.6% (p<0.05). The bone marrow failure and development of cancer in Fanconi anemia is associated with inflammation, experimental and clinical studies focus future prospects of therapies to finding compounds capable of reducing these characteristics. The results suggest that curcumin can be an effective antioxidant prophylactic and reduces the inflammatory process present in Fanconi anemia.

2292F

Glanzmann thrombasthenia in South Asia: Molecular analysis and identification of novel mutations. *A. Haghighi[1], M. Borhany[2], A. Ghazi[3], N. Edwards[4], A. Tabaksert[4], A. Haghighi[5], N. Fatima[2], T.S. Shamsi[2], J.A. Sayer[4].* 1) Harvard Medical School, Boston, MA; 2) Department of Hematology, Hemostasis & Thrombosis of National Institute of Blood Disease & Bone Marrow Transplantation, Karachi, Pakistan; 3) Chronic Pain Clinic, Wilderman Medicine Professional Corporation, Toronto, Canada; 4) Institute of Genetic Medicine, Newcastle University, Newcastle, UK; 5) Toronto General Hospital, University of Toronto, Toronto, Canada.

Glanzmann thrombasthenia (GT) is an inherited genetic disorder affecting platelets, which is characterized by spontaneous mucocutaneous bleeding and abnormally prolonged bleeding in response to injury or trauma. The underlying defect is failure of platelet aggregation due to qualitative and/or quantitative deficiency of platelet integrin αIIbβ3 resulting from molecular genetic defects in either ITGA2B or ITGB3. Here, we examine a Pakistani cohort of 20 patients with clinical symptoms of GT who underwent laboratory and molecular genetic analysis. In patients with a broad range of disease severity and age of presentation, we identified pathogenic mutations in ITGA2B in 14 patients from 9 different families, including 2 novel homozygous mutations and 1 novel heterozygous mutation. Mutations in ITGB3 were identified in 6 patients from 4 families, two of which were novel homozygous truncating mutations. A molecular genetic diagnosis was established in 11 families with GT, including 5 novel mutations extending the spectrum of mutations in this disease within a region of the world where little is known about the incidence of GT. Mutational analysis is a key component of a complete diagnosis of GT and allows appropriate management and screening of other family members to be performed.

2293W

Sweet's syndrome in a patient with compound heterozygous mutations in the Mediterranean fever gene (*MEFV*). *M. Michelson-Kerman[1,2], C. Vinkler[1,2], D. Lev[1,2].* 1) Inst Med Genetics, Wolfson Medical Ctr, Holon, Israel; 2) Maccabi Health Service, Tel-Aviv, Israel.

Sweet's syndrome (SS) or acute febrile neutrophilic dermatosis is a rare disorder that often occurs in association with other systemic diseases. The disorder is characterized by development of nonpruritic, painful erythematous plaques with pseudovesicles, occasional pustules and rare bullae. SS consists of a triad of erythematous plaques infiltrated by neutrophils in association with fever and leukocytosis. The pathological features of SS involve the dermis. The condition presents in three clinical settings: classic (or idiopathic) SS, malignancy-associated SS and drug-induced SS syndrome. The treatment of choice for SS are systemic corticosteroids, although colchicine and potassium iodide are also considered to be effective for SS. We present an unusual recurrent course of SS in a 38 year old man who carries compound heterogous mutations in the MEFV gene. A 38 year old, generally healthy man from Sephardic Jewish ancestry, had suddenly developed fever, malaise, artralgia and painful erythematous plaques with pustules and bullae on the anterior aspects of the upper extremities. Diagnostic evaluation included moderate leucocytosis, elevated erythrocyte sedimentation and C-reactive protein rate and normal liver functions. antistreptolysin O-titer. Blood cultures and tuberculosis test were negative. Chest radiography was normal. The symptoms exacerbated despite treatment with systemic corticosterids. Clinical improvement appeared after administration of colchicine . Mutational analysis of the MEFV gene revealed compound heterozygous M694V and V726A mutations. Familial Mediterranean fever (FMF) is an autosomal recessive disease characterized by recurrent attacks of fever with serosal inflammation. The FMF gene (MEFV) encodes the protein pyrin that plays an important role in modulating the innate immune response. MEFV mutations have been identified primarily in patients from Mediterranean populations and in Israel the carrier state is as high as1: 5. Sweet's syndrome has been described in a patient with classical FMF as a possible new cutaneous feature and has never been described as a presenting sign of FMF. Although various skin lesions have been described with FMF, erysipelas-like erythema (ELE) has been reported the only pathognomonic cutaneous manifestation. Our patient, carrying compound heterozygous mutations in MEFV presented with Sweet's syndrome. We suggest that SS skin lesions might be an only cutaneous presentation of FMF. .

2294T

Genomic characterization of Diamond-Blackfan anemia reveals known and novel ribosomal causes. *S.A. Savage[1], P. Khincha[1], S.R. Ellis[2], N. Giri[1], S. Brodie[3], S. Chandrasekharappa[4], W. Zhou[3], B.D. Hicks[3], J.F. Boland[3], M. Yeager[3], K. Jones[3], B. Zhu[3], M. Wang[3], B.P. Alter[1], L. Mirabello[1].* 1) Div Cancer Epidemiology & Genetics, Natl Cancer Inst, Bethesda, MD; 2) Department of Biochemistry and Molecular Biology, University of Louisville, Louisville, KY; 3) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD; 4) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, Bethesda, MD.

Diamond-Blackfan anemia (DBA) is a rare inherited bone marrow failure syndrome (IBMFS) characterized by erythroid hypoplasia typically presenting with macrocytic anemia in infancy, often with normal white blood cell and platelet counts. It is associated with a spectrum of congenital anomalies, including flat nasal bridge, high arched or cleft palate, short stature, abnormal thumbs, genitourinary and cardiac defects, and webbed neck. DBA patients are at high risk of AML, MDS, colon adenocarcinoma and osteosarcoma. Many patients with DBA have autosomal dominant germline pathogenic variants or deletions in genes affecting ribosomal biogenesis. Associated genes are part of the small (*RPS19, RPS24, RPS17, RPS7, RPS10, RPS26, RPS27, RPS28, RPS29*) and large (*RPL35A, RPL5, RPL11, RPL26, RPL15, RPL27*) ribosomal subunits. Pathogenic variants in two X-linked genes, *GATA1* and *TSR2,* cause DBA in a minority of patients. However, about 45% of DBA families are yet to have a pathogenic variant in a disease-causing gene identified. 87 DBA families have enrolled in the National Cancer Institute's IBMFS study (NCT00027274) to date. Study participants completed detailed questionnaires, medical records were reviewed, and a subset of families underwent evaluations at the NIH Clinical Center. Clinical testing performed prior to 2014 identified the genetic cause in 26 families. Genetic testing information or DNA samples were not available on 26 families. Blood or buccal cell DNA from 35 families was subjected to whole exome sequencing (WES), targeted deletion analysis and genome-wide assessment of copy number variation to identify the pathogenic variant. The causative genetic change was discovered in 17 of the 35 gene unknown families. Pathogenic variants in known DBA-associated ribosomal genes were present in 15 (43%) families. We discovered 2 novel genes with pathogenic variants in 2 families: *RPL35* (p.K77N) and *RPL18* (p.L51S). These variants were proven to be deleterious by functional assays showing characteristic pre-rRNA processing defects and differential gene expression when compared with wild-type. The genetic cause was identified in 43 (71%) of the 61 families available for testing. The combination of WES and deletion analyses was efficient and led to the discovery of 2 novel causes of DBA, in addition to the 17 known causes. Studies are ongoing to identify the etiology in the families with a yet unknown genetic cause of DBA.

2295F

Identification of the molecular causes of hereditary spherocytosis. *L.J. Drury[1], S. Butchart[2], M. Fantauzzi[2], J. Baker[3], M. Kirby-Allen[3], V. Blanchette[3], J. Pugi[3], J. Langer[4], M. Carcao[3].* 1) PreventionGenetics, Marshfield, WI; 2) Department of Nursing, Hospital for Sick Kids, Toronto, Canada; 3) Division of Haematology/Oncology, Hospital for Sick Kids, Toronto, Canada; 4) Division of General Surgery, Hospital for Sick Kids, Toronto, Canada.

Hereditary spherocytosis (HS) is form of hemolytic anemia that affects 1 in 3,000 individuals. The *ANK1, SPTB, SLC4A1, SPTA1,* and *EPB42* genes encode proteins important for maintaining the membrane cytoskeleton allowing normal red blood cells to adopt a biconcave disk morphology. Mutations in these genes lead to HS which can be inherited in both autosomal dominant and recessive manners. Despite HS being a common form of hemolytic anemia, the mutations resulting in HS are incompletely defined. Here we present genetic findings, including genotype and presumed inheritance patterns, from 80 patients with HS. In collaboration with Sick Kids Hospital (Toronto, Canada), we conducted Next Generation sequencing for the *ANK1, SPTB, SLC4A1, SPTA1,* and *EPB42* gene and characterized variants based on ACMG guidelines. All pathogenic, likely pathogenic, and variants of uncertain significance were verified by Sanger sequencing. Potential molecular causes of HS were identified in 70 of 80 patients with 65 mutations being novel. The majority of mutations characterized were private truncating (nonsense, frameshift, or splicing) defects within the *ANK1* and *SPTB* genes resulting in autosomal dominant forms of the disease. In one case we were able to identify a deletion encompassing exons 18-26 in the *ANK1* gene using Next Generation sequencing data. This deletion was confirmed using end point PCR to characterize the precise breakpoints for the deletion. Interestingly, there were 12 cases where mutations were identified in two genes. With these genes participating in the spectrin membrane anchoring complex, it is possible that corresponding mutations both contribute to disease. In all, understanding the genetic etiology for HS families is important for determining modes of inheritance and aiding in the genetic counseling for prospective parents. Importantly, our results are one of the most comprehensive genetic analyses for an HS cohort to date and will add to the growing mutation databases for these genes.

2296W

An update on the diagnostics, phenotype and treatment of deficiency of ADA2 (DADA2). *N. Deuitch, Q. Zhou, A. Ombrello, D. Stone, T. Romeo, A. Jones, D.I. Kastner, I. Aksentijevich.* Inflammatory Disease Section, National Human Genome Institute, Bethesda, MD.

DADA2 is a recessively inherited disease caused by loss-of-function mutations in the *CECR1* gene, which encodes adenosine deaminase 2 (ADA2). Patients present with fevers, rashes, polyarteritis nodosa (PAN) and/or recurrent ischemic strokes leading to neurological impairment. Although the precise molecular mechanism of this disease is not clear, it appears that in addition to its deaminase activity, ADA2 plays an important role in the development of hematopoietic cells. We performed Sanger sequencing of all 9 exons of *CECR1* in the patients and unaffected family members. RT-PCR and cDNA sequencing was done in patients who carry only one coding mutation in the gene to evaluate the presence of both alleles. We tested serum ADA2 enzyme activity in patients using a modified kit from Diazyme. Since our initial publication we have expanded our cohort to include 39 patients, 33 of whom carry bi-allelic homozygous or compound heterozygous mutations in *CECR1*. The remaining patients carry only a single detectable pathogenic mutation suggesting the presence of non-coding and genomic deletion disease-causing variants. We identified 26 mutations, 16 of which had not previously been associated with DADA2. All variants were rare and reported at low frequencies (<.001) in the ExAC, dbSNP, 1000 Genomes, and ClinSeq databases, consistent with the recessive inheritance. All variants were predicted to be damaging by bioinformatics tools and likely affect catalytic activity and protein stability. The 15 patients tested for ADA2 activity were found to have significantly decreased protein activity compared to healthy age-matched controls and unaffected family members. This biochemical test could be used to support the clinical diagnosis of DADA2, particularly in cases of patients with only one demonstrable disease-causing variant. We have also expanded the spectrum of phenotypes associated with DADA2 to include skin nodules diagnosed as erythema nodosum, digital osteolysis, medium vessel vasculitis, hematuria, ischemic bowel disease and immunodeficiency. Treatment with TNF inhibitors efficiently reduced inflammation. 12 patients had experienced at least one ischemic stroke. Over a collective 1,064 patient months before treatment these 12 patients had a total of 46 strokes. In the 323 post-treatment patient months there have been no strokes. We have established the use of anti-TNF medications as the standard of care for all DADA2 patients in our clinic. .

2297T

Inherited hemoglobin disorders are associated with high serum hepcidin concentrations in Cambodian women of reproductive age. *C.D. Karakochuk[1,2], M.K. Barker[1,2], K. Naguib[1,2], R. Devenish[3], C. Oberkanins[4], H. Puhringer[4], R.E. Aleliunas[2,5], A.M. Devlin[2,5], T.J. Green[6,7].* 1) Human Nutrition, University of British Columbia, Vancouver, British Columbia, Canada; 2) Child and Family Research Institute, Vancouver, British Columbia, Canada; 3) National Pediatric Hospital, Phnom Penh, Cambodia; 4) ViennaLab Diagnostics GmbH, Vienna, Austria; 5) Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada; 6) Pediatrics and Reproductive Health, University of Adelaide, Adelaide, Australia; 7) Division of Healthy Mothers, Babies and Children, South Australian Health and Medical Research Institute, Adelaide, Australia.

   Background: Certain inherited hemoglobin (Hb) disorders are associated with altered iron metabolism. Our aims were to assess the frequency of deletions and/or mutations in the 4 α-genes (*HBA1*, *HBA2*) and 2 β-genes (*HBB*) encoding Hb, and to compare hepcidin (a peptide that regulates iron metabolism) concentrations among women with normal and abnormal Hb genotypes in Cambodia. **Methods:** This analysis includes 263 randomly chosen women of 809 who were participating in a larger trial investigating the causes of anemia from Kampong Chhnang province, Cambodia. To be eligible for the trial, women had to be non-pregnant, between 18-45 y, and have a Hb <117 g/L based on a portable hemoglobinometer. Hb (g/L) was measured on a fasting venous blood sample using an automated hematology analyser (Sysmex XN-1000). Serum hepcidin (nmol/L) was measured using an immunoassay kit (DRG International, USA). Hb typing for Hb variants was assessed using capillary Hb electrophoresis using a MINICAP analyser (Sebia, France). DNA extracted from buffy coat was assessed for 21 α-globin gene deletions and point mutations using a StripAssay® SEA kit (ViennaLab Diagnostics, Austria). **Results:** Mean ± SD age of women was 30 ± 8 y, 66% of women had 1-2 children. Overall, mean ± SD Hb was 118 ± 13 g/L, anemia prevalence (Hb <120 g/L) was 47% (*n*=124/262), and median (IQR) hepcidin was 6.1 (1.7, 11.4) nmol/L. A total of 72% (*n*=189/263) of women had some type of abnormal Hb, with 4 variant Hb types identified by Hb electrophoresis (Hb E, CS, H and F) and 7 α-gene deletions/mutations identified by the α-globin Strip Assay® (-3.7, -4.2, --SEA, Hb CS, Hb QS, Hb PS, and anti -3.7). The Hb E trait (β^{E}β) and α-3.7 trait (-3.7α/αα) were most common (24% and 13%, respectively). Median (IQR) hepcidin concentrations were over twice as high among women with an abnormal Hb genotype (6.7 [2.9, 12.7] nmol/L), as compared to women with a normal Hb (3.1 [0.4, 8.9] nmol/L; p<0.0001). **Interpretation:** We highlight the high frequency and diverse heterogeneity of Hb disorders among Cambodian women. Serum hepcidin concentrations were significantly higher among Cambodian women with an abnormal Hb genotype, as compared to women with normal Hb. When hepcidin levels are high, iron becomes trapped in cells and dietary iron absorption is reduced, which may alter iron utilization and iron availability in the body. This has implications for the accurate measurement and interpretation of iron biomarkers in this population.

2298F

Genetic features of myelodysplastic syndrome and aplastic anemia in pediatric and young adult patients. *S.B. Keel[1], A. Scott[2,3,4], M. Sanchez-Bonilla[5], P.A. Ho[2,3,4], S. Gulsuner[6], C.C. Pritchard[7], J.A. Abkowitz[1], M.C. King[6], T. Walsh[6], A. Shimamura[5].* 1) Medicine, Division of Hematology, University of Washington, Seattle, WA; 2) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Pediatric Hematology/Oncology, Seattle Children's Hospital, Seattle, WA; 4) Pediatrics, University of Washington, Seattle, WA; 5) Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; 6) Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 7) Laboratory Medicine, University of Washington, Seattle, WA.

   Clinical and histopathological distinction between the inherited versus acquired bone marrow failure (BMF) or myelodysplastic syndromes (MDS) is challenging. To determine whether a subset of pediatric and young adults undergoing hematopoietic stem cell transplantation for aplastic anemia (AA) or MDS have germline mutations in BMF/MDS genes, we performed a targeted next-generation genetic screen of peripheral blood DNA samples from children and young adults with AA or MDS obtained from a biorepository of patients transplanted at the Fred Hutchinson Cancer Research Center and retrospective chart review. Pathogenic variants were validated by Sanger sequencing and confirmed as constitutional by sequencing DNA isolated from a second tissue or by chromosome fragility testing of patient-derived lymphoblasts. Compound heterozygous variants were confirmed to be in trans by subcloning and Sanger sequencing. Mutations in inherited BMF/MDS genes were found in 5.1%(5/98) of AA patients and 13.6%(15/110) of MDS patients. While the majority of mutations were constitutional, a *RUNX1* mutation present in the blood at a fifty-percent variant allele fraction was confirmed to be somatically acquired in one MDS patient, highlighting the importance of distinguishing germline versus somatic mutations by sequencing DNA from a second tissue or from parents. Pathologic mutations were present in *DKC1, MPL,* and *TP53* among the AA cohort, and in *FANCA, GATA2, MPL, RTEL1, RUNX1, SBDS, TERT, TINF2,* and *TP53* among the MDS cohort. Family history or physical exam failed to reliably predict the presence of germline mutations. Only 2 of the 5 AA patients with mutations had a physical anomaly and only 1 patient had a family history of cancer or a related phenotype. Among the MDS cohort, only 3 of the 14 patients with mutations had a physical anomaly and while patients with mutations were more likely to have a positive family history of cancer or a related phenotype than those patients without mutations [71%(10/14) vs. 49%(47/96)], the absence of a family history did not exclude the possibility of a mutation. The identification of these inherited disorders can inform donor selection, family counseling, transplant preparatory regiment selection, and surveillance for disease-specific extrahematopoietic complications. Our work suggests that genetic screening to evaluate inherited BMF/MDS genes collectively should be considered in pediatric and young adults presenting with AA and MDS.   .

2299W

Genetic susceptibility to neonatal group B streptococcal disease. *A. Borghesi[1,2,3], S. Asgari[1,2], C.W. Thorball[1,2], N. Chaturvedi[1,2], I. Mazzucchelli[3], M. Stronati[3], L. Schlapbach[4], J. Fellay[1,2].* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Lausanne, Vaud, Select a Country; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Neonatal Intensive Care Unit, San Matteo Hospital, Pavia, Italy; 4) Paediatric Critical Care Research Group, Mater Children's Hospital, Brisbane, Australia.

   Background: group B streptococcus (GBS) or *Streptococcus agalactiae* is the most frequently isolated microorganism in neonatal sepsis. Clinical observation demonstrates that only a small proportion of infants exposed to GBS develops overt disease, suggesting that host determinants play a major role in individual susceptibility. Monogenic disorders causing inborn errors of the innate immune responses to microorganisms are thought to give an important contribution to susceptibility to infectious diseases early in life. **Statement of purpose**: we aim to test the hypothesis that rare or novel genetic variants causing single-gene inborn errors of the protective immunity to GBS account for the individual susceptibility in a subset of patients with neonatal GBS disease. **Methods**: We enrolled otherwise healthy, full-term newborn infants who suffered from life-threatening, late-onset (after 6 days of life) neonatal GBS infection. We performed exome sequencing in 30 patients and their parents when available. After read alignment and variant calling, we ran single variant and gene-based case-control tests, comparing our GBS cases to an in-house control population of 120 individuals. We also prioritized variants by filtering against the ExAC database (keeping only variants with a minor allele frequency < 0.01) and by genetic and functional annotation (keeping only missense or predicted loss-of-function (LOF) variants observed in homozygous or hemizygous state). **Summary of results**: The percentage of the bait regions covered 20x or more was higher than 95% for more than 90% of the samples. Five genes were significantly enriched for potentially deleterious variants in the GBS cohort: *SLC35E2B, OR9G1, MAP2K3, CDK11B, MUC3*. By variant prioritization, we identified 168 LOF or missense variants, including 19 in genes predicted to be involved in primary immunodeficiency by the human gene connectome. Further analysis is needed to explore the *de novo* autosomal dominant and compound heterozygous autosomal recessive genetic models.

2300T

Exome sequencing of 200 patients with primary immunodeficiencies: Translational genomics with direct clinical implications. *P. Arts[1], A. Simons[1], E. Yilmaz[1,2], A.R. Mensenkamp[1], W.A. van Zelst-Stams[1], M. van Deuren[3], A.K. Simon[3], J.A. Veltman[1], F.L. van de Veerdonk[3], M.G. Netea[3], A. Hoischen[1].* 1) Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Medical Biology, Faculty of Medicine, Akdeniz University, Antalya, Turkey; 3) Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, Netherlands.

   Introduction: Primary immunodeficiencies (PIDs) are severe medical conditions caused by non-redundant defects of innate or adaptive immune responses, leading to recurrent, severe infections. Severely affected PID patients with an unknown genetic defect were referred for diagnostic exome-sequencing. Methods: Exome data of 200 PID cases have been analyzed following these criteria: As a first tier, variation in 301 genes known to cause PIDs are studied using an *in silico* filter during data-analysis. In addition, 131 patients without a genetic diagnosis after step 1 gave permission for exome-wide analysis of the data. In selected cases *in vitro* stimulations of mononuclear blood cells identified cytokine production capacity defects and provided additional functional evidence for the diagnosis. Results: Diagnostic exome analysis revealed (likely) pathogenic mutations in known PID genes in 32 cases (19%). Exome-wide analysis led to identification of additional causative mutations in 9 patients, for which underlying genes have a known role in specific immunological sub-pathways. One example with direct clinical implication of this approach was the characterization of a genetic, and IFNγ production defects in a patient with recurrent mucocutaneous HSV infections. For this patient IFNγ replacement therapy successfully prevented infections. Conclusion: Exome sequencing can provide a genetic diagnosis in a significant number of patients with PIDs. It is important to perform exome-wide analyses after an analysis of the known PID genes because many new PID genes are only just being discovered. The molecular diagnosis provided novel insight in therapeutic options for a subset of our patients.

2301F

Mutational spectrum of a series of mexican patients with hereditary angioedema: Report of three novel mutations in *SERPING1* **gene.** *D.E. Cervantes-Barragan[1], M.J. Gaytan Garcia[1], J.I. Navarrete Martinez[1], A. Limón Rojas[2], A. Lupian Sanchez[2], F. Campos Romero[3].* 1) Genetics, Hospital Central Sur de Alta Especialidad, Mexico city, Mexico city, Mexico; 2) General Dictatorate, Hospital Central Sur de Alta Especialidad, Mexico city, Mexico city, Mexico; 3) Allergy, Hospital Central Sur de Alta Especialidad, Mexico city, Mexico city, Mexico.

Hereditary angioedema (HAE) is an autosomal dominant inheritance disease with a worldwide prevalence of 1:50.000, to date there are over 300 known mutations in C1- inhibitor (*SERPING1*) gene that means a decreased levels or funtion in the first component of the complement pathway, increased production of bradikinin and restrict the action of factor XII an Kallikrein. The phenotype is characterized by recurrent attacks of severe swelling of limbs, face, intestinal tract, genitals and airway, tipically without hives, 89 % have prodromal symptons like a non itchy rash, called *erythema marginatum*, flu-like, nausea, fatigue, rumbling, and tingling caused by minor trauma, infections, drugs but often with out an specific trigger. The aim of the study is to report 3 novel mutations in 5 patients with hereditary angioedema. Sequencing and MLPA was performed in the 5 patients. The molecular analysis revealed, two novel deletions c.116delA (p.Asp39Alafs*40) in exon 3, c.773delA(p. Asn258Thrfs*21) in exon 5, and a missense variant c.1328A>C (p.Hys443Pro) in exon 8. This three reports are added to the previous *SERPING1* mutations. Genotype-phenotype correlations are important for prognosis and treatment for these patients. .

2302W

Identification of novel candidate genes for susceptibility to tuberculosis by identifying disease-causing mutations in individuals with PIDs. *E.G. Hoal[1], N. Schlechter[1], M. Möller[1], P.D. van Helden[1], M. Schoeman[1], M. Urban[1], M. Esser[2], A. Franke[3], C.J. Kinnear[1].* 1) Mol Biol & Hum Genetics, Stellenbosch Univ, Tygerberg, Western Cape, South Africa; 2) Pathology, Stellenbosch Univ, Tygerberg, Western Cape, South Africa; 3) Institute for Clinical Molecular Biology, Christian-Albrechts University, Germany.

Although approximately 33% of the world's population is infected with *M.tuberculosis*, the causative agent of tuberculosis (TB), only 10% of infected individuals will develop active disease. While the genotype of invading strain and environmental factors are crucial in disease outcome, host genetic factors can be decisive. Several investigations have successfully identified genes involved in TB susceptibility. Nevertheless, it is certain that more susceptibility genes exist, but identifying them in a complex disease such as TB is challenging. Answers may lie in the genomes of individuals suffering from a group of inherited primary immunodeficiency disorders (PIDs) for which multiple *M.tuberculosis* infections is a common feature. We hypothesize that the genes involved in these PIDs could be candidate genes for increased TB susceptibility in the general population. We aimed to identify novel TB susceptibility genes by finding gene mutations in patients suffering from PIDs characterized by increased TB susceptibility. The exomes of three PID patients as well as their healthy parents, where possible, were sequenced using the Illumina HiSeq. Bioinformatic techniques were used to identify a large number of variations from the reference human genome for each patient. We prioritized the genes based on OMIM and HGMD database entries which resulted in the identification of three novel putative disease-causing variants, one per patient, situated in two novel susceptibility genes namely Mitogen-activated protein kinase kinase kinase 14 (*MAP3K14* [MIM 604655]) and Transporter associated with antigen presentation (*TAP1* [MIM 170260]). Functional studies are currently being done to investigate their involvement in disease, after which case-control association studies will be conducted in a cohort of TB patients and healthy controls. The identification of these disease-causing mutations can provide us with novel candidate genes to screen for TB susceptibility in the general population.

2303T

Mendelian genetic studies of immune disorders to identify novel targets for therapeutic intervention. *J. McElwee[1], J. Maranville[1], J. Hughes[2], Y. Huang[1], M. Jordan[4], S. Snapper[5], W. Dunn[5], J. Ouahed[5], M. Lenardo[7], Y. Zhang[7], H. Su[7], J. Milner[7], J. Cohen[7], H. Matthews[7], J. Lyons[7], G. Sun[7], X. Yu[7], R. Goldbach-Mansky[3], A. Almeida de Jesus[3], Z. Deng[3], Y. Liuv[3], R. Geha[6], J. Chou[6], H. Runz[1].* 1) Genetics and Pharmacogenomics, Merck Research Labs, Boston, MA; 2) Computational Genomics & Genetics, Merck Informatics IT, Boston, MA; 3) National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD; 4) Divisions of Immunobiology and bone marrow transplant and immune deficiency, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 5) Inflammatory Bowel Disease Center, Boston Children's Hospital, Boston, MA; 6) Division of Immunology, Boston Children's Hospital, Boston, MA; 7) National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD.

Mendelian disorders that selectively affect the immune system provide a means to identify and understand new genes and pathways involved in human immunology which could represent new routes for therapeutic intervention for human disease. As part of a collaborative effort to identify novel therapeutic opportunities in immunobiology, Merck Research Labs and a network of investigators from the NIH Clinical Center, Boston Children's Hospital, and Cincinnati Children's Hospital Medical Center, have undertaken focused studies of Mendelian forms of primary immunodeficiencies, syndromic atopic disease, very-early onset inflammatory bowel disease, and severe auto-inflammatory conditions using whole-exome (WES) and/or whole-genome (WGS) sequencing approaches. Thus far, we have sequenced 1445 subjects representing 520 independent kindreds exhibiting a range of suspected Mendelian immune diseases. As part of these studies, we have sequenced 208 samples from 42 kindreds using both WES and WGS. While WGS provides clear technical superiority in a number of areas, we find that it did not lead to substantial improvements in the discovery of causal genes in our cohort for unsolved kindreds previously sequenced using WES. Within our cohort, we have used Mendelian filtering techniques, further supported by extensive molecular and cellular validation studies, to successfully identify causal mutations in known and novel immune genes for over 20% of kindreds that have been analyzed to date. From these studies, we have identified and characterized a number of new genes important for immune function and causal for human immune disease, such as LRBA, CTLA4, PIK3R1, and STAT3. Additional novel causal genes and high-quality candidates in several families are currently being functionally validated, and several novel mechanisms are being explored as therapeutic targets for autoimmune disease. We will present an overview of our recent novel discoveries, a comparison of WES and WGS approaches, and our experience of Mendelian genetic studies in this unique discovery cohort.

2304F

Biallelic missense mutations in *NSMCE3,* **encoding a subunit of the SMC5/6 complex, cause a chromosome breakage syndrome with severe lung disease.** *G. van Haaften[1], S.N. van der Crabben[1], M.P. Hennus[1], G. McGregor[2], D.I. Ritter[3], S.C.S. Nagamani[4,6], O.S. Wells[2], M. Harakalova[1], I.K. Chinn[5,6], A. Alt[2], L. Vondrova[7], R. Hochstenbach[1], J.M. van Montfrans[1], S.W. Terheggen-Lagro[1], S. van Lieshout[1], M.J. van Roosmalen[1], E. Hennekam[1], J.S. Orange[5,6], P.M. van Hasselt[1], W.P. Kloosterman[1], I. Renkens[1], K. Duran[1], I.J. Nijman[1], D.A. Wheeler[3,4], J.J. Palecek[7], A.R. Lehmann[2], A.W. Oliver[2], L.H. Pearl[2], S.E. Plon[3,4,5,6], J.M. Murray[2].* 1) University Medical Center Utrecht, Utrecht, the Netherlands; 2) Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, Brighton, United Kingdom; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston TX, USA; 6) Texas Children's Hospital, Houston TX, USA; 7) Central European Institute of Technology and Faculty of Science, Masaryk University, Brno, Czech Republic.

We describe a new chromosome breakage syndrome associated with severe lung disease in early childhood. Four children from two unrelated kindreds died during infancy of severe pulmonary disease following viral pneumonia with evidence of combined T- and B-cell immunodeficiency. Whole exome sequencing revealed biallelic missense mutations in the *NSMCE3* gene encoding a subunit of the SMC5/6 complex, essential for DNA damage response and chromosome segregation. The *NSMCE3* mutations disrupt interactions within the SMC5/6 complex, leading to destabilization of the complex. Patient cells showed chromosome rearrangements, micronuclei, sensitivity to replication stress and DNA damage, and defective homologous recombination. This study is an example of the power of genematcher.org in uniting researchers studying rare genetic diseases.

2305W

Somatic mosaicism of an intragenic duplication in X-linked *FANCB* observed in a Fanconi anemia patient and his mother, leads to an unusual phenotype. *S.C. Chandrasekharappa[1], A.D. Auerbach[2], F.P. Lach[2], D.C. Kimble[1], F.X. Donovan[1], J. Thomas[1], M. Park[1], P. Chines[1], A. Vlachos[3], A. Smogorzewska[2], R.S. Asur[1].* 1) National Institutes of Health, Bethesda, MD; 2) The Rockefeller University, New York, NY; 3) Hofstra Northwell School of Medicine, New Hyde Park, NY.

Fanconi anemia (FA) is a rare, mostly recessive, disorder characterized by congenital malformations, progressive bone marrow failure, and predisposition to cancer. FA is a genetically and phenotypically heterogeneous disorder, with 19 known FA genes. Patients harboring X-linked *FANCB* mutations usually present with a severe clinical phenotype, resembling VACTERL syndrome with hydrocephalus. We describe here a patient determined to be complementation group FA-B by rescue of crosslink hypersensitivity of patient fibroblasts by expression of *FANCB*, with much milder phenotype (cafe-au-lait spots, failure to thrive, hypopigmented skin patches, tracheomalacia, no hematologic manifestations at age 12 yrs.). The patient exhibited mosaicism detected during the DEB test, a diagnostic for FA, where patient LCLs were resistant to DEB while fibroblasts were sensitive. Employing aCGH and targeted next-gen sequencing, we identified an intragenic duplication in *FANCB*. The 9154 bp duplication included the first coding exon, exon 3, and the flanking intronic regions (chrX:14877976-14887129). Presence of a 4 bp homology (GTAG) at both ends of the breakpoint appears to indicate the duplication may have been mediated via the alt-EJ mechanism. A ddPCR assay was developed for the junction region to evaluate the extent of the duplication. The duplication was present in 90% of the patient fibroblast cell line but in only 5% of the patient LCL. This is consistent with the DEB test, indicating that the allele with duplication has reverted back to wild-type (WT) in LCL and, to a lesser extent, in fibroblast cells. The assay also detected the duplication in DNA from the mother, which also exhibited a similar mosaic reversion in which only 22% of DNA carried the duplication. RT-PCR of RNA from patient fibroblast cells showed the WT transcript, along with an aberrant transcript, 1 kb larger. PacBio sequencing of the larger transcript revealed the aberrant transcript had insertion of the entire exon 3, predicted to introduce a stop codon at the junction. Unlike sequence variants, duplications are difficult to define precisely and quantitate accurately. We demonstrate that detailed characterization of the disease-causing variant is critical for better understanding the patient phenotype.

2306T

Modeling dyskeratosis congenita in the mouse. *C. Keegan[1], K. Bisht[2], J. White[1], J. Chase[3], G. Balbin-Cuesta[3], S. Savage[4], J. Nandakumar[2], I. Maillard[3].* 1) Dept. of Pediatric Genetics, Univ Michigan, Ann Arbor, MI; 2) Dept. of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI; 3) Dept. of Internal Medicine, University of Michigan, Ann Arbor MI; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD.

Dyskeratosis congenita (DC [MIM 127550]) is a telomere biology disorder characterized by hematopoietic failure and classically associated with the clinical triad of oral leukoplakia, dysplastic nails, and pigmentary skin changes. Classic DC is caused by mutations in telomerase genes (*TERC* and *TERT*) and associated proteins (*DKC1*). In recent years, the clinical spectrum of DC has expanded to include rare, severe variants such as Hoyeraal Hreidarsson (HH) syndrome, and more common variants such as aplastic anemia and pulmonary fibrosis that can present during adulthood. Likewise, the genetic spectrum has expanded to include proteins that regulate telomerase trafficking, replication, stability and telomere length. Shelterin is a six-protein complex that regulates access of telomerase to telomeres and protects chromosome ends from unregulated DNA repair. Mutations in two shelterin components, *TINF2* and *ACD*, have been identified in individuals with DC. Although the basic mechanisms by which shelterin protects telomeres have been identified, the pathogenic mechanisms that lead to DC phenotypes remain largely unidentified. The goal of this project is to dissect the in vivo mechanisms leading to expression of the DC phenotype by modeling human DC mutations in the mouse, both by studying the function of hematopoietic stem and progenitor cells expressing specific shelterin protein variants and by engineering these variants into the mouse genome. Our group previously identified biallelic mutations in *ACD* in a family with HH syndrome, c.508_510delAAG (p.K170del) and c.1471A>C (p.P491T), which affect recruitment of telomerase and binding to TIN2, respectively. Using CRISPR/Cas9 genome editing, we have generated a mouse line with a mutation in the orthologous residue (K82) in mice, and we are in the process of generating P491T (P365T) in mice. Homozygous K82del mice are viable and exhibit no identifiable external phenotype. Because mice have longer telomeres than humans, we predict that we will observe a DC-like phenotype in successive generations, and these crosses are in progress. We have also followed the growth of K82del/+ intercross mice and have found that homozygous K82del mice exhibit increased weight gain compared to K82del/+ and +/+ mice beginning at age 2-3 months. Modeling specific human mutations in the mouse will provide a better understanding of the in vivo mechanisms that lead to phenotypes within the DC spectrum. Supported by R01-AG050509.

2307F

Assessment of risk of developing pancreatic risk with ATM gene mutation based on the genetics of hereditary pancreatitis by using saliva: A multi-center evaluation. *S. Singh[1], S. Ghavimi[2], H. Azimi[4,5,8], J. Brown[3,4], A. Razavi[5,7], P. Sealy[2], A. Afsari[2], S. Emmamian[6], A. Asghar Kolahi[9].* 1) Internal Medicine , University of Illinois at Chicago , Chicago , IL; 2) Howard University Hospital, Department of Medicine, Washington, DC; 3) Carleton University, Division of Human Science, Ottawa, Ontario, CA; 4) PsychoGenom, Ottawa, Ontario, CA; 5) Genomics On Call, Washington DC; 6) University of Indiana, Department of Health Sciences and Physics, Bloomingdale, IN; 7) Islamic Azad University of Tehran, Tehran, Iran; 8) University of Ottawa, Ottawa, Ontario, CA; 9) Shaheed Beheshti School of Medicine, Genetic statistician, Tehran, Iran.

   Objective: To evaluate the genetics of hereditary pancreatitis (HP) in adult patients with Chronic pancreatitis, and heterozygous mutations of ATM gene. **Methods:** This outpatient study was conducted on 200 Patients, of whom (n=100) from Psychogenome, Canada, and (n=100) came from Tajrish's Hospital, which wore diagnosed with Chronic Pancreatitis. We examined the patients and their parents for Hereditary Pancreatitis. HP defined as those with trypsinogen gene (PRSS1) mutation on the long arm of chromosome seven (7q35), and Heterozygous ATM mutation was defined with those with 1 or more mutations in our panel [(8 variants) 5762ins137, 5762-1050A>G, E1978*, Q1970*, R2506Tfs*3, R35*, c.3576G>A, c.7638_7646del9, p.H1082Lfs*14). Patients were also categorized in groups of having mutation in their ATM gene DNA was extracted from peripheral blood leukocytes, and exons 2 and 3 of the gene of PSSR2 and exon 2-6 of ATM were individually amplified by polymerase chain reaction and sequenced by NGS. **Results:** Of 200 adult patients with pancreatitis (median age 55), 64 (32%) were diagnosed with adult onset of HP. From the 64 patient with positive mutation on the trypsinogen (PRSS1) gene, 34 of them had one parent with mutation of the trypsinogen gene (PRSS1). The route of inheritance is Autosomal Dominant. From the 34 adult patients who had parents with the mutations of the trypsinogen gene (PRSS1), and 37 (19%) were heteroz, for ATM gene mutation, 39 (61%) recalled of their babies or young children, with having been diagnosed with epigastric pain, severe vomiting and nausea both after birth and during early childhood. From the 37 patients who wore diagnosed with having the mutations in the ATM, (PRSS1) gene and HP, 11 (30%) indicated that one of their parents has been deceased or currently diagnosed with Pancreatic Cancer. From this 37 patients with (HP) and genetic mutations of ATM and PRSS1 gene 14 (38%) also had Type 1 Diabetes Mellitus, 7 (19%) had a sibling or family member which was diagnosed with breast cancer and 4 (11%) had colorectal cancer. **Conclusions:** Patients with HP who have a heterozygous mutation of the ATM gene have a higher incident of developing Pancreatic Cancer. We suggest GI-Fellows and physicians to work closely with genetic counselors for families with history of HP or pancreatitis and cancer . Better risk models and Genetic panel screening will help for a precise diagnostic and preventive interventions early in the life of individual patients.

2308W

Hypomorphic autosomal dominant polycystic kidney disease (ADPKD): A rare case of lethal biallelic ADPKD and asymptomatic heterozygotes. *A.L. Goetsch[1], B. Angle[2], S. Peters[3], M.J. Hussey[4], J.C. Lane[5].* 1) Genetics, Birth Defects & Metabolism, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 2) Pediatric Genetics, Advocate Lutheran General Hospital, Park RIdge, IL; 3) Genetic Medicine, Northwestern Medicine, Winfield, IL; 4) Maternal Fetal Medicine, Northwestern Medicine, Winfield, IL; 5) Nephrology, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL.

   Autosomal Dominant Polycystic Kidney Disease (ADPKD) is one of the most common genetic conditions with prevalence of 1:400 to 1:1000. The disease is considered nearly 100% penetrant. Typically, adults over 30 years of age known to have pathogenic variants in *PKD1* and *PKD2* are found to have bilateral renal cysts. It has been established in the literature that rare *biallelic* inheritance of *PKD1* pathogenic variants mimic Autosomal Recessive Polycystic Kidney Disease (ARPKD) within families characterized by otherwise typical ADPKD. To our knowledge, individuals reported with *biallelic* ADPKD have at least one affected parent. We describe an infant with lethal *biallelic* ADPKD and two unaffected heterozygote parents, thus supporting the theory of *hypomorphic* alleles in ADPKD.   Prenatal ultrasound was significant for enlarged, echogenic kidneys with oligohydramnios and hydrops fetalis, findings typically consistent with ARPKD. Prenatal testing was inconclusive, as a single paternally inherited variant of uncertain clinical significance in *PKHD1* was identified (c.275G>A, p.Arg92Gln). The infant was born at 30 2/7 weeks gestation due to premature labor. His health was complicated by kidney failure, respiratory failure due to pulmonary hypoplasia, and congenital heart defects. The infant died on day of life four. His autopsy was significant for large kidney size (493g) with presence of glomerulocystic disease (GKD). These findings were suggestive of either ADPKD or isolated GKD in the absence of specific syndromic features. Additional genetic testing revealed two novel likely pathogenic variants in *PKD1* (c.5480G>A, p.Gly1872Glu; c.6542A>G, p.Tyr2181Cys). Parental studies confirmed the infant inherited the *PKD1* variants in *trans*. At this time, both parents are in good health with no evidence of renal cysts on ultrasound imaging. Their two living children and current pregnancy are known to carry the *PKD1* c.6542A>G (p.Tyr2181Cys) likely pathogenic variant. All family members are currently undergoing appropriate ADPKD surveillance given the lack of data on *hypomorphic* alleles in ADPKD.   In summary, this case supports the possibility of *hypomorphic* alleles in ADPKD, such that *PKD1* variants with reduced or incomplete penetrance in heterozygotes can present as lethal disease in compound heterozygotes. It raises the need for additional research to determine appropriate medical management guidelines and genetic counseling for seemingly asymptomatic heterozygotes.

2309T

Molecular characterization of Peruvian families with Fabry Disease. *G. Manassero-Morales, K. Franco-Bustamante.* Genetics, Instituto Nacional de Salud del Niño San Borja, Lima, Lima, Peru.

  Introduction: Fabry disease [MIM 301500], is a progressive, X-chromosomal inherited lysosomal storage disorder, of the catabolism of glycosphingolipids produced by a deficit of lisosomal enzyme α-galactosidase A (α-GAL A). The Human Gene Mutation Database, Fabry mutation database, and Clin Var database have hundreds of mutations of GLA gene encoding α-GAL A registered. The estimated incidence of Fabry disease is □1 in 50,000 males in the world; although the incidence or prevalence in Perú has not been estimated. **Objective:** To present the molecular characterization of fourteen Peruvian families with Fabry disease. **Methods:** A screening program using α-GAL A activity in Blood Spot was performed in patients undergoing hemodialysis at the largest hospitals in the country; sixteen patients had reduced enzyme activity confirmed in leukocytes. Complete molecular analysis of the GLA gene showed mutations in fourteen patients. A family tree was built for each proband including all members of at least four generations; enzymatic and molecular familial mutation targeted testing was performed in all available at risk family members. **Results:** Of the fourteen mutant alleles found in these families, eleven different mutations were identified; three families shared mutations (p.D109G, p.K130 and p.R363H). Four novel mutations were detected, (p.G35A, p.I64F, p.K130T and p.G171S). Of the 364 at risk members tested for their specific familial mutation, 128 (35.16%) heterozygous cases were detected, from which, 44 (34.37%) were male patients and 84 (65.63%) female. Twenty male patients and one symptomatic woman are actually undergoing Enzyme Replacement Therapy (ERT). Results are summarized in table 1. **Conclusions:** The molecular GLA full gene analysis in fourteen Peruvian families showed eleven different genotypes; four novel mutations were found. Results confirm the importance of pedigree analysis as a useful tool to determine at risk family members in order to identify symptomatic patients and recommend early replacement treatment. **Disclosure:** This study was sponsored by Sanofi-Genzyme.

Table 1: Genotype characterization in Peruvian Families with Fabry Disease						
Family	Genotype	Members at risk tested	Heterozygous mutation detected	Male	Female	ERT
1	M1T	78	19	10	9	12
2	R363H	52	22	9	13	1
3	G171S	5	3	1	2	1
4	M1K	18	4	1	3	2
5	D109G	21	10	5	5	3
6	I64F	12	5	1	4	0
7	N215S	15	5	2	3	1
8	G35A	25	12	2	10	0
9	D109G	9	20	4	16	0
10	K130T	7	3	1	2	1
11	W220R	60	11	4	7	0
12	I198T	9	2	1	1	0
13	R363H	24	5	1	4	0
14	K130T	29	7	2	5	0
Total		364	128	44	84	21

2310F

Identification and molecular modeling of novel mutations in Congenital Adrenal Hyperplasia patients (*CYP21A2* **gene).** *R. Prasad[1], R. Khajuria[1], R. Walia[2], A. Bhansali[2].* 1) Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Endocrinology, PGIMER, Chandigarh, Chandigarh, India.

  Congenital Adrenal Hyperplasia (CAH) is autosomal recessive disease, with a wide range of clinical manifestations from severe classical form to late onset form. Most cases of CAH, the inherited inability to synthesize cortisol, aldosterone with subsequent overproduction of androgens, are caused by mutations in steroid 21 Hydroxylase enzyme encoded by CYP21A2 gene. Our aim was to identify novel mutations subsequently their functional implication in enzyme by in silico analysis in 55 patients. Novel mutations were identified by SSCP technique and subsequently sequencing of amplified product. Multiple sequence alignment (MSA) was done to evaluate the functional impact of the regions where novel mutations were present. Functional implication of novel mutations was analysed by various mutation severity prediction softwares. 3-D analysis of the interaction of wildtype and mutant amino-acid with other interacting partners within enzyme was analysed by swiss-PDB viewer. We identified 4 novel mutations viz., F306V, D234D, P357P and H365N.They accounted for 1.8% in Indian population. It was deduced after the alignment that Histidine in case of H365N mutant is present in conserved region and is involved in binding of 21-OHase enzyme to heme. Histidine at 365 position is conserved in different species. Although phenylalanine at 306 position is not conserved residue as seen by the result of multiple sequence alignment by CLUSTAL W but this residue lies in the region forming I-helix of the protein. The I-helix of the protein is involved in substrate recognition. The severity prediction softwares viz., PROVEAN, SIFT and PolyPhen predicted novel mutations to be deleterious, affect protein function and probably damaging to the protein. 3D structure of 21-hydroxylase depicted that Histidine at 365 position interacts with the heme- ligand. Upon substitution with asparagine i.e.H365N, the interaction between heme and the mutated amino acid residue is interrupted leading to unstable protein. Phenylalanine at 306 position is illustrated that upon substitution with valine that causes interruption in interaction as present in wild type protein. Disruption of the π-π stacking interaction among the interacting amino acid residues W302, F306, and F304 with F306 prevent stable packing interactions. Hence, the relevant information about the mutation severity can be acquired from the bioinformatic tools within a period of time.

2311W

Association of airway abnormalities with 22q11.2 deletion syndrome.
R. Sacca[1], K.B. Zur[2,4], K.D. Valverde[1], D.M. McDonald-McGinn[3,4]. 1) Genetic Counseling Master's Program, Arcadia University, Glenside, PA; 2) Pediatric Voice Program and The Center for Pediatric Airway Disorders, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) 22q and You and Clinical Genetics Centers, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) The Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA.

Background: 22q11.2 Deletion Syndrome (DS) presents with complex but variable symptoms most frequently including cardiac, immune, palatal, endocrine, cognitive and psychiatric issues. Structural airway abnormalities have not been formally described. Given the significant morbidity and mortality observed in patients with 22q11.2DS and symptomatic airway disease, we set to document the frequency of this association. **Methods:** We retrospectively reviewed the medical records of patients with 22q11.2DS evaluated at the Children's Hospital of Philadelphia between 1999 and 2015 who were referred to otolaryngology for an airway assessment. The type of airway abnormality observed and the presence of comorbidities such as congenital heart disease, the incidence of tracheostomies, and association with prenatal polyhydramnios were noted. **Results:** 104 patients who underwent an otolaryngology procedure (laryngoscopy or bronchoscopy) were identified. Of these, 77%, ranging in age from 5 months to 37 years, had airway anomalies. Males and females were equally affected. Findings included: tracheomalacia (35%), subglottic stenosis (27%), laryngomalacia (23%), glottic web (23%) and bronchomalacia (15%). The majority of patients (84%) had an associated congenital heart defect with Ventricular Septal Defect and Tetralogy of Fallot being the most prevalent. Importantly, 28% required a tracheostomy and 18% were noted to have a history of polyhydramnios. **Conclusion:** Our findings suggest that airway abnormalities are a common feature of 22q11.2DS leading to significant morbidity particularly when combined with complex cardiac disease. Further, polyhydramnios may be a useful prenatal clue to the diagnosis of 22q11.2DS with airway anomalies. Lastly, postnatal assessment of airway structure and function should be a key component of the evaluation of patients with 22q11.2DS that will help guide patient management. .

2312T

X-chromosome mutations and deletions in a female with premature ovarian insufficiency.
O. Migita[1,2], N. Shimbashi[3,4], S. Igarashi[3], N. Suzuki[1,3], K. Hata[2]. 1) Department of Clinical Genetics, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan; 2) Department of Maternal-Fetal Biology National Research Institute for Child Health and Development, Setagaya, Tokyo, Japan; 3) Department of Obstetrics and Gynecology, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan; 4) Nanako Ladies' Clinic, Ota, Tokyo, Japan.

Women with premature ovarian insufficiency (POI) have repeatedly been reported X chromosome abnormalities. X chromosome terminal deletion and FMR1 gene premutation are well known risk factors for POI. But it is still unclear what caused POI. The application of new molecular genetic analyzing technology would be help in resolving unknown mechanisms of POI. Cytogenetic and molecular characterization of POI is also important for understanding which genes are crucial for normal ovarian development. We conducted karyotyping and whole exome sequence to address the mechanisms of POI. Cytogenetic study with 643 female who have been diagnosed POI were done. Whole exome sequencing was used to screen for mutation in 47 POI individuals from unrelated families. Our results show 12.1% of patients have chromosomal abnormalities. X chromosome abnormalities are most often observed, which have reported in previous studies. Partial deletions in long arm of chromosome X support the hypothesis that haploinsufficiency of X-linked genes can be on the basis of POI. We intend to determine the causable genetic factors in X chromosome by next generation sequencing technology. Using exome sequencing to find rare variants that could not be found in control data. It is still unkown whether affective mutations are included in our observed variant, or not. Our exome sequencing had identified 157 candidate variants on X chromosome.

2313F

Rare nephronophthisis related ciliopathy identified by next generation sequencing in ten non-consanguineous families. *N. Morisada[1,2], A. Shono[2], K. Nozu[2], T. Ninchoji[2], K. Nagatani[3], T. Ohta[4], J. Shimizu[5], T. Yoshikawa[6], K. Saida[6], S. Ishimori[7], M. Yasui[8], C. Nagano[9], K. Kamei[6], K. Ishikura[6], S. Ito[10], R. Tanaka[11], K. Iijima[2].* 1) Clinical Genetics, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan; 2) Pediatrics, Kobe University Graduate School of Medicine; 3) Department of Pediatrics, Uwajima City Hospital; 4) Pediatric Nephrology, Hiroshima Prefectural Hospital; 5) Pediatrics, Okayama Medical Center; 6) Nephrology and Rheumatology, National Center for Child Health and Development; 7) Pediatrics, Kakogawa West City Hospital; 8) Pediatrics, Fukuyama Capital Hospital; 9) Pediatrics, Japanese Red Cross Nagoya Daini Hospital; 10) Pediatrics, Yokohama City University Graduate School of Medicine; 11) Nephrology, Hyogo Prefectural Kobe Children's Hospital.

Purpose Nephronophthisis related ciliopathy (NPHP-RC) is a rare hereditary renal disorder manifested by cystic kidney and end stage renal disease (ESRD) in childhood. The patients of NPHP-RC are often complicated by extra-renal symptoms; such as retinitis pigmentosa, skeletal anomaly, liver disorder, cerebellar anomaly and intellectual disability. More than twenty genes have been reported as the cause of NPHP-RC and most of genes associated with NPHP-RC are autosomal recessive inheritance. The most frequent causative gene in NPHP-RC is *NPHP1*, which account for more than 90% of NPHP-RC. Therefore, the causative genes of NPHP-RC other than *NPHP1* are extremely rare. Comprehensive genetic analysis using next generation sequencing (NGS) may be useful to diagnose of rare NPHP-RC. **Methods** We performed genetic analysis for the patients with congenital anomalies of the kidney and urinary tract (CAKUT), NPHP-RC or cryptogenic ESRD in childhood using NGS. We used TruSight One Sequencing Panel or HaloPlex custom kits for gene targeted analyzing in MiSeq platform. This study was approved by the Institutional Review Board of Kobe University Graduate School of Medicine. **Results** We identified seven genes in 10 patients from 10 non-consanguineous families as candidates for their NPHP-RC. Male to female ratio was 3:7, and average diagnosed age was 7.86 (± 8.55 standard deviation, 0.58-27). Identified genes were *OFD1* (2 patients), *WDR19* (2), *WDR35* (2), *SDCCAG8*, *TMEM67*, *MKKS*, and *NPHP3*. A girl with 4 base pairs deletions in *SDCCAG8* and a woman with *MKKS* splice site mutations showed homozygous mutations. Some boys showed hemizygous mutations in *OFD1* gene. Rest of them had compound heterozygous mutation in each gene. In clinical manifestations, 8 patients in this study fell into ESRD and required renal replacement therapy. Other two patients had cystic kidneys with renal dysfunction. 9 patients showed mild to severe intellectual disability. 4 patients had polydactyly and thoracic deformity. 3 had hepatic disorder. Only one patient had congenital heart anomaly. **Conclusion** We presented the rare NPHP-RC related disorders in Japan. NGS is useful for undiagnosed renal disorders.

2314W

Identification of mutations in *GANAB*, a new gene causing polycystic kidney disease. *C. Ferec[1,2,3,4], S. Mastri[3], A. Bertholet-Thomas[5], M. Hourmant[6], E. Cornec-Le Gall[7], Y. Le Meur[7], J.M. Chen[1,2,4], M.P. Audrézet[1,2,3].* 1) Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, France; 2) Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France; 3) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Universitaire (CHU) Brest, Hôpital Morvan, Brest, France; 4) Etablissement Français du Sang (EFS) – Bretagne, Brest, France; 5) Néphrogones, Centre de Référence des Maladies Rénales Rares, Hôpital Femme-Mère-Enfant, Bron, France; 6) Service de Néphrologie et Immunologie Clinique, CHU de Nantes, France; 7) Service de Néphrologie, Hémodialyse et Transplantation Rénale, CHU Hôpital de la Cavale Blanche, Brest, France.

Autosomal Dominant Polycystic Kidney Disease (ADPKD; MIM#s 173900, 613095) is the most common form of Polycystic Kidney Diseases (PKD) and one of the most common monogenic diseases. ADPKD is characterized by progressive renal cyst formation during the whole life, ultimately resulting in End Stage Renal Disease (ESRD) (Torres et al. Lancet 2007; Harris and Torres. J Clin Invest 2014). ADPKD is genetically heterogeneous with two genes identified: *PKD1* (16p13.3) which encodes polycystin 1 (PC1) and *PKD2* (4q22) which encodes polycystin 2 (PC2). Very recently, a collaborative study coordinated by Peter C. Harris has recruited 327 ADPKD families in which no mutations in the *PKD1* and *PKD2* genes had been identified. Whole exome sequencing (WES) of six families firstly identified a candidate gene named *GANAB* (11q12.3); the gene encodes the catalytic alpha subunit of glucosidase 11 and is expressed in human kidney and liver. The analysis of the other 321 families then led to the identification of eight additional families carrying *GANAB* mutations (Porath et al. Am J Hum Genet. June 2016). Here we analyzed the *GANAB* gene in some 50 French ADPKD families in which no mutations in the *PKD1* and *PKD2* mutations had previously been identified. We found two obvious loss-of-function mutations in the *GANAB* gene that segregated perfectly with the renal phenotype of the disease in two families. The first is a frameshift mutation, c.2733delC, which was found in a family with three affected patients. The second mutation is a splice mutation, c.39-1G>C, which was identified in a family with four affected patients. In addition, we identified two *GANAB* missense mutations each in a single family. None of the *GANAB* mutation-positive patients have so far developed ESRD and the great majority of them displayed also liver cysts. In summary, our results supported the recent finding that *GANAB* is a new gene causing PKD. All *GANAB* mutation-positive patients had rather small renal cysts and reached no ESRD indicated that these patients may be diagnosed as having atypical ADPKD. Genotype/phenotype correlations using larger independent cohorts will improve our understanding of the etiology of the disease. ADPKD patients with few renal cysts and no enlarged kidneys but having also liver cysts should now be analyzed for mutations in the *GANAB* gene.

2315T

Genetic diagnosis of endocrine disorders: Implications for bench to bedside studies. *S. Birla, E. Malik, An. Sharma, T. Grover, A. Sharma.* All India Institute of Medial Sciences, delhi, New Delhi, India.

Successful translation of molecular research in clinical practice is evident by increasing patient referrals for genetic testing in various medical specialties. Impact of genetics is apparent in endocrine disorders (growth hormone deficiency, obesity, endocrine neoplasias, disorders of sexual dysgenesis, skeletal dysplasias etc) which involve confirmation of diagnosis, treatment, management, prevention, effective surveillance, pre- and post- natal diagnostic identification of genetic causes in rare disorders etc. Ours being a tertiary care referral centre, patients with a variety of disorders are referred for treatment and diagnostic testing. We report here the results of cases referred to our laboratory during the period 2011 to 2016 which either ruled out or supported a diagnosis and thus helped in providing effective patient care. A total of 360 patients were tested for more than twenty genes which include *PRKAR1A* (Carneys complex)*, MC4R* (Obesity), *CYP19A1*(Aromatase deficiency), *MENIN*(Multiple endocrine neoplasia-1)*, RET(*Multiple endocrine neoplasia-2*), CYP21A2*(Congenital adrenal hyperplasia), *XOR*(Xanthine cholelithiasis), *TNSALP*(Hypophosphatasia), *CASR*(Hypocalcemia), *GH1* and *GHRHR* (Isolated growth hormone deficiency), *PROP1, POU1F1* and *HESX1*(Combined pituitary hormone deficiency)*, GHR*(Larons Syndrome)*, SRY, NR5A1, DAX* and *DHH*(Disorders of sexual dysgenesis), *COMP* and *COL9A1*(Multiple epiphyseal dysplasia), *COL2A1*(Spondyloepiphyseal dysplasia), *EXT1* (Hereditary multiple exostosis)and *FGFR3*(Achondroplasia) based on the diagnosis. Molecular analysis revealed novel and reported mutations/variations in 88 (24%) cases which were absent in the control database ExAC as well as in 100 healthy controls screened. Some rare disorders like xanthine cholelithiasis and hypophosphatasia were also analysed but no mutations were identified in the *XOR* and *TNSALP* genes. The patient group comprised more of sporadic compared to familial cases. Novel variations were mostly identified in sporadic cases which need characterization. Primary outcome of the study was creation of a database for the northern part of India, providing information on the mutation/variation pattern and their population-specific distribution. Knowledge of molecular lesions not only contributed towards an effective support secondary to clinical assessment and management but also identification of the novel mutations added to the universal pool of genetic variation.

2316F

De novo **missense mutation in** *GRB10* **in a patient with growth restriction and ketotic hypoglycemia.** *A.E. Koehler[1], P. Sharma[1], G.A. Golas[1], J. Fraser[2], T.C. Markello[1,3], D.R. Adams[1,3], M.C.V. Malicdan[1], W.A. Gahl[1,3], C.J. Tifft[1,3].* 1) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, NIH and National Human Genome Research Institute, NIH, Bethesda, MD; 2) Medical Genetics and Genomics Medicine Training Program, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Office of the Clinical Director, National Institutes of Health and Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Growth factor receptor-bound protein 10 (GRB10) is an adaptor protein that interacts with mitogenic receptor tyrosine kinases. Mutations in *GRB10* have not been definitively linked to a human disorder. Through the NIH Undiagnosed Diseases Program (NIH UDP), we describe a growth-restricted child with a monoallelic *GRB10* mutation. The proposita is a 3-year-old girl of European descent admitted to the NIH UDP for failure to thrive, developmental delay, dysmorphic features, and growth restriction (<1st centile height and weight) with normal head circumference. She was previously treated for acquired cataracts, secondary conductive hearing loss, and acquired hypothyroidism. Extensive clinical and biochemical testing revealed hypotonia, macrocytic anemia, ketotic hypoglycemia with low serum insulin, and hepatic steatosis. Additionally, she had a low free T4 with elevated TSH, low IGF1 and tryptophan, and elevated urine organic acids. Whole exome analysis including the parents and two unaffected siblings revealed a *de novo* mutation in *GRB10*. Maternal uniparental disomy of chromosome 7, which includes *GRB10* (7p11.2-p13), is associated with Silver-Russell Syndrome (SRS; OMIM #180860), a disorder marked by intrauterine growth restriction and facial dysmorphisms, thus implicating over-expression of a maternally-expressed gene in growth inhibition. The role of GRB10 in the etiology of SRS, however, is debated as growth plate cartilage does not express GRB10. The protein possesses a C-terminal Src homology 2 (SH2) domain, which binds phosphotyrosine residues of proteins in pathways including mTORC1 and insulin, allowing GRB10 to inhibit insulin and IGF-I stimulated mitogenesis. The mutation in our patient is near the beginning of this SH2 domain. Functional analysis showed decreased GRB10 abundance in fibroblasts of the proband compared to controls, suggesting haploinsufficiency. Additionally, insulin receptor substrate 1 and 2 (IRS1/2) abundances were decreased, but phosphorylation of mTORC1 targets was increased. We hypothesize *GRB10* haploinsufficiency causes a loss of mTORC1 inhibition; consequent mTORC1 hyperactivation could lead to ER stress. We therefore propose that constitutive induction of ER stress, which is linked to low birth weight in humans and fetal growth restriction and hypoglycemia in mice, is a plausible contributor to our patient's phenotype.

2317W

A novel *LHX4* mutation is associated with hypogonadotropic hypogonadism, not combined pituitary hormone deficiency. *M. Takagi[1,2,3], T. Daitsu[4], C. Numakura[5], T. Sato[3], S. Narumi[3], T. Hasegawa[3].* 1) Department of Pediatrics Kawasaki City hospital; 2) Department of Endocrinology and Metabolism, Tokyo Metropolitan Children's Medical Center; 3) Department of Pediatrics, Keio University School of Medicine; 4) Department of Pediatrics, Yamagata City Hospital Saiseikan; 5) Department of Pediatrics, Faculty of Medicine, Yamagata University.

  Background: Mutations in *HESX1*, *SOX3* responsible for combined pituitary hormone deficiency (CPHD) have been identified in a small number of hypogonadotropic hypogonadism (HH), suggesting that the genetic overlap between CPHD and HH. **Case Presentation:** A 2-month-old boy was referred because of micropenis (stretched penile length 1.0 cm) with intrascrotal testes (1 ml). Hormone assays revealed very-low plasma testosterone levels (0.06 ng/mL). LHRH stimulating test performed at the age of 3 months revealed LH peak 7.3 mIU/mL, and FSH peak 20.7 mIU/mL, suggesting pre-pubertal response. Plasma concentrations of thyroxine and insulin-like growth factor-1 were within normal ranges. No episode of hypoglycemia was noted. Clinical diagnosis was HH. Brain MRI showed a normal size anterior pituitary with a visible stalk. **Method:** Using a next-generation sequencing strategy, we sequenced 9 genes implicated in CPHD, including *POU1F1*, *PROP1*, *LHX3*, *HESX1*, *OTX2*, *SOX3*, *SOX2*, *GLI2*, *LHX4*, and 12 genes implicated in HH, including *CHD7*, *FGFR1*, *FGF8*, *GNRH1*, *GNRHR*, *KAL1*, *KISS1*, *KISS1R*, *PROK2*, *PROKR2*, *TAC3*, and *TACR3*. Transcriptional activity of identified LHX4 variant was evaluated by luciferase reporter assays constructed by inserting the *POU1F1* or *αGSU* promoters. We also performed subcellular localization analyses and EMSA assays. **Results:** We identified a novel heterozygous *LHX4* sequence variation, c.661C>T (p.R221W). R221W LHX4 had reduced transactivation and no dominant negative effect. Subcellular localization revealed no significant difference between wild-type and mutant LHX4. EMSA experiments showed that the R221W LHX4 abrogated DNA-binding ability. **Conclusion:** The R221W was proven to be loss-of function mutation. We showed for the first time that *LHX4* mutation is associated with HH, not CPHD. This study extends our understanding of the phenotypic features, molecular mechanism, and developmental course associated with mutations in *LHX4*.

2318T

Utility of targeted exome sequencing as a diagnostic tool for disorders of sex development. *H. Yoo[1,2], E. Kang[1], B. Lee[1,2], G. Kim[2], J. Jang[3], E. Cho[3], J. Choi[1,2].* 1) Dept Pediatrics, Asan Medical Ctr, Seoul, South Korea; 2) Medical Genetics, Asan Medical Ctr, Seoul, South Korea; 3) Green Cross Genome, Yongin, South Korea.

  Purpose: Patients with disorders of sex development (DSD) can present with a large phenotypic spectrum and caused by a number of different genetic defects. Therefore, it is difficult to reach a specific diagnosis using traditional approaches including biochemical analysis and Sanger sequencing of candidate genes in a number of patients with DSD. Recently, next-generation sequencing technologies have revolutionized the identification of causative genes with diseases with genetic heterogeneity using massive parallel sequencing of multiple samples simultaneously. This study was performed to investigate the genetic etiology of DSD patients using targeted exome sequencing of 67 known human DSD associated genes. **Methods:** This study included 37 patients with 46,XY DSD and 8 patients with 46,XX DSD. Exomes were captured using SureSelect kit (Agilent Technologies) and sequenced on the Miseq platform (Illumina, Inc.), which includes 2914 probes targeting a 152.953 kbp region spanning 67 genes. Mean coverage was over 150X for each sample, and approximately 99.46% of the targeted bases were read. We classified variants into five main categories (pathogenic, likely pathogenic, uncertain significance, likely benign, and benign) according to the American College of Medical Genetics and Genomics guidelines. **Results:** We identified known pathogenic mutations or deletion in 11 (24.4%) in the *AR*, *CYP17A1*, *POR*, and *DMRT1/2* genes. Novel variants were identified in 8 patients (17.8%) in the *AR*, *ATRX*, *MAP3K1*, *NR5A1*, *PROP1*, and *WWOX* genes. Of the patients with novel variants, one patients harbored pathogenic frameshift mutation in *AR* (p.Leu446Phefs*56). Two patients had *PROP* exons 1–3 duplication and *WWOX* duplication, respectively. The remaining 5 patients had variants of uncertain significance. **Conclusions:** Whole exome sequencing is expensive and laborious. Thus, targeted strategies provide efficient methods for identification of the genetic causes in DSD patients. This approach allows for early diagnosis of a genetic cause, which could influence clinical management and genetic counseling. *In vitro* or *in vivo* functional analysis is needed to verify functional implications of the novel variants.

2319F

Identifying candidate mutations for rare birth defects through family trio-based exome sequencing. *T. Nicholas[1], D. Koboldt[1], D. Wegner[2], R. Fulton[1], B. Ozenberger[1], J. Wambach[2], D. Grange[2], R. Wilson[1], F. Cole[2].* 1) McDonnell Genome Institute, St. Louis, MO; 2) Department of Pediatrics, Washington University at St. Louis, St. Louis MO.

Genomic information is increasingly relevant for clinical care, particularly with the availability of rapid, inexpensive tools for genome and exome sequencing and analysisand analysis. The McDonnell Genome Institute and the Department of Pediatrics at the Washington University School of Medicine have worked jointly to develop a program for the identification of candidate pathogenic variants associated with rare birth defects. It is estimated that as many as 80% of such cases can be attributed to inherited or *de novo*underlying genetic variation. Families recruited into the program are selected by a board of medical geneticists, clinicians, developmental biologists, and genome scientists, based on phenotypic information, relevant clinical diagnostic studies, family history, and availability of appropriate informed consented samples. This program provides a streamlined approach for combining clinical expertise with genomic methods to better identify causative genetic variants that can inform and aid the diagnosis and treatment of patients with rare birth defects, and reveal the genetic risk of family members. Our strategy primarily involves exome sequencing of the proband and immediate family members (parents and available siblings). The sequence data are aligned to the human reference sequence and assessed for variants, including both single nucleotide variants and indels, which are further annotated with gene and functional information. Candidate causal variants (*de novo* mutations and rare recessive or compound heterozygous variants) are reported and prioritized based on relevant segregation patterns, predicted functional consequences and population allele frequencies. Here, we present an update to findings from exome sequencing of 12 families (44 samples total) with rare birth defects including congenital heart disease, congenital diaphragmatic hernia, genitourinary defects, sternal malformation, and neuromuscular disease. In particular, we highlight a particularly compelling case family featuring two siblings exhibiting ambiguous genitaliapseudohermaphroditism. Each sibling was found to be compound heterozygous for damaging mutations (one previously identified and one novel) in the *HSD17B3.* gene, which Model system and in vivo testing confirmed pathogenicity and helped predict need for testosterone replacement at puberty.

2320W

Impact of genetic testing on the diagnosis and management of congenital diarrheal disorders. *R. Lombardo, D. Montiel, M. Goldschmidt, M. Aub-El-Haija, R. Hopkin.* Cincinnati Children's Hospital and Medical Center, Cincinnati, OH.

Congenital diarrheal disorders (CDD) comprise a group of rare enteropathies that are typically monogenic and autosomal recessive in nature. Over 50 genes have been implicated in this group of disorders with substantial overlap in phenotype. For many of these conditions, patients present within the first few days to weeks of life with severe diarrhea and dehydration. The non-specific presentation, however, makes diagnosis challenging and is complicated by heterogeneous etiology and a broad differential. Nevertheless, rapid identification of the disorder can have significant impact on management and outcome. Traditional diagnostic approach relies heavily on defining the clinical characteristics of the illness and histological evaluation of intestinal tissue. Diagnostic yield of endoscopy, however, is highly variable. We are reviewing our institution's experience with genetic evaluation of CDD. To date, ten patients with congenital diarrhea or intestinal failure have been identified in whom genetic testing established diagnosis and helped guide disease specific management. Seven patients had negative or non-specific findings on esophagastroduodenoscopy and/or colonoscopy, one patient did not receive endoscopy. Molecular sequencing identified several rare disorders including tufting enteropathy (1), congenital sodium diarrhea (1), tricohepatoenteric syndrome (2), mutation in the EFR3B-PI4KA-TTC7A pathway (2), Aicardi-Goutieres syndrome (1), *IARS2* mutation (1), very early onset inflammatory bowel disease (1), and sucrase-isomaltase deficiency (1). Six patients had genetic syndromes for which disease specific management was available. Targeted intervention led to improvement in both symptoms and growth. In two families, a previous child had died prior to diagnosis of the specific disorder. This series demonstrates how genetic testing in conjunction with detailed phenotyping has a direct impact on patient management, leading to improved outcomes. Molecular testing should be considered as a first line diagnostic tool, and may be more appropriate than traditional endoscopy in clinically tenuous patients.

2321T

Identification of a novel variant in X-linked Alport syndrome. A case report. *R. Ortega[1], A.I Sanchez[1], J.C Prieto[1,2].* 1) Bogota, Instituto de genética humana, Universidad Javeriana, Bogota, Bogota, Colombia; 2) Hospital la Victoria SDS, Bogotá Colombia.

　　Introduction: Alport syndrome (AS) is an inherited heterogeneous disorder characterized by hematuria and progressive nephropathy, often associated with high-tone sensorineural deafness and ocular lesions such as lenticonus and maculopathy. Prevalence has been estimated at 1:50,000 live births. It is caused by mutations encoding the collagen IV alfa3-alfa5 chains. Approximately 80% of AS is X-linked; 15% is autosomal recessive, and 5% is autosomal dominant. The X-linked form is caused by mutations in the COL4A5 gene. **Case report:** this is a 9 y/o male patient who was born by uncomplicated spontaneous vaginal delivery, after an uneventful first full term gestation from healthy non-consanguineous parents. Birth weight was 3.7 kg. Significant family history was denied. At age 2 y/o, he started having episodic gross hematuria which became almost persistent at age 9. Overt proteinuria was seen without diminished glomerular filtration rate. Diminished hearing was detectable at age 8 by audiometry, with bilateral reduction in sensitivity to high tones. Ophthalmic examination revealed no abnormalities. Renal biopsy showed diffuse thinning of the glomerular capillary wall with some areas of thickening and multilamellation of the glomerular base membrane. Physical examination revealed no abnormalities. Sequence analysis of COL4A3, COL4A4 and COL4A5 genes was performed. The hemizigous variant c.393dup (p.Gly132Trpfs*26) was identified in COL4A5 gene. **Discussion**: COL4A5 gene is located at Xq22 and encodes for collagen IV that represent a major component of the basement membranes in the glomerulus, cochlea and lens which explains the clinical phenotype. To our knowledge, the variant found in this patient hasn't been report previously in literature. It is known that affected patients with the X-linked form may have severe and progressive disease. **Conclusions**: We present a patient with consisted phenotype of AS; therefore we propose a novel variant in COL4A5 gene for AS. It is important to identify early clinical features to establish an early diagnosis and initiate management to avoid complications and final renal stage disease.

2322F

Genetic and clinical spectrum of a large cohort of patients with distal renal tubular acidosis. *V. Palazzo[1], A. Provenzano[1], F. Becherucci[2], B. Mazzinghi[2], V. Rlandini[1], L. Giunti[3], G. Sansavini[2], R.M. Ropert[2], M. Pantaleo[3], R. Artuso[3], E. Andreucci[3], S. Bargiacchi[3], G. Traficante[3], S. Stagi[4], G. Colussi[5], L. Garavelli[6], S. Penco[7], F. Emma[8], M.R. Caruso[9], S. Andrulli[10], E. Benetti[11], L. Murer[11], G.M. Ghiggeri[12], M. Materassi[2], P. Romagnani[1,2,13], S. Giglio[1,3].* 1) Department of Biomedical Experimental and Clinical Sciences "Mario Serio", University of Florence. Viale Pieraccini, 24. Florence; 2) . Nephrology and Dialysis Unit, Meyer Children's University Hospital, Viale Pieraccini 24, Florence, Italy; 3) Medical Genetics Unit, Meyer Children's University Hospital, Viale Pieraccini 24, Florence, Italy; 4) Endocrinology Unit, Department of Health Sciences, University of Florence, Viale Pieraccini 24, Florence, Italy; 5) Division of Nephrology, Dialysis and Renal Transplantation, Niguarda Ca'Granda Hospital, Milan, Italy; 6) Genetics Unit, Department of Obstetrics and Paediatrics, IRCCS S. Maria Nuova Hospital, Reggio Emilia, Italy; 7) Medical Genetics Unit, Department of Laboratory Medicine, Niguarda Ca'Granda Hospital, Milan, Italy; 8) Nephrology, Bambino Gesù, Children's Hospital, Roma, Italy; 9) Nephrology and Dialysis Unit, Papa Giovanni XXIII Hospital, Bergamo, Italy; 10) Department of Nephrology and Dialysis Alessandro Manzoni Hospital, Lecco, Italy; 11) Pediatric Nephrology, Dialysis and Transplant Unit Department of Pediatrics, University of Padua, Padua, Italy; 12) Division of Nephrology, Dialysis and Transplantation, Giannina Gaslini Institute, Genoa, Italy; 13) Excellence Center DENOTHE, University of Florence, Florence, Italy.

　　Primary distal renal tubular acidosis is a rare and complex genetic disease characterized by inability to maximally acidify urine, thus resulting in metabolic acidosis. Mutations in *SLC4A1*, *ATP6V0A4* and *ATP6V1B1* genes have been described as the cause of the disease, transmitted as either an autosomal dominant or an autosomal recessive trait. Particular clinical features, such as sensorineural hearing loss, have been mainly described in association with mutations in one gene instead of the others. Nevertheless, the diagnosis of distal renal tubular acidosis is essentially based on clinical and laboratory findings and the series of patients described so far are small. Therefore, a strict genotype-phenotype correlation is still lacking and questions about whether clinical and laboratory data should direct the genetic analysis remains open. In this study, we applied next-generation sequencing in 105 patients with a clinical diagnosis of distal renal tubular acidosis, analyzing the prevalence of genetic defects in *SLC4A1*, *ATP6V0A4* and *ATP6V1B1* genes as well as the clinical phenotype. A genetic cause was determined in 60% of cases. In contrast to previous observations, in our group of sporadic cases, the association of the disease with specific clinical features, including sensorineural hearing loss, is not a good indicator of the causal underlying gene. Mutations in *ATP6V0A4* gene are quite as frequent as mutations in *ATP6V1B1*, in patients with an autosomal recessive pattern of inheritance. Our results suggest that when distal renal tubular acidosis is suspected, a complete genetic testing should be performed, irrespective of the clinical phenotype of the patient.

2323W

Biallelic *NUP107* mutations in early childhood-onset steroid resistant nephrotic syndrome. *N. Miyake[1], H. Tsukaguchi[2], E. Koshimizu[1], A. Shono[3], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Second Department of Internal Medicine, Kansai Medical University, Osaka, Japan; 3) Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan.

Nephrotic syndrome is a renal disorder caused by disruption of the glomerular filtration barrier, resulting in massive proteinuria, hypoalbuminemia, and dyslipidemia. At least 27 genes have been reported to be associated with steroid resistant nephrotic syndrome (SRNS), but approximately 70% of the patients with childhood-onset SRNS are genetically unknown. To identify the novel responsible gene for SRNS, we performed WES for 18 families (17 Japanese and 1 Korean families) in whom the causative mutations have not been identified in the known SRNS genes. As a result, we identified biallelic *NUP107* mutations in nine affected individuals from five unrelated families who showed early-onset SRNS. Seven of nine affected individuals had compound heterozygous mutation of one truncating mutation (c.1079_1083del or c.969+1G>A) and commonly sheared missense mutation (c.2492A>C). They developed nephrotic syndrome from 2 to 3 years of age and progressed to end stage renal disease until 10 years old. The rest of two patients in one Korean family had compound heterozygous mutation of two missense mutation (c.469G>T and c.2492A>C) and showed milder phenotype with later onset (10 to 11 years old). Interestingly, all patients harbor the identical missense mutation of c.2492A>C (p.Asp831Ala). Then we constructed the haplotype by using informative microsatellite markers and SNPs, and found a 412-kb haplotype shared by all five families. Since they are all East Asian population and this variant is extreme rare except Human Genetics Variation Database (the public exome database of the Japanese population), c.2492A>C could be a founder mutation in East Asians. *NUP107* encodes Nucleoporin 107kDa (NUP107) is a component of the nuclear pore complex embedded in the nuclear envelope, and ubiquitously expressed including in glomerular podocytes. Our *in vitro* experiment showed the mutant NUP107 protein decrease the binding with NUP133 and altered the intracellular localization. Furthermore, *nup107* knockdown zebrafish generated by morpholino oligonucleotides displayed hypoplastic glomerulus structures and abnormal podocyte foot processes, which could partly recapitulate the renal changes in the patients with *NUP107* mutations. In this study, we identified biallelic *NUP107* mutations caused early childhood-onset SRNS and highlighted the importance of nuclear pore complex in human renal disease.

2324T

Mutations of *CFHR* genes in Chinese atypical hemolytic uremic syndrome (aHUS) with CFH autoantibody positive. *Z. Hao[1,2], W. Li[1,2], X. Liu[3].* 1) Center for Medical Genetics, Beijing Children's Hospital, Capital Medical University, Beijing, China; 2) Beijing Pediatric Research Institute, Beijing, China; 3) Department of Nephrology, Beijing Children's Hospital, Capital Medical University, Beijing, China.

Atypical hemolytic uremic syndrome (aHUS) is a severe kidney disease that is characterized by microangiopathic hemolytic anemia, thrombocytopenia, acute renal failure and is associated with defective regulation of complement activation. A special form of aHUS designated as DEAP (DEficiency of CFHR proteins and CFH Autoantibody Positive)-HUS has been described in recent years, but reports from the Chinese patients are still very limited. In the present study, we aim to explore the relationship between clinical features, outcomes and genotypes of DEAP-HUS from a Chinese cohort. Using enzyme linked immunosorbent assay (ELISA), complement factor H (CFH) autoantibodies were detectable in 31 (50.8%) out of 61 patients with aHUS. Plasma levels of CFH were decreased while C3a, C5a, and sC5b-9 were increased significantly in these CFH autoantibody positive patients compared with the healthy controls. Autoantibody positive individuals treated with immunosuppressants in conjunction with plasma exchange had more favorable outcomes. In addition, CFH level was higher in remission stage than in acute stage. Genetic variants of *CFH*, *CFHR1* and *CFHR3* were analyzed in 12 CFH autoantibody positive Chinese patients by real-time PCR and sequencing of genomic DNA. Among these patients, three carried homozygous deletion of *CFHR1/CFHR3*, four carried homozygous deletion of *CFHR1*, one carried novel homozygous nonsense mutation (p.C129X) of *CFHR1*, one carried novel homozygous missense mutation (p.I214N) of *CFHR1*, and one carried heterozygous deletions of *CFHR1*. No pathological mutation was found in the remaining two patients. The copy number of the *CFH* gene was normal in all 12 patients. In summary, we have identified mutations of the *CFHR* genes in 10 CFH autoantibody positive aHUS patients, suggesting that dysregulation of the complement system plays a major role in Chinese DEAP-HUS patients. Based on our experience, the precision genotyping of DEAP-HUS is a prerequisite for personalized treatment of aHUS. *Corresponding authors: liwei@bch.com.cn (W. Li); desin2000@sina.com(X. Liu).

2325F

Population history of the *NPHS2* **p. V260E mutation, a cause of childhood SR-FSGS, in Southern Africa.** *G.W. Nelson[1], R. Bhimma[2], K. Asharam[2], S. Limou[1], C.A. Winkler[1].* 1) Leidos Biomedical Research, Frederick National Laboratory, Frederick, MD; 2) University of KwaZulu-Natal, Durban, SA;.

Mutations in genes critical to podocyte structure are common recessive causes of childhood steroid resistant focal segmental glomerulosclerosis (SR-FSGS), a severe kidney disease often leading to kidney failure. These mutations occur sporadically in the population, with disease observed in consanguineous families; in population studies no variant has been reported to account for more than a small percentage of disease. Strikingly, a study of 44 children with SR-FSGS in Durban, South Africa, revealed that 30% carried two copies of a missense mutation, p. V260E, in the podocin gene *NPHS2*. This mutation has been observed associated with FSGS in consanguineous families in Oman and in regions associated with the Omani empire of the 19th century, suggesting spread with this empire. We speculated the variant was recently introduced into Southern Africa, and that the children in our study had cryptically consanguineous parents, reflecting a recent founder event. Parents inheriting a variant from a recent common ancestor will have an extended region of shared haplotype around the variant, resulting in an extended stretch of homozygosity in children inheriting the variant from both parents. To test for this we genotyped 10 individuals homozygous for the mutation, and 74 individuals homozygous for the wild type variant, with the Illumina exome chip. After QC, this chip provided 1674 markers within 20 megabases (Mb) of V260E. Two of the ten individuals homozygous for the V260E mutation had regions of homozygosity of 13 and 14 mB around V260E, suggestive of relatively recent common ancestry of these individual's parents, but the other eight had segments of homozygosity from 1.9 to 3.6 mB. To predict the distribution of lengths of homozygous regions around the locus as a function of number of generations between parents and their common ancestor carrying the mutation, we performed 10,000 coalescent simulations of recombination in each generation, generating a distribution of lengths of homozygosity for each n generations from 6 to 48. Comparing with this simulation, the distribution of observed lengths of homozygosity is inconsistent with a common ancestor within the last 19 generations, and most consistent with the most recent common ancestor being 31 generations removed, ruling out a recent founder effect and suggesting that the mutation originated in Africa and spread from there to the Omani Empire.

2326W

mTOR inhibitor therapy demonstrates reduction in renal cysts volume in a 6 year old girl with TSC. *N. Urraca[1], A. Pounders[1], S. Weatherspoon[1,2], J.J. Bissler[1,3].* 1) Le Bonheur Children's Hospital, Memphis, TN; 2) Pediatric Neurology Department, UTHSC, Memphis, TN; 3) Pediatric Nephrology Department, UTHSC, Memphis, TN.

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder affecting approximately 1 in every 6,000 newborns. TSC is frequently associated with de novo mutations in *TSC1* or *TSC2* genes encoding the proteins hamartin and tuberin respectively. The hamartin-tuberin complex inhibits the mammalian-target-of-Rapamycin (mTOR) pathway, which controls cell growth and proliferation. TSC is characterized by hamartomatous growth abnormalities in many organs. Renal disease is a major cause of morbidity and mortality. Approximately 75-80% of TSC patients have angiomyolipomas and 50% have renal cysts. The *TSC2* gene lies immediately adjacent to *PKD1*, the major gene causing autosomal dominant polycystic kidney disease (ADPKD). Approximately 2% of TSC patients will present with polycystic kidney disease as part of the contiguous gene deletion syndrome that involves *TSC2* and *PKD1* genes. Here we present a 6 year old female that was sent to our center with a history of contiguous gene syndrome. She was diagnosed with TSC prenatally when cardiac rhabdomyomas were detected by ultrasound. She meets TSC diagnostic criteria due to the presence of mild facial angiofibromas, subependymal nodules, cortical tubers, two retinal astrocytomas, and multiple hypomelanotic macules. Genetic testing confirmed a deletion in *TSC2* (exons 15-41) and *PKD1* genes. At 5 years of age she complained occasionally of right-sided abdominal pain and early satiety. An abdominal MRI demonstrated two very large cysts in her right kidney which were progressively getting larger and multiple small cysts bilaterally. She also developed hypertension secondary to cystic disease. The two very large cysts were marsupialized and everolimus was initiated twelve weeks after surgery to prevent rapid growth of the other cysts. Everolimus inhibits intracellular mTOR complex and has been shown to shrink angiomyolipomas in patients with TSC. It is FDA approved in TSC for adults with angiomyolipomas not requiring immediate surgery. In this patient everolimus was used off label. Six months after starting everolimus renal MRI revealed resolution of the large cysts and reduction of the smaller cysts. This case raises the possibility that an mTOR inhibitor can be a treatment option for renal cysts in children.

2327T

Detection of early end-organ damage by endothelial dysfunction with reactive hyperemia-digital peripheral arterial tonometry in patients with Fabry disease. *S. Yano[1], K. Moseley[1], C. Azen[2]. 1) Genetics/Pediatrics, University of Southern California, Los Angeles, CA; 2) Clinical Trial Unit Children's Hospital Los Angeles, University of Southern California, Los Angeles, CA.*

Introduction: Progressive renal and cardiovascular lesions have been known to be the major causes of death in individuals with Fabry disease. Based on recent data from the Fabry Registry, proteinuria has been established as the major risk factor for renal disease progression. Microalbuminuria is considered one of the first signs of impairment of renal function in Fabry disease. There is compelling evidence that proteinuria is an indicator of renal dysfunction and requires immediate intervention with enzyme replacement therapy (ERT) and/or antiproteinuric medications; however, there are no clear guidelines for when to initiate ERT in children with Fabry disease. Currently, there are no non-invasive methods or biomarkers for early diagnosis of kidney changes before they become irreversible and subsequent deterioration of kidney function. Endothelial dysfunction precedes the development of microalbuminuria. Plethysmographic studies of forearm blood flow in individuals with Fabry disease show impaired NO-dependent vasodilatation. The evaluation of endothelial function based on reactive hyperemia-digital peripheral arterial tonometry (RH-PAT) with the EndoPAT 2000, a FDA approved instrument to assess endothelial function by reflecting endothelial NO production, may diagnose endothelial dysfunction before irreversible vascular and renal pathologic changes occur, possibly earlier than detection of microalbuminuria. Objective: To evaluate endothelial function with RH-PAT with Endo-PAT 2000 to detect the earliest pathologic changes in the cardiovascular system prior to detection of microalbuminuria in individuals with Fabry disease. Methods: Young male individuals with Fabry disease were enrolled: study subjects n=14 (age: 6-20 years, untreated with ERT n=3). Evaluation of endothelial function with Endo-PAT 2000 was performed. Microalbumin was measured with first void urine specimens. Results: (1) all study subjects showed abnormal endothelial function (Reactive Hyperemia Index: 1.175 ± 0.225, ref: > 1.67), (2) all study subjects showed normal urine microalbumin levels (5.14 ± 3.35 ug/mg Creatinine, n=14, reference range < 30). Conclusions: Endothelial dysfunction was diagnosed by RH-PAT with EndoPAT 2000 in subjects with Fabry disease with normal microalbuminuria. Early diagnosis of endothelial dysfunction by this non-invasive method provides objective clinical information to determine the timing of initiation for ERT. This study was funded by Sanofi Genzyme.

2328F

Mutations in the nuclear bile acid receptor FXR cause progressive familial intrahepatic cholestasis. *R. Xiao[1,2,14], N. Gomez-Ospina[3], C.J. Potter[4], K. Manickam[4], M-S. Kim[2], K. Kim[2], B.L. Shneider[5], J.L. Picarsic[6], T.A. Jacobson[4], J. Zhang[1,14], W. He[1,14], P. Liu[1,14], A.S. Knisely[7], M.J. Finegold[8], D.M. Muzny[9], E. Boerwinkle[9], J.R. Lupski[1], S.E. Plon[1], R.A Gibson[1,9], C.M. Eng[1,14], Y. Yang[1,14], G.C. Washington[10], M.H. Porteus[10], W.E. Berquist[11], N. Kambham[12], R.J. Singh[13], F. Xia[1,14], G.M. Enns[3], D.D Moore[1,2]. 1) Dept. of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept. of Molecular & Cellular Biology, Baylor College of Medicine, Houston, TX; 3) 3Lucile Packard Children's Hospital, Divisions of Medical Genetics, Stanford University Medical Center, CA; 4) Section of Human and Molecular Genetics, Nationwide Children's Hospital, Columbus, OH; 5) Dept. of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Dept. of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 7) Institute of Liver Studies, King's College Hospital, London SE5 9R5, UK; 8) Dept. of Pathology and Immunology, Baylor College of Medicine, Houston, TX; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 10) Pediatric Stem Cell Transplantation, Stanford University Medical Center, Stanford, CA; 11) Pediatric Gastroenterology, Stanford University Medical Center, Stanford, CA; 12) Anatomic and Clinical Pathology, Stanford University Medical Center, Stanford, CA; 13) Immunochemistry Core Laboratory, College of Medicine, Mayo Clinic, Rochester, MN; 14) Baylor Miraca Genetics Laboratories, Houston, TX.*

Neonatal cholestasis is a potentially life-threatening condition requiring prompt diagnosis. Mutations in several different genes can cause progressive familial intrahepatic cholestasis, but known genes cannot account for all familial cases. Here we report four individuals from two unrelated families with neonatal cholestasis. Clinical features of these patients include neonatal onset with rapid progression to end-stage liver disease, vitamin-K independent coagulopathy, low-to-normal serum gamma-glutamyl transferase activity, elevated serum alpha-fetoprotein, and undetectable liver bile salt export pump (ABCB11) expression. Clinical whole exome sequencing for these patients identified mutations in the NR1H4 gene, which encodes the farnesoid X receptor (FXR), a bile acid-activated nuclear hormone receptor that plays central roles in bile acid metabolism. Patients 1 and 2 from the first family both carry a homozygous c.526C>T (p.R176*) pathogenic variant and patients 3 and 4 from the other family carry an in-frame insertion c.419_420insAAA (p.Tyr139_Asn140insLys) and a 30kb gross deletion of the NR1H4 gene. Functional studies showed that the in-frame insertion disrupts DNA binding and transcriptional regulation functions of FXR. In addition, we demonstrated that FXR directly regulates ABCB11 as well as members of the complement and coagulation cascades. Our findings demonstrate a pivotal function for FXR in bile acid homeostasis and hepatoprotection in humans.

2329W

HSP90AA1: A new candidate for Camurati-Engelmann-like phenotype.
S. Ehresmann, J. Gauthier, J. Rousseau, J. Michaud, P.M. Campeau. CHU Sainte Justine Universite de Montreal, Montreal, Canada.

   Camurati-Engelmann disease is a condition with hyperostosis of the long bones and the skull, and dominant mutations in *TGFB1* (encoding transforming growth factor beta-1) are found in a majority of patients. We performed whole exome sequencing in an individual with skeletal features of Camurati-Engelmann disease, hypertension and celiac disease, from consanguineous parents. This led to the identification of a homozygous mutation in *HSP90AA1* (NM_005348:c.1486+4A>G), which encodes a chaperone that protects TFGBRs from Smurf2 ubiquitination, and subsequent degradation of the TGFBR complex. Data from the International Mouse Phenotyping Consortium shows that *HSP90AA1* knockout mice have higher Bone Mineral Content to Body Mass ratio, which supports our hypothesis of *HSP90AA1* being a new candidate gene for Camurati-Engelmann disease. The mutation identified is in the splice donor site of exon 9. A decrease splice donor activity could lead to the inclusion of intron 9 in the coding sequence, and a loss of protein function. A minigene assay is underway to assess the effect of the mutation on splicing, and thus substantiate *HSP90AA1* as a new candidate gene for *TGFB1*-negative Camurati-Engelmann-like phenotype.

2330T

Sclerostin antibody treatment improves the bone phenotype of *Crtap$^{-/-}$* mice, a model of recessive osteogenesis imperfecta. *S. Alexander[1], I. Grafe[1], T. Yang[1], C. Lietman[1], E. Homan[1], E. Munivez[1], Y. Chen[1], M. Jiang[1], T. Bertin[1], B. Dawson[1], F. Asuncion[2], H. Ke[3], M. Ominsky[2], B. Lee[1].* 1) Department of Genetics, Baylor College of Medicine, Houston, TX; 2) Amgen, Inc., Thousand Oaks, CA, USA; 3) UCB Pharma, Slough, UK.

   Osteogenesis imperfecta (OI) is a genetic osteodysplasia producing low bone mass, bone deformities, and fractures. Most cases of OI are caused by autosomal-dominant mutations in type I collagen or by autosomal-recessive variants in genes that modulate post-translational type I collagen modifications, including cartilage associated protein (CRTAP). Bisphosphonates, the current standard OI treatment, do not fully correct the bone phenotype, only moderately reduce fracture risk, and are less effective in adult patients, indicating a need for more effective therapies. Sclerostin-neutralizing antibodies (Scl-Abs) improve bone mass and bone formation in animal models of osteoporosis and dominant OI, but have not been evaluated in recessive OI. The purpose of this study was to investigate the effects of Scl-Ab treatment in the *Crtap$^{-/-}$* mouse model of recessive OI. We treated two age cohorts of *Crtap$^{-/-}$* mice (1-week and 6-week-old mice as models of "pediatric" and "young adult" OI, respectively) with Scl-Ab for 6 weeks (25 mg/kg, s.c., twice per week). We assessed spines and femurs by microCT and histomorphometry and femur biomechanical properties by 3-point bending. MicroCT at L4 vertebrae showed Scl-Ab increased bone mass (BV/TV) from 28% to 71% of WT in the pediatric cohort and from 35% to 72% of WT in the young adult cohort. In femurs, Scl-Ab improved trabecular BV/TV from 45% to 69% of WT in the pediatric cohort and from 32% to 100% of WT in the young adult cohort. On histomorphometry, young adult *Crtap$^{-/-}$* mice had increased osteoclasts and decreased bone formation rate (BFR). Scl-Ab rescued the osteoclast parameters to WT levels and increased BFR from 76% to 128% of WT. The pediatric cohort had similar trends, however while Scl-Ab reduced osteoclast parameters, there was no significant effect on BFR. In the young adult cohort, biomechanical testing of femurs showed that Scl-Ab improved maximum load from 65% to 91% and stiffness from 60% to 88% of WT, but had no significant effect on energy to failure. Similar effects were seen in the pediatric cohort, but to a lesser extent. In summary, Scl-Ab treatment improved bone mass and strength in both pediatric and young adult cohorts of *Crtap$^{-/-}$* mice. However, the less pronounced responses in the pediatric cohort suggests possible age-dependent differences in the effects of Scl-Ab treatment in recessive OI. These findings suggest that Scl-Ab could be a promising new treatment option for patients with recessive OI.

2331F

Effect of bisphosphonates on bone mineral density and other health outcomes in type I osteogenesis imperfecta. *E. Carter[1], J. Bains[1], K. Citron[1], D. Cuthbertson[2], J. Shapirl[10], R. Steiner[3], P. Smith[8], M. Bober[4], T. Hart[9], J. Krischer[2], P. Byers[7], F. Glorieux[5], F. Rauch[5], S. Nagamani[6], V. Sutton[6], C. Raggio[1], Members of the BBD Consortium.* 1) Ctr Skeletal Dysplasias, Hosp Special Surgery, New York City, NY; 2) Health Informatics Institute, University of South Florida, Tampa, FL; 3) University of Wisconsin School of Medicine and Public Health, Madison, WI; 4) A.I. DuPont Hospital for Children, Wilmington, DE; 5) Shriners Hospitals for Children, Montreal, Quebec; 6) Baylor College of Medicine, Houston, TX; 7) University of Washington School of Medicine, Seattle, WA; 8) Shriners Hospitals for Children, Chicago, IL; 9) Osteogenesis Imperfecta Foundation, Gaithersburg, MD; 10) Suburban Hospital, Bethesda, MD.

Osteogenesis imperfecta (OI) is a heritable disorder characterized by low bone mass and bone fragility. This study uses data from the Linked Clinical Research Centers to examine the treatment effect of bisphosphonates (BP; a class of antiresorptive drugs commonly used to treat OI) on bone mineral density (BMD) and other clinically relevant outcomes. Principal methodology: linear regressions of BMD on age were used to construct expected BMD age progressions for untreated individuals with type I (TI) OI. A 2-sided t-test was then used to assess whether BMDs of individuals with TI OI treated with BP differed from their age-matched untreated counterparts. Treatment increased the BMD of individuals with TI OI by 9% relative to those untreated; this was driven by pre-pubescent (<14yr) individuals whose mean increase with treatment was 12% (p<0.01). There was no significant effect on individuals >14yr. Intravenous BP and treatment durations of >2yrs were particularly effective. The treatment effect did not differ markedly by sex. In males, BMDs in those treated with BP were higher than in the untreated individuals across the entire cohort. In females, this effect was observed only in the pre-pubescent years. All results were corroborated by regression analysis. Several sources of downward bias affect this study (*e.g.* not a RCT, age of treatment initiation is not reported). As a result, the treatment effect on BMD is likely underestimated. In addition to those with T1 OI, BP improved BMD z-scores in individuals with types IV, V, and VIII. The treatment effect also extended to non-BMD outcomes. Using logistic regression models, we found that for a 1% increase in BMD, in individuals with TI was associated with a 4% decrease in fracture probability while those untreated had an increase by 1% increase (p<.05). Similarly, scoliosis probability exhibited larger drops in treated pre-pubescent individuals with TI OI for a given BMD increase (p<0.01). BP also slowed the acceleration in scoliosis probability with age in the pre-pubescent individuals with TI OI (p<.01). This suggests that treatment may improve aspects of bone structure more difficult to measure than bone mass, such as bone architecture and mineralization. Linear regression analysis showed that BP treatment also increased mobility in TI OI over the whole age sample (p<.01), providing further evidence that improvements in bone mass and structure raise patient quality of life in multiple ways.

2332W

A founder mutation in the *ZNF687* gene is responsible for giant cell tumor associated with Paget's disease of bone. *G. Divisato[1], F. Scotto di Carlo[1], T. Esposito[1], D. Merlotti[2], L. Pazzaglia[3], E. Siris[4], P. Orcel[5], J.P. Brown[6], M.S. Benassi[3], M.L. Cancela[7], L. Michou[6], D. Rendina[8], L. Gennari[2], F. Gianfrancesco[1].* 1) Institute of Genetics and Biophysics, National Research Council of Italy, Naples, Italy; 2) Department of Medicine, Surgery and Neurosciences, University of Siena, Italy;; 3) Laboratory of Experimental Oncology, Rizzoli Orthopaedic Institute, Bologna, Italy; 4) Department of Medicine, Columbia University Medical Centre, New York, USA; 5) Pole Appareil Locomoteur, Service de Rhumatologie B, Hopital Lariboisiere, Assistance Publique–Hopitaux de Paris, France; 6) Division of Rheumatology, Department of Medicine, Universite´ Laval, Quebec, Canada; 7) Department of Biomedical Sciences and Medicine and Centre of Marine Sciences, University of Algarve, Faro, Portugal; 8) Department of Medicine and Surgery, Federico II University, Naples, Italy.

Paget disease of bone (PDB) is a skeletal disorder characterized by focal abnormalities of bone remodeling, which result in enlarged and deformed bones in one or more regions of the skeleton. In some cases, the pagetic tissue undergoes neoplastic transformation, resulting in osteosarcoma and, less frequently, giant cell tumor of bone (GCT/PDB). We performed whole-exome sequencing in a large family with 14 PDB affected members, four of whom developed GCT at multiple pagetic skeletal sites, and we identified the c.2810C>G (p.Pro937Arg) missense mutation in the zinc finger protein 687 gene (*ZNF687*). The mutation precisely co-segregated with the clinical phenotype in all affected family members. The sequencing of seven unrelated GCT/PDB individuals, identified the same mutation in all unravelling a founder effect. The evidence of a common ancestral chromosome was also corroborated by the segregation of rare variants in genes surrounding the *ZNF687* mutation, suggesting that c.2810C>G mutation originated from a unique haplotype. Our results also showed that *ZNF687*, indicated as a target gene of the NFkB transcription factor by ChIP-seq analysis, was also upregulated in the peripheral blood of PDB-affected individuals with or without mutations in *SQSTM1*, encouraging additional studies to investigate its potential role as a biomarker of PDB risk.

2333T

Identification of novel *COL2A1* mutations that results in dysspondyloenchondromatosis: Support for expansion of *COL2A1* associated disorders. *L.M. Hurd[1], M. Fathy[2], S.M. Kirwin[1], K. Rogers[2], C.P. Ditro[2], W.G. Mackenzie[2], V.L. Funanage[1], M.B. Bober[3].* 1) Molecular Diagnostics Laboratory, Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware; 2) Department of Orthopaedics, Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware; 3) Division of Genetics, Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware.

Enchondromatosis is a rare disorder distinguished by the presence of ectopic, benign cartilaginous growth within the bone. Dysspondyloenchondromatosis (DSC) is characterized as a subtype of enchondromatosis with vertebral involvement. Clinical symptoms of this disorder can include vertebral irregularities (anisospondly) and profuse enchondromas in the long and flat bones. Other remarkable characteristics include short stature, progressive kyphoscoliosis and asymmetric limb shortening. Phenotypic similarity with other collagen type II disorders, including spondyloepimetaphyseal dysplasia-Strudwick type have alluded to deleterious variants in *COL2A1* as a potential cause of DSC. We report on two unrelated patients with this rare type of skeletal dysplasia where thoraco-lumbar kyphoscoliosis and windswept deformity of the lower limbs are the major abnormalities encountered. We performed traditional Sanger sequencing to identify potential pathogenic variants within the *COL2A1* gene. Using DNA isolated from bone and cartilage tissues, we have identified likely pathogenic missense mutations in the triple-helical region of *COL2A1* in both patients. Patient one harbors a c.1555 G>C, p.Gly519Arg substitution in exon 24, and patient two exhibits a c.3572G>A, p.Gly1191Glu substitution in exon 50. *COL2A1* as the cause of DSC has remained controversial. Cases of DSC have presented both with and without the presence of a deleterious variant in *COL2A1,* suggestive of genetic heterogeneity or possible mosaicism. We could hypothesize that patients which present clinically with DSC, but fail to exhibit deleterious variants in *COL2A1* may be mosaic for a deleterious allele. Genotyping DNA isolated from tissues affected by disease may play a crucial role in genetic diagnosis. We provide further support for the expansion of the collagen type II spectrum of disorders to include DSC by discovery of these novel mutations.

2334F

Identification of a novel osteogenesis imperfecta gene. *M.K. Hytönen[1,2,3], G. Vidgren[4], K. Kyöstilä[1,2,3], A.K. Lappalainen[5], S. Malkamäki[2], K. Vainio-Siukola[4], M. Arumilli[1,2,3], M. Anttila[4], H. Lohi[1,2,3].* 1) Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; 2) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 3) The Folkhälsan Institute of Genetics, Helsinki, Finland; 4) Pathology Unit, Finnish Food Safety Authority, Evira, Helsinki, Finland; 5) Department of Equine and Small Animal Medicine, University of Helsinki, Helsinki, Finland.

Osteogenesis imperfecta (OI) is a group of genetic disorders affecting the connective tissue and featured by fragile bones resulting in easily fractured bones and skeletal deformities. The disorder is commonly caused by mutations affecting the synthesis or post-translational modification of the collagen type I. As a model of human OI, we have studied a severe form of OI in the Finnish Lapphund dog breed to identify its genetic cause. The affected dogs die neonatally or have to be euthanized soon after birth due to multiple fractures. The postmortem examination of 15 affected puppies in several litters demonstrated osteopenia as a prominent feature. A pedigree analysis suggested an autosomal recessive disease. A genome-wide association study using 15 cases and 15 controls mapped the locus at region on chromosome 21 that did not contain known OI genes. Whole genome and exome sequencing data analysis under recessive model in three cases identified a case-specific coding structural variant in a gene related to calcium-phosphate metabolism. Our ongoing analyses aim to validate the candidate variant in a large cohort of cases and controls in the breed. Our preliminary data has revealed a novel OI locus and suggests a physiologically relevant candidate gene with a novel disease mechanism. Canine OI patients will serve as large animal models for additional functional and clinical intervention studies.

2335W

Achondroplasia: PEDICAT as an assessment of the impact on daily functionality. *M.D. Irving, C. Coleman, M.S. Cheung, J.A. Hulse.* Guy's and St Thomas" NHS Trust, London, United Kingdom.

Achondroplasia is the most common cause of genetic disproportional short stature. It is caused by a recurrent mutation in *FGFR3*, the gene encoding the fibroblast growth receptor type 3, which is a transmembrane protein expressed in growth plate cartilage chondrocytes. Achondroplasia is associated with a number of physical and medical complications, as a consequence of abnormal bone formation, resulting in elbow contractures, middle ear dysfunction and the impact upon activities of daily living through skeletal disproportion. The significant morbidity and mortality risks of these can be largely overcome through anticipatory surveillance, bit it remains that the physical effects of the condition can have a significant impact upon affected individuals in their daily activities, including functionality. PEDICAT (Pediatric Evaluation of Disability Inventory – Computer Adaptive Test) measures functional abilities in three domains: daily activities, mobilities and social/cognition. It is a tool that has established applicability across a number of pediatric conditions illustrating the impact on the activities of daily living. Here we demonstrate the use of PEDICAT as a tool in highlighting the specific cognitive and functional needs of children with achondroplasia. It applies especially to those requiring earlier intervention, with a view to developing further a multidisciplinary service that addresses the functional needs of this group of patients, as well as addressing their needs based on medical complications. The PEDICAT tool was administered to over 30 pediatric patients attending the dedicated achondroplasia clinic at the Evelina London Children's Hospital. Mostly the data depicts single episodes, but, also, in a limited number of patients, longitudinal assessment of functionality. Here we summarise the data based upon PEDICAT assessment of daily functionality in children aged between 5-16 with achondroplasia and highlight the main factors used to determine improvement of our clinical service to address the specific needs highlighted.

2336T

Farber disease: Acid ceramidase deficiency is more common than previously thought, and slowly progressive phenotypes may only be diagnosed in adulthood. *J. Mitchell[1], A. Solyom[2], X. He[3], C. Simonaro[3], E.H. Schuchman[3].* 1) Pediatric Endocrinology, Montreal Children's Hospital, Montreal, Canada; 2) Plexcera Therapeutics LLC, New York, USA; 3) Genetics and Genomic Sciences, Icahn School of Medicine at Mt. Sinai, New York, USA.

Farber disease is caused by mutations in the *ASAH1* gene, resulting in acid ceramidase deficiency and accumulation of the pro-inflammatory and pro-apoptotic lipid, ceramide. Patients usually present in infancy or late childhood with one of three characteristic symptoms: early-onset polyarticular arthritis, subcutaneous nodules and dysphonia. Acid ceramidase deficiency can also cause another disease, Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME). The prevalence of both Farber disease and SMA-PME is currently unknown, and awareness is extremely limited due to their rarity and lack of specific treatments; both are likely underdiagnosed. Data including genotypes, clinical, biochemical and immunologic phenotypes, is being collected on a growing cohort of 40 Farber patients from around the world. This is the largest Farber cohort yet reported. Findings to date reinforce the validity of the characteristic symptoms of Farber disease. However, they also reveal that there are patients who present with only one or two of these symptoms, and that the spectrum of disease includes remarkably attenuated forms with relatively little associated disability. Such patients are likely to be misdiagnosed or experience a significant delay in diagnosis. Three siblings were recently diagnosed with Farber, by whole exome sequencing, after developing peripheral osteolysis at over 40 years of age. Currently, Farber disease can be treated with hematopoietic stem cell transplantation, which has shown variable results and carries a severe burden for the patients. Acid ceramidase enzyme therapy is under development, and a natural history study is planned to better understand disease progression and factors influencing the different phenotypes. Based on the information gathered so far from this cohort, screening of selected patient populations (such as patients with early-onset, therapy-resistant polyarticular arthritis, or with peripheral osteolysis of unknown cause) for Farber disease is indicated.

2337F

Evidence that bi-allelic mutations in *NPR3* result in a peculiar phenotype with tall stature, arachnodactyly, long halluces and multiple extra epiphyses in hands and feet. *G. Mortier[1], T.R. de Jong[2], H.S.A. Heymans[3], J.M. Wit[4], W. Van Hul[1], E. Boudin[1].* 1) Center of Medical Genetics, University of Antwerp & Antwerp University Hospital, Antwerp, Belgium; 2) Department of Plastic and Reconstructive Surgery and Hand Surgery, Isala Clinics, Zwolle, The Netherlands; 3) Department of Pediatrics, Emma's Childrenshospital- Academic Medical Centre, Amsterdam, The Netherlands; 4) Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands.

   Studies in mice and humans have shown that natriuretic peptides and their corresponding receptors control endochondral ossification and linear bone growth. Reciprocal translocations resulting in an increased production of the C-type natriuretic peptide (CNP) and heterozygous activating mutations in the natriuretic peptide receptor 2 (NPR2) have been reported in individuals with tall stature. In contrast, bi-allelic inactivating mutations in *NPR2* have been identified in patients with severe short stature (acromesomelic dysplasia Maroteaux type). With this paper we provide strong evidence that bi-allelic hypomorphic mutations in the clearance receptor *NPR3* can also result in enhanced growth. Compound heterozygosity for a missense (NM_001204375.1, c.442T>C;p.Ser148Pro) and nonsense mutation (NM_001204375.1, c.1524delC, p.Tyr508*) was identified by WES in two affected sibs born to non-consanguineous, healthy parents of Dutch origin. Both the parents and the unaffected younger sib are heterozygous carriers. Using mRNA isolated from whole blood, we demonstrated that the frameshift mutation resulted in nonsense mediated mRNA decay while localization studies revealed that the mutant NPR3 receptor with the p.Ser148Pro substitution was not incorporated into the plasma membrane. These data suggest that both mutations have a loss-of-function effect on the protein. The phenotype observed in both affected boys was characterized by tall stature (Marfanoid habitus), generalized joint hyperlaxity with hypotonia, mild pectus excavatum, long fingers (arachnodactyly) and markedly long halluces. In addition, hand radiographs revealed the presence of extra (pseudo)epiphyses in first metacarpals and all proximal and middle phalanges. Extra (pseudo)epiphyses were also visible in the proximal phalanges of both feet. This remarkable radiographic finding is retrospectively also visible in the family with an activating mutation of NPR2 reported by Miura K et al. (PLoS ONE 2012). It suggests that enhanced growth due to defects in either the NPR2 or NPR3 controlled signaling pathways may be partially mediated through the creation of extra growth plates in tubular bones.

2338W

Raine syndrome (OMIM #259775), caused by mutations in *FAM20C*, is congenital sclerosing osteomalacia with cerebral calcification (OMIM 259660). *S. Mumm[1,2], W.H. McAlister[3], M.D. Fallon[4], M.E. Pierpont[5], V.N. Bijanki[2], S. Duan[1], G.A. Otaify[1,2,6], W.S. Sly[7], M.P. Whyte[1,2].* 1) Div of Bone and Mineral Diseases, Washington Univ Sch of Med, St Louis, MO, USA; 2) Center for Metabolic Disease and Molecular Research, Shriners Hospital for Children, St. Louis, MO, USA; 3) Dept of Pediatric Radiology, Washington Univ Sch of Med, St. Louis, MO, USA; 4) Dept of Surgical Pathology, Univ of Pennsylvania Sch of Med, Philadelphia, PA, USA; 5) Dept of Pediatrics, Univ of Minnesota Sch of Med and Children's Hosp of Minnesota, Minneapolis, MN, USA; 6) Dept of Clinical Genetics, Centre of Excellence of Human Genetics, National Research Centre, Cairo, Egypt; 7) Doisy Dept of Biochemistry, St. Louis Univ, St. Louis, MO, USA.

   In 1985, we briefly reported two sisters with a new, lethal, autosomal recessive disorder that we called congenital sclerosing osteomalacia with cerebral calcification. In 1986, our findings entered *Mendelian Inheritance In Man* (MIM) as osteomalacia, sclerosing, with cerebral calcification (MIM 259660). However, no attestations followed. Instead, in 1989 Raine and colleagues published a concise description of an affected neonate and considered the striking clinical and radiographic features unprecedented. In 1992, after two additional unrelated newborns were reported, the disorder became "Raine syndrome" and entered MIM formally as osteosclerotic bone dysplasia, lethal (MIM #259775). In 2007, the etiology was discovered to be loss-of-function mutations of the FAM20C gene that encodes family with sequence similarity 20, member C. *FAM20C* is highly expressed in calcified tissues and its translated protein is a kinase for most of the secreted phosphoproteome including FGF23, osteopontin, and several additional regulators of skeletal mineralization. Here, we detail the clinical, radiological, biochemical, histopathological, and *FAM20C* findings of the sisters. Following pre-mortem tetracycline labeling, the proposita's post-mortem skeletal histopathology revealed no rickets but documented osteomalacia accompanying severe osteosclerosis. Her sister showed evidence of hypophosphatemia. Archival DNA from the proposita and her parents and brother indicated the etiology of the disorder was compound heterozygosity for a unique missense defect and a novel deletion in *FAM20C*. Heterozygous family members carrying the missense mutation seemed susceptible to trigonocephaly. Our findings clarify the function of *FAM20C* and show that Raine syndrome is congenital sclerosing osteomalacia with cerebral calcification.

2339T

Defect in phosphoinositide signaling through a homozygous mutation in *PLCB3* causes a new form of spondylometaphyseal dysplasia with corneal clouding. *S. Robbins[1], S. Ben-Salam[2], N. Sobreira[1], A. Lyon[3], A. Al-Shamsi[4], B. Islam[5], N. Akawi[6], A. John[2], P. Thachillath[7], D. Valle[1], B. Ali[2], L. Al-Gazali[7].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Pathology, College of Medicine and Heath Sciences, United Arab Emirates University; 3) Chemistry and Biological Sciences, 560 Oval Drive, West Lafayette, IN 47907, BRWN 3130C, USA; 4) Department of Paediatrics, Tawam Hospital, Al-Ain, United Arab Emirates; 5) School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX; 6) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK; 7) Department of Paediatrics, College of Medicine and Heath Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates.

Skeletal dysplasias are a large group of osteochondrodysplasias affecting the growth and structure of the skeletal system. Here, we report the clinical and molecular delineation of a new form of spondylometaphyseal dysplasia with cloudy corneas in two affected male cousins from a consanguineous Emirati family. Both patients had postnatal growth deficiency with profoundly shortened limbs with involvement of proximal and distal segments, irregular iliac abnormalities, narrow chest, hypertelorism, prominent eyes, depressed nasal bridge, short upturned nose and corneal clouding. Whole genome SNP genotyping and homozygosity mapping identified a region of homozygosity spanning 19.32 Mb over 11q12.1-q13.1. Whole exome sequencing in the 2 affected cousins and an unaffected sib of each revealed homozygous variants in 5 autosomal genes (*CELF1, PRPF19, FEN1, INCENP, PLCB3*), the latter 4 of which are in the 11q12.1-q13.1 region, plus 1 hemizygous X-linked variant (*CETN2*) shared by the affected cousins but not present in the unaffected sibs. Because pathogenic variants in genes involved in phospholipid metabolism, *PLCB4* and *PCYT1A,* have also been reported to cause bone disease with or without eye anomalies, we focused on *PLCB3* as our main candidate gene. The *PLCB3* encodes phospholipase C beta which catalyzes the conversion of PIP2 to IP3 and DAG. The *PLCB3* missense variant *(*p.A878S) we identified alters single amino acid in the Ha2' element of the proximal CRD domain. We transfected wild type *PLCB3* and variant *PLCB3*-A878S into COS7 cells and compared their enzymatic activity. The enzymatic activity encoded by the *PLCB3*-A878S allele was ~7% of control, indicating the missense variant produced a hypomorphic allele. Interestingly, Lowe syndrome (OMIM 309000), an X-linked recessive phenotype, characterized by thick clouding of the lenses in both eyes, delayed development, and renal Fanconi syndrome, is caused by variants in *OCRL* which encodes PIP2 5-phosphatase that catalyzes an alternate reaction in PIP2 degradation. Suchy and Nussbaum (2002) suggested that elevated PIP2 levels in patients with Lowe syndrome cause abnormalities in the actin cytoskeleton. Here, we hypothesize that *PLCB3* loss of function variants may lead to elevated levels of PIP2 in affected cells causing the eye and skeletal phenotype in our patients. To test our hypothesis, we are performing F-actin staining to quantify stress fiber formation in patient fibroblasts.   .

2340F

Deletion of the first exon of *COL1A2* results in a mild form of osteogenesis imperfecta mediated by loss of transcription from the mutant allele. *J. Schleit[1], T. Tran[1], Y. Liu[1], T. Fields[1], D. Leistritz[1], E. Carter[2], J. Davis[2], P. Fernhoff[3], P.H. Byers[1,4].* 1) Department of Pathology, University of Washington, Seattle, WA; 2) Division of Genetics, Hospital for Special Surgery, New York, NY; 3) Department of Human Genetics, Emory University, Atlanta, GA; 4) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA.

More than 90% of individuals with osteogenesis imperfecta (OI) have variants in the type I procollagen genes, *COL1A1* and *COL1A2,* that encode the proα1(I) and proα2(I) chains, respectively. Approximately 70% of individuals with OI have variants within *COL1A1*. Almost half of the variants identified in *COL1A1* are heterozygous null variants due to introduction of a premature termination codon by any of several mechanisms. In contrast, null variants in *COL1A2* are rarely reported, perhaps because the phenotypic consequences are usually minor. We identified a heterozygous 851bp deletion within one allele of *COL1A2* that includes 54bp upstream of the translation start site, exon 1, and 760bp of intron 1in all members of a four generation family who had a history of fractures. This deletion prevented expression of one allele of *COL1A2*, confirmed by sequencing of cDNA, and resulted in reduced synthesis of proα2(I) chains of type I procollagen. Analysis of type I procollagen under non-reducing conditions showed increased formation of proα1(I) trimers. Identification of this variant indicates that *COL1A2* null variants can cause bone fragility and deletion/duplication analysis of both *COL1A1* and *COL1A2* should be considered when a genetic etiology of bone fragility is suspected.   .

2341W

Identification of novel compound heterozygous *BMP1* mutations in osteogenesis imperfecta with increased bone mineral density. *A. Shimada[1,2], M. Takagi[2,4], H. Fujita[3], S. Narumi[4], T. Hasegawa[4], G. Nishimura[5].* 1) Department of Pediatrics, Tama-Hokubu Medical Center, Tokyo, Tokyo, Japan; 2) Department of Endocrinology and Metabolism, Tokyo Metropolitan Children's Medical Center; 3) Department of Rehabilitation, Hokkaido Medical Center for Child Health and Rehabilitation; 4) Department of Pediatrics, Keio University School of Medicine; 5) Department of Radiology, Tokyo Metropolitan Children's Medical Center.

    Osteogenesis imperfecta (OI) is a spectrum of genetic disorders characterized by bone fragility and increased fracture risk. More than 90% cases of OI are caused by heterozygous mutation in *COL1A1* and *COL1A2* encoding for the two α-chains of collagen type I. In most of the cases, patients show markedly decreased bone mineral density reflecting abnormal synthesis of type I collagen. Here, we report an OI patient with high bone mineral density harboring mutation in *BMP1*. The patient is a 2-year-old Japanese boy who suffered from recurrent fractures from the neonatal period. He was the first child of nonconsanguineous healthy parents and born at term after uneventful pregnancy. Birth weight, length, and OFC were 3542g (+1.2 SD), 51.0cm (+0.8 SD) and 34cm (+0.4 SD) respectively. After the first fracture admitted on day 3, he experienced recurrent fractures without any episodes of significant trauma. Radiographs showed dense bone of lower limbs and no remarkable finding as OI was identified. Unexpectedly, he showed high bone mineral density at 11 months (BMD 0.26 g/cm^2, Z score +1.5 SD). On the suspicion of OI from the patient's clinical history, we sequenced 10 causative genes of OI by next generation sequencing and identified *BMP1* mutations (E214Q/X731S). Glutamine 214 is a highly evolutionarily conserved amino acid. BMP1 remove the C-propeptides from procollagen precursors and enable mature collagen monomers to form collagen fibrils. So far, fourteen patients of OI caused by *BMP1* mutation in 11 families were reported. BMD were evaluated in 7 families and increased bone mineral density was noted in 5 families. We also experience another case of OI with high bone mineral density harbouring a mutation in COL1A1 C-propeptide cleavage site. We should keep in mind that OI patients not always show decreased bone mineral density.

2342T

Whole exome sequencing ends the diagnostic odyssey for an infant with osteopetrosis. *C. Trujillo[1], L.C. Burrage[2], M. Jain[2], M.K. Alwasiyah[4,5], S. Jhangiani[3], D.M. Muzny[3], H. Bouazzi[6], B.H. Lee[2], R. Gibbs[3].* 1) Dept Gen, Erfan Hospital, Jeddah, Saudi Arabia; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 4) Aziziah Maternity & Children Hospital, Jeddah, Saudi Arabia; 5) Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia; 6) Necker-Enfants Malades Hospital, Paris Descartes University, Paris, France.

    Osteopetrosis is a rare skeletal disorder associated with increased bone density secondary to impaired osteoclast differentiation or function. Osteopetrosis is genetically heterogeneous with multiple genes associated with the phenotype. Here, we present a 2 month old girl born to a consanguineous couple who was referred to us for pallor, growth failure and hepatosplenomegaly. On hematological examination, thrombocytopenia, leukocytosis and myelophthisic anemia were noted. Bone marrow studies suggested a possible diagnosis of "osteopetrosis." Radiographs, which showed uniformly increased bone density, confirmed the diagnosis. In particular, radiographs of the skull demonstrated thickened and dense bone, especially at the base, and sinuses were small and underpneumatized. By six months of age, macrocephaly with progressive blindness and deafness were noted Initially, Sanger sequencing of *TNFS11* was performed and was negative. Then, a next-generation sequencing panel of 14 genes associated with high bone mass was sent and was also negative. Consequently, whole exome sequencing was performed and revealed an apparently homozygous pathogenic variant in *SNX10* (p.Glu128fs) which is predicted to result in a frameshift in this gene and which was confirmed using Sanger sequencing. Pathogenic variants in *SNX10* have been associated with a diagnosis of osteopetrosis, autosomal recessive 8 (OMIM #615085), and this gene was not included on the previously performed gene panel. This diagnosis ended the diagnostic odyssey for this patient, and moreover, this case demonstrates the utility of whole exome sequencing in the setting of a negative gene panel for disorders with increasing genetic heterogeneity.

2343F

Mutation in the gene for osteoprotegerin in two families with calcium pyrophosphate deposition arthropathy. *C. Williams[1], A. Ortiz[1], U. Qazi[2], A. Kaell[2], M. Bernstein[2], G. Eshel[3], A. Rosenthal[4].* 1) Biomedical Sciences, Cooper Medical School of Rowan University, Camden, NJ; 2) John T Mather Memorial Hospital-SUNY Stony Brook Medicine, Port Jefferson, NY; 3) Department of Medicine, Assaf Harofe Medical Ctr, Tel Aviv, Israel; 4) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI.

Calcium pyrophosphate deposition disease (CPDD) is characterized by the deposition of calcium-containing crystals in affected joints. When the disease occurs in families, it is inherited in an autosomal dominant manner and, in some cases, the occurrence of CPP crystals coincides with premature primary generalized osteoarthritis (PGOA). Two distinct genetic loci have been linked to the disease. The first locus, CCAL1, is on chromosome 8q while the second locus, CCAL2, is on chromosome 5p. To date, eight families with CPDD have been linked to the CCAL2 locus and missense, insertion, and frameshift mutations in the progressive ankylosis (ANKH) gene have been demonstrated. In 2015, a second gene for CPDD was reported with the discovery of a substitution mutation in the Stop codon of the TNFRSF11B gene in a Dutch CPPD/PGOA family (Ramos YF et al, Ann Rheum Dis 74(9):1756-1762, 2015). This gene codes for osteoprotegerin (OPG), a decoy receptor for RANKL that inhibits the binding of osteoclast-bound RANK thus inhibiting the differentiation of osteoclasts. We examined two CPDD families, one from Israel and one from the Northeastern US, for mutations in both ANKH and TNFRSF11B. While no mutations were observed in ANKH for either family, both families displayed the same read-through mutation of the termination codon (Stop402Leu) of the TN-FRSF11B gene as observed in the Dutch family. Genotyping of SNPs in and around the TNFRSF11B gene in affected members of each family demonstrated that the kindreds are apparently unrelated. Studies performed by Ramos et al suggest that this TNFRSF11B mutation results in a gain of function with respect to the ability of mutant OPG to inhibit osteoclastogenesis. In order to assess the potential impact of mutant OPG on pyrophosphate (PPi) generation, we performed metabolic labeling studies on fibroblasts from two affected members of our Israeli family and observed that levels of extracellular PPi were similar to those obtained for cultured fibroblasts from several unaffected controls. Furthermore, adding OPG to chondrocytes did not increase levels of extracellular ATP, a precursor of extracellular PPi. Finally, steady-state expression levels of ANKH transcript were unchanged in comparison to unaffected controls. These observations suggest that the Stop402Leu mutation in TN-FRSF11B does not significantly impact ANKH expression or function. Future studies will elucidate the role of OPG in CPDD.

2344W

An activating *IDH1* variant leading to inherited bilateral atypical hyaline cartilage neoplasms of the mastoid. *P.R. Blackburn[1], D. Oglesbee[2], T.R. Caulfield[3], N.J. Boczek[4,5], M.A. Cousin[4,5], J.M. Carter[2], R.H. Gavrilova[6,7], C.Y. Inwards[2], E.W. Klee[4,5].* 1) Center for Individualized Medicine, Mayo Clinic, Jacksonville, FL 32224, USA; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA; 3) Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA; 4) Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA; 5) Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 6) Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA; 7) Department of Clinical Genomics, Mayo Clinic, Rochester, MN 55905, USA.

We report a family (mother and two sons) presenting in childhood with atypical cartilaginous tumors of the mastoid consistent with low-grade chondrosarcoma. The mother developed a unilateral middle ear mass that was resected with a mastoidectomy at 12-years-of-age. Her sons at 9- and 11-years-of-age presented with headaches, facial paresis, and conductive hearing loss requiring bilateral tympanomastoidectomy after the confirmation of bilateral mastoid tumors. The clinical presentation and radiologic features were suggestive of an atypical hyaline cartilaginous neoplasm with no clonal abnormality detected within the tumors. Research whole exome sequencing was performed on the germline of the proband, mother, father, and the proband's half-brother identifying a novel heterozygous missense variant in isocitrate dehydrogenase 1 (*IDH1*, NM_001282386.1, c.299G>A, p.R100Q) that was unique to the three affected individuals. IDH1 is a NADP-dependent enzyme that decarboxylates isocitrate into alpha-ketoglutarate (α-KG), an intermediate that is utilized by DNA and histone demethylases. Recently, somatic variants have been observed in conserved IDH1 residues that confer gain-of-function activity producing 2-hydroxyglutarate (2-HG) with neomorphic effect. In particular, the IDH1 p.R100Q protein variant has been shown to produce this oncometabolite and is observed in ~ 1% of gliomas. It is presumed that 2-HG inhibits the activity of α-KG-dependent dioxygenases and plays a role in tumor development through global dysregulation of the cellular epigenetic environment. Other somatic heterozygous variants in *IDH1* leading to missense changes at p.R132 are frequently found in Ollier disease (MIM# 166000), Maffucci syndrome (MIM# 614569), and metaphyseal enchondromatosis with urinary excretion of D-2-hydroxyglutaric acid (MIM# 614875). Tumors of the mastoid bone are exceedingly rare, with only a few cases reported to date and there have been no described cases of bilateral tumors developing independently at the same location. Here, we report the first case of a germline *IDH1* variant leading to development of rare bilateral chondrosarcoma of the mastoid. We hypothesize that p.R100Q protein variant may lead to low levels of 2-HG production, impede endochondral ossification, and promote tumor formation within embryonic cartilage rests.

2345T

'Corner fracture' type spondylometaphyseal dysplasia – heterogeneous disorder overlapping with type II collagenopathies. *K. Machol[1], M. Jain[1], M. Almannai[1], T. Orand[1], J. Lu[1,4], A. Schlesinger[3], R. Gibbs[1,4], S. Unger[2], L. Bonafe[2], A. Superti-Furga[2], B. Lee[1], P. Campeau[5], L. Burrage[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine , Houston, TX; 2) Center for Molecular Diseases, University of Lausanne, Lausanne, Switzerland; 3) Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, CHU Sainte-Justine, University of Montreal, Montreal, QC, Canada.

Spondylometaphyseal dysplasias are a heterogenous group of disorders characterized by vertebral and metaphyseal abnormalities. Although the molecular basis for most of these disorders is known, the causative gene has not yet been identified for several of these disorders. Spondylometaphyseal dysplasia (SMD) 'corner fracture' type (also known as SMD 'Sutcliffe' type, MIM 184255) is a rare skeletal dysplasia that presents with mild to moderate short stature, developmental coxa vara, mild platyspondyly, corner fracture-like lesions and metaphyseal abnormalities with sparing of the epiphyses. The molecular basis for this disorder has yet to be clarified. We describe two patients with SMD 'corner fracture' type who have been found to harbor heterozygous pathogenic variants in *COL2A1* by Whole Exome Sequencing (WES). Patient 1 was found to have a de novo heterozygous missense variant in *COL2A1* gene (p.Gly345Asp) that was previously reported in a patient with spondyloepimetaphyseal dysplasia (SEMD) 'Strudwick' type. This is a type II collagenopathy characterized by defective growth and modelling of the spine and long bones with clinical features including disproportionate short stature, malformed vertebrae and abnormal epiphyses or metaphyses. Patient 2 was found to have a novel de novo heterozygous missense variant in *COL2A1* gene (p.Gly876Ser). Both variants are predicted to be pathogenic as they change a glycine residue in the type II collagen protein. These two cases, together with a third case of SMD 'corner fracture' type with a heterozygous *COL2A1* variant (p.Gly181Arg) previously described in the literature, suggest that this disorder is heterogeneous and overlaps with type II collagenopathies. The finding of one of the pathogenic variants in a previously reported case of SEMD 'Strudwick' type and the significant clinical similarity between the two syndromes, suggest an overlap between SMD 'corner fracture' and SEMD 'Strudwick' types.

2346F

Nonsense mutation in WTX gene and osteopathy striata with cranial sclerosis: Early onset of complications. *J. Prieto[1,2], M. Garcia-Acero[1].* 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana , Bogota Cundinama, Bogota, Colombia; 2) Hospital La Victoria, SDS, Bogota, Colombia.

Introduction: Osteopathy striata with cranial sclerosis is a bone dysplasia, dominant X-linked inheritance caused by mutations in WTX a repressor of Wnt, which mediates bone formation, the alteration affects bone density by imbalance between osteoblasts and osteoclasts. Presents with skeletal abnormalities with radiographic findings in form of regular hyperdense linear bands from the metaphysis to diaphysis of long bones, fan-shaped striations in the iliac bones and sometimes is associated with craniofacial hyperostosis and extra-skeletal anomalies. The linear striations of long bones do not cause symptoms, while the cranial sclerosis is often symptomatic due to cranial deformities and injuries due to cranial nerve compression. To date, approximately 100 have been described cases. **Case presentation:** We present a female patient of 4 years with prenatal history of macrocephaly and history of global developmental delay, short stature and congenital ventricular septal defect, at physical examination had macrocephaly, prominent forehead, hypertelorism and cutaneous syndactyly in hands associated with headache, bilateral sensorineural hearing loss and decreased visual acuity. In the context of investigation of short stature, radiographs were requested with evidence of thickening of the cranial vault, longitudinal linear grooves in long bones and spiral fluted ilia. Exome sequencing identified a heterozygous nonsense mutation c.1072C>T in WTX gene, which leads to production of a truncated protein (p.R358X), the change not was evident in the parents, suggesting a de novo change confirming the clinical and radiological diagnosis of this patient. **Discussion.** Osteopathia striata with cranial sclerosis (OS-CS) is a bone dysplasia characterized by longitudinal striations of the metaphysis of long bones, sclerosis of the craniofacial bones and extra-skeletal symptoms. Diagnosis is based on clinical and radiological examinations. The clinical presentation is highly variable, skeletal manifestations ranging from mild to multisystem organ disorders. This patient had early onset of complications with prenatal evidence of cranial anomaly and auditory visual impairment secondary to severe cranial sclerosis. **Conclusion.** The OS - CS is a rare, lethal entity in men and highly variant in women. It should be considered as a differential diagnosis in the fetus/child with macrocephaly. The prognosis depends on the associated anomalies and multidisciplinary management. .

2347W

Vosoritide in children with achondroplasia: Updated results from an ongoing Phase 2, open-label, sequential cohort, dose-escalation study. *J. Hoover-Fong[1], M. Irving[2], C. Bacino[3], X. Cao[4], J. Charrow[5], V. Cormiere-Daire[6], W. Dummer[4], P. Harmatz[7], K. Jayaram[4], L. Katz[4], K. Larimore[4], J. Phillips[8], S. Vaidya[4], R. Savarirayan[9].* 1) Inst Genetic Med, Johns Hopkins Univ, Baltimore, MD; 2) Guy's and St. Thomas' NHS Foundation Trust, Evelina Children's Hospital, London, UK; 3) Baylor College of Medicine, Houston, TX, USA; 4) BioMarin Pharmaceutical Inc., Novato, CA, USA; 5) Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA; 6) Institut Imagine, Universite Paris Descartes, Hopital Necker - Enfants Malades, Paris, France; 7) UCSF Benioff Children's Hospital Oakland, Oakland, CA, USA; 8) Vanderbilt University Medical Center, Nashville, TN, USA; 9) Murdoch Children's Research Institute, Royal Children's Hospital Victoria, University of Melbourne, Parkville, Victoria, Australia.

Background and objectives Achondroplasia (ACH), the most common form of dwarfism, is caused by a mutation in the fibroblast growth factor receptor 3 gene (*FGFR3*) which leads to inhibition of endochondral bone growth. Vosoritide is an analog of C-type natriuretic peptide (CNP), a potent stimulator of endochondral bone growth. This global phase 2, multi-center, open-label, sequential cohort, dose-escalation study was designed to evaluate the safety, tolerability and efficacy of vosoritide for up to 24 months in children with ACH aged 5-14 years. Design and methods Subjects completed at least 6 months of pre-treatment growth measurements in an observational study prior to enrollment to establish baseline annualized growth velocity (AGV). 26 children (mean age 7.8 ± 1.8 years) were enrolled into 3 dose cohorts: 2.5 μg/kg/day (Cohort 1, n=8), 7.5 μg/kg/day (Cohort 2, n=8), and 15 μg/kg/day (Cohort 3, n=10) given daily by subcutaneous route. Based on favorable safety and efficacy data in Cohort 3 after 6 months of vosoritide at 15 μg/kg/day, 12 subjects at lower doses were subsequently escalated to 15 μg/kg/day for ≥6 months. Cohort 3 subjects received vosoritide at 15 μg/kg/day for at least 12 months. Results Vosoritide was generally well-tolerated at a dose of 15 μg/kg/day. The majority of adverse events (AEs) were mild and no serious AEs were reported as study drug-related. Common study drug-related AEs reported for Cohort 3 subjects during 12 months of treatment were injection site reactions (90% of subjects) and asymptomatic hypotension (40% of subjects), which were all mild. The safety profile in subjects escalated to 15 μg/kg/day for ≥6 months was similar, although two AEs of transient symptomatic hypotension were reported in two subjects of which one event was study drug-related. Both subjects continued to receive vosoritide without dose modification and the event did not recur. After 12 months at 15 μg/kg/day, Cohort 3 subjects had 1.9 ± 2.0 cm/year (46%) increase in mean AGV from baseline (p=0.02). Twelve subjects escalated to 15μg/kg/day for ≥6 months had an increase of 2.3 ± 1.9 cm/year (65%) in mean AGV from baseline. No worsening of body proportions was observed as measured by upper-to-lower segment ratio. Conclusions Vosoritide was generally well-tolerated at a dose of 15 μg/kg/day for up to 12 months. Safety and efficacy data support further evaluation and development of vosoritide for the treatment of children with ACH.

2348T

Whole genome sequencing identifies a paracentric mcro-inversion in a Chinese family with isolated split-hand/foot malformation type I. *W. Yang[1,2], Y.Z. Huang[2], D. Lv[2], Y.M. Xu[2], X.L. Zhao[2], X. Zhang[2].* 1) Center for Medical Genetics, Beijing Children's Hospital, Beijing, Beijing, China; 2) Department of Medical Genetics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China.

Backgrounds: Six genetic loci have been identified for Split-hand/foot malformation (SHFM). Among them, Split-hand/foot malformation 1 (SHFM1[MIM 183600]) is caused by deletion, duplication, or rearrangement of chromosome 7q21.3, involving the *DLX5* (MIM 600028), *DLX6* (MIM 600030) genes and possible regulatory elements in the region. Some of those mutations can be detected directly by chromosome banding, chromosome microarray, but others of them, the micro or complex ones, maybe difficult to determine. **Purpose:** We collected a Chinese family with dominantly inherited non-syndromic SHFM and tried to find the causative mutation of the family. **Methods:** Blood samples were gathered from 12 family members including 4 patients. High resolution karyotype analysis and copy number variation (CNV) analysis by SNP array were carried out. Two point linkage analysis were performed on all known SHFM loci with dominant inheritance pattern. Candidate genes (*DSS1*[MIM 601285], *DLX5*, *DLX6*, *DYNC1I1*[MIM 603772]) and some conserved elements around *DLX5/DLX6* were screened by PCR-Sanger sequencing and qPCR. Finally, whole-genome sequencing (WGS) was performed on the proband, and variations within the 17Mb minimal shared haplotype of the family were analyzed. For a structure variation revealed by WGS, the breakpoint junctions were sequenced, then co-segregation analysis were performed by the junction PCR. **Results:** No causative abnormality was revealed through karyotyping or CNV analysis. After linkage analysis, five known SHFM loci were excluded, while SHFM1 was suggested linked. By Sanger sequencing or qPCR, no causative mutations were found at candidate genes or elements in the SHFM1 region. WGS revealed a 2.03Mb inversion at 7q21.3 (AC_002540.1:g65233_AC_091654.4:g16inv), which demonstrated co-segregating with the SHFM malformation in the family. *DLX5*, *DLX6* gene were found almost in the midst of the inverted fragment, and the proximal breakpoint located in intron 13 of the *SLC25A13* gene. At least three elements have been reported with enhancer activity in embryonic limb and just centromeric to the proximal breakpoint. The results above suggested that the 2.03Mb inversion was the cause of the limb malformation of the family, and it may result in the deformity by disturbing the proper interaction between the target genes such as *DLX5*, *DLX6*, and their limb-specific enhancers during embryo development.

2349F

Role of next generation sequencing (NGS) in genetic testing and counseling of hereditary skeletal dysplasias in developing countries. *L. Bhatnagar[1], K.J. Keerthi[1], M. Suraj[1], K. Das[1], K.A. Priya[1], S. Poornima[1], R. Jaiswal[2], S. Bopanna[3], Q. Hasan[1].* 1) Department of Genetics and Molecular Medicine, Kamineni Hospital, L.B. Nagar, Hyderabad, India; 2) Department of Orthopedics, KAMS & RC, L.B. Nagar, Hyderabad, India; 3) Department of Radiology, KAMS & RC, L.B. Nagar, Hyderabad, India.

‘Skeletal Dysplasia (SD)’ is an umbrella term for disorders, which affect the growth of bone and cartilage resulting in abnormal skeletal shape, size, and disproportionate stature. The current census (2011) documented nearly 3 crore disabled people in India, of which >50 lakh have movement disability, some of this is due to trauma, but a large number are due to neurological deficits or SD. The incidence of SD is approximately 1 in 5000, hence, a large proportion of moderate to severe disability is of this category. In developing countries where most of the population has to cater to their own health needs and there are very few support services for the disabled, having more than one affected child/individual in the family is devastating. SD is currently identified based on clinical and radiographical findings. Due to overlapping features exact diagnosis is difficult without molecular testing, hence most remain undiagnosed. In this pilot study, Next generation sequencing (NGS) was employed for exact diagnosis and appropriate management and prevention of hereditary SD through counseling. Twelve cases of SD identified after clinical evaluation and skeletal imaging were referred to the genetics unit of a tertiary care hospital in South India (2015-2016). Seven of the referred patients after detailed history and pedigree analysis by a genetic counselor were evaluated by NGS. Age of the patients ranged from 6 days to 32 years. A panel of 78 genes were assessed and all cases were diagnosed based on these: Osteogenesis Imperfecta (OI) I (*COL1A1*) and OI type VI (*SERPINF1*), Osteopetrosis 1 (*LRP5*) and 2 (*CLCN7*), Mucolipidosis type III (*GNPTG*), Fibrochondrogenesis 2 (*COL11A2*) and Mucopolysaccharidosis (*MPS*) type II (IDS). Specific, targeted mutation analysis was advised to high-risk family members after case-specific genetic counseling based on the type of hereditary disorder i.e., autosomal dominant, autosomal recessive, X-linked recessive. In this way, the benefit of the NGS could be extended to > 32 individuals belonging to these seven families. This pilot study indicates that NGS, though currently expensive, is the best option in developing countries like India, where the subsequent medical costs/disability support can be reduced after proper diagnosis and genetic counseling. The government should include this in its public health policy to both categorize cases of disability and prevent these from increasing to mitigate large-scale suffering.

2350W

A probably new type of mesomelic dysplasia with acral synostoses with scoliosis, antecubital pterigium, and aberrant flexion creases on palms and digits. *J.R. Corona-Rivera[1,2], B. Isidor[3], E. Zapata-Aldana[2], C. Peña-Padilla[2], S. Olvera-Molina[2], L. Bobadilla-Morales[1,2], A. Corona-Rivera[1,2].* 1) Instituto de Genetica Humana, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Servicio de Genética, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 3) CHU Nantes, Service de Génétique Médicale, Nantes, F44093, France.

Introduction. Mesomelic dysplasia (MD) with acral synostoses (AS) includes only the Verloes-David-Pfeiffer syndrome (VDPS, [MIM 600383]), and the mesomelic dysplasia Kantaputra type (MDK, [MIM 156232]). VDPS is characterized by short stature, typical facial gestalt, palatal abnormalities, and MD with AS, this last with proximal carpometacarpal fusion, proximal fusion of metacarpals 4 and 5, and brachydactyly (complex, postaxial). In MDK, the facial features and hands are normal, the fibula, talus, and calcaneus are small and the AS affects predominantly the lower limbs, including fibulocalcaneal synostosis, tibio-talus fusion, and the feet are described as ‘ballerina-like standing’. **Objective**: Here, we describe a mother and its daughter with a form of MD with AS that could correspond to a new type, since both have other previously unreported abnormalities. **Clinical reports.** *Patient 1*. The 21-year-old mother of this report had a height of 138 cm, normal intelligence, expressionless face, mild eyelid ptosis, micrognathia, mesomelic shortening in both upper and lower limbs (mild), antecubital pterigium, short fifth fingers (only), limited flexion of fingers, absence of digital flexion creases, aberrant palmar creases, and bilateral club foot. She had a progressive limited motion of joints. On radiographs, she had thoracic scoliosis, short ulna and radius with bowing (mild) and dislocated at elbow; fusion of carpal bones, proximal fusion metacarpals 4 and 5; coxa valga, short femoral necks, mild tibial and fibular shortening (without bowing), cuboid and cuneiform synostoses, and short metatarsals and phalanges (postaxial). *Patient 2*. She is the unique proposita's daughter and was born at term, with a birth length of 44.5 cm (<3rd percentile). She showed an overall phenotype almost similar to his mother, excepting for scoliosis and those age-related manifestations. In both, no evidence for deletions of *SULF1* and *SLCO5A1* genes at 8q13 was found using whole-genome oligonucleotide array CGH. **Discussion.** Although the phenotype in our patients resembles more the VDPS, array CGH ruled out this diagnosis. Associated abnormalities previously unreported in VDPS or MDK, including scoliosis, antecubital pterigium, and aberrant flexion creases on palms and digits, suggest that the disorder in this family probably represents a new form of autosomal dominant MD with AS. .

2351T

MEP1A mutation identified by whole exome sequencing in a family with slipped capital femoral epiphysis. *Y. Guo[1], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 3) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Slipped capital femoral epiphysis (SCFE) is an important pediatric and adolescent hip disorder in which a fracture through the growth plate (physis) results in slippage of the overlying end of the femur (epiphysis). Risk factors include sex (boy vs girl ratio around 2:1), ethnicity (Africans and Polynesians have higher rates), overweight/obesity (found in 2/3 of SCFE cases), and metabolic endocrine disorders (e.g. hypothyroidism, panhypopituitarism, hypogonadism, renal osteodystrophy, and growth hormone abnormalities). In a family with dominantly inherited bilateral SCFE and severe obesity, we performed whole exome sequencing in two affected kids and found a missense mutation in an enzyme gene MEP1A (meprin A subunit alpha; also known as endopeptidase-2, N-benzoyl-L-tyrosyl-p-aminobenzoic acid hydrolase, PABA-peptide hydrolase, or PPH). Sanger validation showed that this mutation is shared by two affected children and their affected mother. The mutation is located at a highly conserved genomic region (astacin) across vertebrates with very deleterious predictions, and is absent from any public database including ExAC of ~60,000 samples. The human BMP1 gene (bone morphogenetic protein-1) has the same domain astacin, mutations in which can cause osteogenesis imperfecta with SCFE relevant manifestations such as lack of bone modeling with wide distal metaphyses of femora. There are previous reports associating MEP1A gene variants with insulin metabolism, showing the importance of metalloendopeptidases in the breakdown of the gut hormone, and also linking metalloproteases meprin α and meprin β which are C- and N-procollagen proteinases to collagen assembly and tensile strength. Functional experiments are ongoing to characterize the role of MEP1A in developing SCFE related phenotypes.

2352F

Studying the genetic basis of idiopathic short stature using whole exome sequencing. *C.T. Thiel[1], N.N. Hauer[1], E. Schöller[1], S. Schuhmann[1], M.T. Wittmann[1], B. Popp[1], S. Uebe[1], C. Buettner[1], A.B. Ekici[1], R. Abou-Jamra[2], K. Kutsche[3], D. Wieczorek[4], H. Sticht[5], H.-G. Doerr[6], A. Reis[1].* 1) Institute of Human Gentics, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany; 2) Institute of Human Genetics, University of Leipzig, Leipzig, Germany; 3) Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4) Institute of Human-Genetics, University Duisburg-Essen, Essen, Germany; 5) Institute of Biochemistry, FAU Erlangen-Nürnberg, Erlangen, Germany; 6) Department of Pediatrics and Adolescent Medicine, FAU Erlangen-Nürnberg, Erlangen, Germany.

Shortness of stature is a common medical concern in childhood and has an incidence of 3 % in the general population. After excluding defects of the growth hormone pathway and recognizable syndromes the underlying cause remains unknown in approximately 80-85 % of individuals. We now recruited and clinically characterized in detail a large study group of 560 families with idiopathic short stature and selected 200 individuals where common genetic causes were excluded. We first performed trio whole exome sequencing (WES) of affected individuals and their parents of normal height for 100 individuals (trio analysis). Variants were selected unbiased based on all possible modes of inheritance in agreement with the segregation in the families and their potential effect on protein function. To further evaluate candidate genes we performed WES in 100 further individuals with targeted follow-up of single variants in parents. In 21 individuals we found mutations in known short stature genes including *COL2A1*, *CUL7*, *PDE3A* and *KDM5A* (59% dominant, 14% recessive and 27% X-linked inheritance). The underlying diagnosis was missed clinically as characteristic clinical features of the syndromes were missing. The most common mutated gene was ACAN. In the remaining 89 trios we found potential protein affecting variants in 126 novel candidate genes in 62 of the affected individuals. Within our study, a second variant with a compatible inheritance was found for 16 of the 126 candidate genes in the 100 patients of the affected only analysis. Two further genes were also recently listed in another WES study for idiopathic short stature in agreement with our observed dominant *de novo* model of inheritance. The functional analysis of the remaining 120 candidate genes reveled 31 genes with gene-level relevance, suggesting that they are involved in idiopathic short stature. In conclusion, exome analyses of 200 patients with idiopathic short stature identified a known cause of shorted stature in 11% of cases. As the clinical spectrum of most known genetic defects is yet to be explored, an unbiased genetic analysis of patients with idiopathic short stature can help establish a diagnosis in these cases. Furthermore, we found 126 potential novel candidate genes in 62% of the individuals including 18 genes with independent mutations in two individuals each. Thus, our data strongly suggest that single gene defects may be a frequent cause for idiopathic short stature.

2353W

Novel mutations in the fourth β-propeller domain of *LRP4* **are associated with isolated syndactyly with fusion of the third and fourth fingers.** *R. Sukenik Halevy, B. Heinz, N. Ahituv.* Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA.

LRP4 mediates *SOST*-dependent inhibition of WNT signaling and therefore has a negative impact on bone formation. Mutations in this gene cause 3 known syndromes; Myasthenic syndrome, Cenani-Lenz syndactyly syndrome and Sclerosteosis 2 with the latter two having various limb malformations including syndactyly. Using exome sequencing on a trio with a child that has isolated bilateral hand syndactyly with complete fusion of the 3rd and 4th fingers, we identified mutations in *LRP4* to be associated with this phenotype. The patient is a compound heterozygote for two non-synonymous variants, each inherited from a different parent: p.Q1564K and p.D1403H. These variants have not been previously reported in online databases and have CADD scores above 20. No other limb malformation associated *de novo* coding mutations were identified. Analysis of a healthy sib showed that he is not a carrier of these mutations. These variants are located in the 16th and 19th repeats of the LDL-receptor class B protein domain which are within the 4th β-propeller of the extracellular domain of this protein. The extracellular ß –propeller domain is thought to be critical for ligand release and recycling of the receptor as well as for the inhibitory action of sclerostin during WNT signaling. Mice carrying two recessive mutations in the *Lrp4* gene were reported to have polysyndactyly and abnormal nails and incisors. Targeted phenotypic evaluation of our patient revealed no phenotypic features related to any of the known syndromes except for growth parameters which were in the upper normal range (probably constitutional), 5th finger clinodactyly and missing three teeth. To functionally validate these mutations, we are using luciferase reporter assays to assess WNT signaling in the presence and absence of these mutations. In addition, we are currently sequencing exons 28-32, which code for the 4th propeller domain, in a cohort of patients with similar isolated syndactyly phenotypes. In summary, we report a novel *LRP4* phenotype in a patient with isolated bilateral third and fourth finger syndactyly. The two mutations found in our patient are located at a domain in which mutations have not been reported before, indicating the existence of a new attenuated LRP4 phenotype. Acknowledgments: Sequencing was provided by the University of Washington Center for Mendelian Genomics and was funded by the NHGRI and the NHLBI grant 2UM1HG006493 to Drs. Nickerson, Bamshad and Leal.

2354T

Severe Osteogenesis Imperfecta (OI) in 3 Vietnamese siblings with a homozygous mutation of prolyl 3-hydroxylase 1 (*P3H1***), also known as** *LEPRE 1*. *E. Elias[1], B. Tucker[2].* 1) Dept Genetics & Pediatrics, Children's Hospital Colorado, Aurora, CO; 2) Colorado Fetal Care Center, Children's Hospital Colorado, Aurora, CO.

Osteogenesis Imperfecta (OI) is a disorder of connective tissue, usually caused by dominant mutations in type 1 collagen. Patients have been reported with recessive mutations in genes associated with post-translational modifications of type 1 collagen, including P3H1. A Vietnamese family is reported with three affected siblings, presenting with a severe but non-lethal form of OI. These patients help to expand the phenotype and ethnic background of patients with confirmed P3H1 mutations. **Case 1:** 1965 g male product of a 37 week gestation to healthy non-consanguineous Vietnamese parents. The pregacy was closely followed due to a history of two prior children with OI as well as prenatally detected short stature and multiple fractures. At birth the baby was vigorous, but required nasal cannula oxygen. The baby had a transverse defect of the R hand. Post-natal films showed multiple fractures of all long bones, gracile ribs with many fractures, and markedly decreased calvarial mineralization. The baby fed well and was discharged on 0.5L oxygen. He had hearing loss on newborn screen. **Case 2:** 5 yr old sister of case 1, with multiple fractures seen pre- and post-natally. She required nasal cannula O2 during infancy but has no current respiratory issues. She had multiple renal stones as an infant, which resolved with bisphosphonate treatment. She can ambulate with assistance following rodding of her lower extremities. She is cognitively normal with a normal brain MRI. Xrays show diffuse osteopenia and gracile bones but no popcorn epiphyses. **Case 3:** 9 yr old sister of case 1, with multiple fractures pre- and post-natally. She had no respiratory issues. She had hypercalciuria during infancy, but no renal stones. Her brain MRI showed basilar invagination which resolved by age 8. She has had multiple rodding procedures and is a wheelchair ambulator. She is cognitively bright. Xrays show no popcorn epiphyses. **Genetic Testing:** a homozygous c.1170+5G>C (IVS6+5G>C) mutation was seen in all 3 siblings, which disrupts the intron 6 splice site causing exon 6 skipping. **Conclusions:** The phenotype of this homozygous mutation in P3H1 includes severe but non-lethal OI, minimal respiratory issues, basilar invagination and hypercalciuria. No popcorn epiphyses were noted, as has been reported in other patients with LEPRE 1 mutations. These cases provide additional proof that DNA testing in patients with OI can help clarify the diagnosis, as well as provide accurate recurrence risk.

2355F

Nager syndrome presenting with osteoporosis in a 10 year-old male with a novel mutation of SF3B4. *A. Labilloy[1], R. Stottmann[1], C. Gordon[2], B. Kline-Fath[3], H. Saal[1], R. Hopkin[1].* 1) Human Genetics, Cincinnati Children's Medical Center, Cincinnati, OH; 2) Plastic Surgery, Cincinnati Children's Medical Center, Cincinnati, OH; 3) Radiology and Medical Imaging, Cincinnati Children's Medical Center, Cincinnati, OH.

Nager syndrome (MIM 154400) is a type of acrofacial dysostosis characterized by hypoplasia of 1st and 2nd branchial derivatives and preaxial limb abnormalities. The condition is most commonly caused by mutations in SF3B4, which encodes a spliceosome-associated protein involved in BMP signaling, a major mediator of bone development and homeostasis. Estimated prevalence is 1:250,000. We report clinical outcomes of a 10 year-old male prenatally diagnosed with Nager syndrome due to sonographic and fetal MRI findings concerning for severe micrognathia and severely shortened upper limbs with apparently normal brain development. Patient was delivered at 35 weeks by scheduled EXIT procedure for upper airway obstruction requiring tracheotomy while on placental bypass. At birth, multiple craniofacial anomalies were noted, including severe micrognathia with cleft palate, hypoplasia of mastication muscles, microtia with bilateral hearing loss, lid coloboma. Other findings included sacral dimple and hypospadias. MLB/FOB revealed dysplastic epiglottis, posterior glottic stenosis and laryngotracheoesophageal cleft. In addition several skeletal abnormalities were observed, including asymmetric foreshortened upper extremities with absent radius and thumb and bifid humerus, genu valgum, pes planus, Morton's toe with sandal gap deformity and C1 ring hypoplasia. He has a history of two fractures, and at age 10 he was found to be osteopenic with gracile bones on foot Xray. A DEXA scan showed severe osteoporosis (Z score -5.7). Patient is gastrostomy and tracheostomy-dependent, and has a history of motor and speech delays. He has a mouth aperture of 35 mm. He is cognitively and behaviorally appropriate for his age and socially active independently. Molecular testing revealed a novel c.34+15A>T transversion in IVS1 of SF3B4 gene, predicted to generate a new donor splice site, with the potential to result in aberrant mRNA processing. This case illustrates the importance of screening patients with Nager syndrome for osteoporosis, especially those with history of fractures. Given a possible pathogenic link between SF3B4 and bone homeostasis, osteoporosis and skeletal involvement beyond upper limbs are predictable, but unreported features of Nager syndrome. Long-term natural history data for patients with Nager syndrome are crucial for management decisions with the most severe patients. Long-term survival and good QOL are possible.

2356W

Association between the *p.Pro392Leu* mutation within *SQSTM1* gene and air pollutants in Paget's disease of bone. *L. Michou[1,2,3], M.S. Numan[1], J.P. Brown[1,2,3], S. Jean[2,4].* 1) CHU de Québec Research Centre, Quebec (Quebec) Canada; 2) Department of Medicine, Laval University, Quebec (Quebec) Canada; 3) Department of Rheumatology, CHU de Québec, Quebec (Quebec) Canada; 4) Institut national de santé publique du Québec, Quebec (Quebec), Canada.

Purpose: The *p.Pro392Leu* mutation in *SQSTM1* gene is the most frequent mutation linked to Paget's disease of bone (PDB). Associations between PDB and environmental factors such as rural residency, tobacco smoke and wood heating, were also reported. We searched for potential association between the *p.Pro392Leu* mutation and air pollutants in PDB. **Methods**: We studied a subgroup of our French-Canadian cohort consisting in 110 patients, including 39 carriers of the *p.Pro392Leu* mutation, and 113 noncarriers healthy controls. We used a questionnaire on the history, during life, of residence and proximity to sources of outdoor air pollutants (highways, bus, train or airport stations, gas stations). For indoor air pollutants, questions on heating combustibles (carbon, wood, oil) used in past residences and tobacco exposure were also used. Associations between air pollutant exposures and PDB were assessed with Chi-squared, Fisher and t tests. Odds ratio (OR) and 95% confidence intervals were calculated. Multivariate analysis was performed by stepwise selection of variable with p<0.1 in univariate analyses. In a subgroup of patients (n=11 to 29) and controls (n=12 to 48), urinary concentration of 17 heavy metals measured by mass spectrometry were compared by the use of a Wilcoxon test. **Results:** In multivariate analysis, PDB patients carrier of the *p.Pro392Leu* mutation were less frequently exposed than controls to residence close to highways (OR=0.04 [0.02-0.13], p<0.0001). They were less frequently exposed to tobacco smoke in the residence in childhood (OR=0.25 [0.07-0.86], p=0.04) and to forced air from central carbon heating system (OR=0.15 [0.04-0.59], p=0.02). Exposure to forced air from central oil heating systems was the only indoor air pollutant positively associated with pagetic patients carrier of the *p.Pro392Leu* mutation (OR=5.04 [0.92-27.52]) but not with sporadic PDB patients (OR=0.64 [0.24-1.77], p=0.05). Only mean levels of urinary lead were found to be higher among patients with the mutation than controls (p=0.07). **Conclusions:** Patients with PDB carrier of the *p.Pro392Leu* mutation were less frequently exposed to outdoor air pollutants during childhood, which is consistent with the known association of PDB with rural residency, whereas they were more frequently exposed to forced air from oil heating systems. Higher levels of urinary lead in patients with PDB carrier of the mutation may support some potential role for air pollutants in PDB.

2357T

Previously unreported *TCIRG1* mutation in a young girl with osteopetros-is. *A. Nastro[1], V. Neerukonda[1], K. Kerr[2], G. Kennedy[2], H. Gao[3], G. Amundson[4], I. Kharode[5], J.A. Brochstein[6], I. Sahdev[6], J. Gross[5], R.R. Lebel[1].* 1) Center for Development, Behavior and Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Division of Pediatric Hematology-Oncology, SUNY Upstate Medical University, Syracuse, NY; 3) Fulgent Diagnostics, Temple City, CA; 4) Division of Pediatric Radiology, SUNY Upstate Medical University, Syracuse, NY; 5) Department of Pediatrics, SUNY Upstate Medical University, Syracuse, NY; 6) Division of Pediatric Hematology-Oncology, Cohen Children's Medical Center of NY, New Hyde Park, NY.

We report a 2-year-old female with infantile malignant osteopetrosis, type 1B (OPTB1), with heterozygous mutations at the *TCIRG1* gene, one previously unreported and the other known. She also has a mutation of unknown significance in the *OSTM1* gene. Pathogenic mutations at both of these loci are recessive for infantile malignant osteopetrosis. Pregnancy and delivery were uneventful. However, the neonate returned a few days after birth due to poor feeding and irritability. She was found to have petechiae, hepatosplenomegaly, thrombocytopenia, anemia, leukocytosis, and hypocalcemia with elevated PTH. All of these are features of OPTB1. Her condition worsened, and infantile malignant osteopetrosis was diagnosed via plain film chest x-ray and clinical correlation. She had not yet displayed any dysmorphic features, but after a few months frontal bossing was noted. Skeletal dysplasias NGS Panel, which sequenced 161 genes, revealed heterozygous mutations at the *TCIRG1* gene, leading to a formal diagnosis of OPTB1. *TCIRG1* encodes for the 116kD subunit of the osteoclast-specific vacuolar proton pump, and mutations in this gene are the most common recessive cause of infantile malignant osteopetrosis. *OSTM1* encodes for osteopetrosis-associated transmembrane protein 1. Mutations in this gene are a recessive cause of a very severe form of infantile malignant osteopetrosis. Our patient is heterozygous for *TCIRG1* mutation NM_006019.3:c2066G>A (p.Trp689*), a variant not previously reported in the HGMD database or the Broad ExAC dataset. This is a nonsense mutation and considered to be pathogenic. She is also heterozygous for mutation NM_006019.3:c.1674-1G>A, which is thought to affect RNA splicing by changing the canonical acceptor splice site at the -1 position in intron 14. This variant has been reported previously as pathogenic for osteopetrosis. Finally, she is heterozygous at the *OSTM1* gene, for mutation NM_014028.3:c.143c>G (p.Ser48Trp). This allele has been noted previously, and has been noted at 0.83% in the African population. This is a missense mutation, likely to be benign. Computation tools (Align GVGD and MAPP) predict it to be benign, but SIFT predicts it to be deleterious. Ser and Trp have a large physiochemical difference, thus further computational analysis and correlation with other instances of *OSTM1* mutations is necessary to determine the extent to which the three mutations together could be exacerbating the pathologic phenotype of our patient.

2358F

Expanding the clinical and mutational spectrum of *B3GALT6*-associated Ehlers-Danlos syndrome and spondyloepimetaphyseal dysplasia with joint hypermobility type 1. *F. Malfait[1], T. Van Damme[1], X. Pang[2], B. Guille-myn[3], S. Gulberti[2], D. Syx[1], C. de Die-Smulders[3], R. Pfundt[4], A. Kariminejad[5], S. Nampoothiri[6], G. Pierquin[7], S. Bulk[7], A. Larson[8], S. Symoens[1], K. Chatfield[9], M. Simon[10], A. De Paepe[1], S. Fournel-Gigleux[2].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, OV, Belgium; 2) MolCelTEG Team, UMR 7365 CNRS-Université de Lorraine (IMoPA), Biopôle UL, Faculté de Méde-cine, Vandoeuvre-lès-Nancy, France; 3) Department of Clinical Genetics, Maastricht UMC, Maastricht, The Netherlands; 4) Department of Human Genetics, Radboud UMC, Nijmegen, The Netherlands; 5) Kariminejad-Na-jmabadi Pathology & Genetics Center, Tehran, Iran; 6) Department of Pediatric Genetics, Amrita Institute of Medical Sciences & Research Centre, Cochin, Kerala, India; 7) Service de Génétique Médicale, CHU liège, Liège, Belgium; 8) Department of Pediatrics, University of Colorado, Colorado, USA; 9) Department of Cardiology, University of Colorado, Colorado, USA; 10) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands.

Proteoglycans (PGs) are among the most abundant and structurally complex biomacromolecules, and are indispensable in many biological functions. They are composed of a core protein onto which one or more glycosaminoglycan (GAG) side chains are covalently attached via a tetrasaccharide linker region. Biallelic mutations in *B3GALT6*, encoding the linker region enzyme β3GalT6 or galactosyltransferase II, were recently shown to cause a spectrum of human Mendelian disorders ranging from spondyloepimetaphyseal dysplasia with joint hypermobility type 1 (SEMDJL1) to a severe pleiotropic Ehlers-Danlos syndrome (EDS) like disorder with joint hyperlaxity, progressive contractures, kyphoscoliosis, muscle hypotonia, skin and bone fragility and spondyloepi-metaphyseal dysplasia. We report on 11 novel patients (8 families) with either SEMDJL1 (4 families) or β3GalT6-deficient EDS (4 families), and expand the clinical spectrum to include severe and potential life-threatening complications, including atlanto-axial instability, aortic root dilatation, tracheomalacia, and respiratory dysfunction. Whole-exome sequencing or traditional Sanger sequencing identified biallelic *B3GALT6* mutations, mostly frameshift and missense mutations, in all families. Using immunofluorescence microscopy, we show that *B3GALT6* mutations reduce the amount of β3GalT6 protein in patient derived cultured dermal fibroblasts. In addition, high performance liquid chromatography showed a (near-) complete loss of galactosyltransferase activity towards the towards Galβ1-4Xyl(2-*O*-phosphate)-O-MethoxyNaphthyle substrate, and a decreased ability to prime GAG synthesis in patient-derived β3GalT6-deficient fibroblasts. We were unable however to correlate protein level or enzymatic activity with observed phenotypes. Finally, we observed impaired glycanation of the small chondroitin/dermatan sulfate PG decorin, and ultrastructural abnormalities in collagen fibril organization which empha-sizes the importance of GAG chains in this process. In conclusion, our study expands the molecular and clinical spectrum of B3GALT6-associated EDS and SEMDJL1.

2359W

Whole-exome sequencing identified novel compound heterozygous mutations in a Japanese patient with Geroderma Osteodysplastica. *R. Takeda[1], M. Takagi[2], H. Shinohara[3], S. Narumi[4], T. Hasegawa[5], H. Yoshihashi[1].* 1) Department of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 2) Department of Endocrinology and Metabolism, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 3) Department of Pediatrics, Ibaraki Seinan Medical Center, Ibaraki, Japan; 4) Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan; 5) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan.

[;Introduction];Geroderma Osteodysplastica (GO) is one of the subtypes of Cutis Laxa syndrome, which is characterized by congenital wrinkly skin, prematurely aged face, severe short stature, and osteoporosis leading to recurrent fractures. GO is inherited in an autosomal recessive pattern, and caused by loss of function mutation in GORAB gene that encodes a protein important for Golgi-related transport process. We report whole-exome sequencing identifies novel compound heterozygous nonsense mutation of the GORAB gene in a patient with GO. [;Case report]; The patient is 14 years old Japanese boy born through normal delivery at 38 weeks of gestation with a birth weight was 2,680g (-0.8 SD) and length was 50.6 cm (+0.8 SD). Wrinkle skin and joint laxity were present at birth. At 1 year of age, he was clinically diagnosed with Cutis Laxa syndrome based on recurrent long bone fracture and clinical features such as wrinkle skin, joint laxity, and distinctive face. His gross motor and cognitive development were not retarded. At 10 years of age, he was treated with oral bisphosphonate and vitamin D because of recurrent multiple spontaneous fractures of vertebral and extremities bone associated with lower bone mineral density (BMD). After the initiation of bisphosphonate treatment, his BMD and fracture rate were significantly improved. Whole-exome sequencing revealed two novel compound heterozygous nonsense mutations in the GORAB gene (p.R60*, p.Q124*), and established the diagnosis of GO.[;Discussion]; GO is a rare connective tissue disorder; with only 40 cases have been described in the literature so far, and this is the first report from Japan. There are few reports on the effects of the bisphosphonate treatment for GO patients with recurrent spontaneous fractures. We propose that oral bisphosphonate and vitamin D were effective and safe treatment option in the management of recurrent fractures with GO. It is important to make a precise diagnosis of GO to prevent recurrent fractures and for selecting better treatment.

2360T

Dominant cutis laxa with progeroid features due to a novel, *de novo* **mutation in** *ALDH18A1***.** *P.T. Bhola[1], S.N. Nikkel[2], T. Hartley[1], D.A. Dyment[1], K.M. Boycott[1], Care4Rare Consortium.* 1) Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Medical Genetics, BC Women's Hospital and Health Centre, Vancouver, British Columbia, Canada.

Autosomal recessive cutis laxa type 3, or De Barsy syndrome, is a rare condition characterized by cutis laxa, a progeroid appearance, ophthalmological abnormalities and intellectual disability. It is caused by biallelic mutations in the gene *PYCR1*, which encodes pyrroline-5-carboxylate reductase 1 or biallelic mutations in *ALDH18A1*, which encodes pyrroline-5-carboxylate synthase (P5CS). These gene products are mitochondrial enzymes that have a role in proline biosynthesis. It was recently recognized that there is a dominant form of De Barsy syndrome that is caused by *de novo* mutations in *ALDH18A1*. The individuals with *de novo* mutations all have a recurrent substitution at p.Arg138. We report the case of an 8 year old male who presented to Genetics with cutis laxa and suspected photophobia at 1 month of age. His medical history includes corneal clouding, cataracts, global developmental delay, bilateral adducted thumbs, bilateral congenital hip dislocation and inguinal hernias. The family is of Somali descent and the patient's parents are consanguineous. At age 8, cutis laxa and distinctive features were present and growth parameters of head circumference, height and weight were 5-6 standard deviations below the mean. Clinical genetic testing revealed a normal chromosomal SNP microarray. Due to the presumed recessive inheritance of the disorder and consanguinity in the family, the patient was enrolled in a whole exome research study at the Children's Hospital of Eastern Ontario, which also identified the heterozygous Arg126His mutation. No other pathogenic or likely pathogenic mutations were found. *In silico* analysis of this mutation predicted that it was deleterious. Segregation analysis in the parents by Sanger sequencing showed that the mutation was *de novo*. This report describes a novel dominant mutation in *ALDH18A1* in a patient with cutis laxa with progeroid features. As with the Arg138 mutation, the Arg126His mutation affects the gamma-glutamyl kinase domain of P5CS. Certain heterozygous mutations in A*LDH18A1* affecting the gamma-glutamyl kinase domain or the glutamyl-phosphate reductase domain of P5CS are causative for spastic paraplegia 9A, which is characterized by progressive spasticity. A greater understanding of dominant cutis laxa with progeroid features and of mutations in *ALDH18A1* will have implications for diagnosis and genetic counselling of these patients.

2361F

De novo missense variant in the *KIT* proto-oncogene in a child with piebaldism. *E. Tan, J.C. Nagaputra, E. Lim, H.W. Lim, M. Brett, M. Koh.* KK Women's & Children's Hospital, Bukit Timah Road, Singapore.

   Piebaldism is a rare congenital pigmentation disorder caused by the absence of melanocytes and melanin in certain areas of the skin. It is characterised by depigmented patches of skin and hair. Physical presentation include white forelock, hypopigmented patches on the face, trunk and limbs, often with islands of hyperpigmentation.   We report the findings in a Eurasian girl conceived via *in vitro* fertilization. Her mother is of Chinese ancestry while her father is German. She was born prematurely at 33 weeks of gestation via emergency caesarean section for IVF dichorionic diamniotic twins in labour. There was no family history of pigmentation disorders and her non-identical twin sister was normal. At birth, she was noted to have a white forelock with patches of depigmentation and multiple café-au-lait macules. When she was reviewed at 4 years of age, she had the white forelock with multiple well-demarcated, depigmented patch over the central forehead, neck and limbs. There were also multiple café-au-lait spots, and inguinal freckling. Her cognitive and physical development is normal for her age.   Genomic DNA was extracted from venous blood and sequenced using the TruSight One panel on the MiSeq System. Analysis revealed no pathogenic mutation for *SPRED1* and *NF1*, but a heterozygous missense mutation was detected in *KIT*. This variant results in the substitution of Leucine with Arginine in the intracellular tyrosine kinase domain of the *KIT* protein. It has not been reported in the Human Gene Mutation Database or in published literature. It is not found in the DNA from the saliva sample of her unaffected sister or the blood samples of her parents. This change is predicted to be pathogenic according to SIFT and Polyphen2. The phenotype of the patient is consistent with the diagnosis of piebaldism with depigmented patches and islands of normal/hyperpigmented skin. The presence of a pathogenic variant in the heterozygous state is consistent with the molecular genetics of this autosomal dominant disorder. Our report of this novel missense variant adds to the spectrum of pathogenic mutations in the *KIT* gene in patients with piebaldism.

2362W

New mutations and genotype/phenotype correlation in patients from consanguineous families from Saudi Arabia and Pakistan with different forms of autosomal recessive congenital ichthyosis. *D.Lima Cunha[1,2,3], O.M. Alakloby[4], K.M. Eckl[1,2,7], M. Rauch[2], R. Casper[3], N. Kakar[5], B. Krabichler[2], R. Gruber[2,6], J. Altmüller[3], P. Nürnberg[3], J. Zschocke[2], G. Borck[5], M. Schmuth[6], K.A. Alnutaifi[4], A.S Alabdulkareem[8], H.C. Hennies[1,2,3].* 1) Dept. of Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; 2) Div. of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 3) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 4) Dept. of Dermatology, University of Dammam, Dammam, Saudi Arabia; 5) Inst. of Human Genetics, University of Ulm, Ulm, Germany; 6) Dept. of Dermatology, Innsbruck Medical University, Innsbruck, Austria; 7) Dept. of Biology, Edge Hill University, Ormskirk, United Kingdom; 8) King Saud Medical City, Riyadh, Saudi Arabia.

   Autosomal recessive congenital ichthyosis (ARCI) is a genetically and phenotypically heterogeneous skin disease with varying features and degrees of severity. Several but not all genes with underlying mutations have been identified, but no clear correlation between the genetic causes and the clinical picture has been described to date. Our study included 13 families from Saudi Arabia and 5 from Pakistan. All patients were diagnosed with a form of ARCI and had a family history of consanguinity. We used homozygosity mapping as an initial screening tool and identified homozygous causal mutations in all patients by massively parallel and/or Sanger sequencing. Mutations were in five different genes and included several known and five unknown mutations: a splice site and a missense mutations in *TGM1*, a splice site mutation in *NIPAL4*, and missense mutations in both *ABCA12* and *CYP4F22*. New insights in phenotype/genotype correlations were accomplished, with *TGM1* mutations resulting in coarse dark brownish scaling, in contrast with *NIPAL4*, *ALOXE3* and *CYP4F22* patients, who presented with more localized, light-colored scales and partly erythematous skin. Our patients with *ABCA12* missense mutations, who were initially diagnosed with congenital ichthyosiform erythroderma, showed the most severe phenotype with generalized fine white scales and severe erythema. Patients with a "self-healing" phenotype, referred to as self-improving congenital ichthyosis, who had hardly any remaining signs of skin disease in early childhood, were also shown to have missense mutations in *TGM1*. Within the spectrum of ichthyosis phenotypes, we can attribute *TGM1* and *ABCA12* mutations to the most severe forms of lamellar and erythematous ichthyoses, respectively. The other mutations confirm the presence of a phenotypic spectrum, however, still lead to a rather lamellar course of the disease. This is in accordance with the assumption that these gene products are involved in ceramide metabolism of terminally differentiating keratinocytes, whereas ABCA12 contributes to transport of various lipids, including glucosylceramides, and inactivation of ABCA12 might result in intracellular accumulation of lipids such as gangliosides. Our findings contribute to expanding the mutational spectrum of ARCI and give further important insights into the genotype/phenotype correlation that is much needed for an easier and more precise diagnosis and the definition of targets for more specific therapies for ARCI.

2363T

Congenital contractural arachnodactyly: Delineation of clinical diagnostic criteria. *I. Meerschaut[1], S. De Coninck[1], S. Garcia Miñaur[2], J.C. Oosterwijk[3], R. Igbokwe[4], M. Suri[5], A. Bayat[6], G. Jones[7], C.I. Dali[6], S. Lynch[8], E. Blair[9], A. Collins[10], V. Laitinen[11], E. Thomas[12], A. Male[13], I. Stolte-Dijkstra[3], K. Keymolen[14], L. Cheryl[15], D. Yadav[16], F. Mckenzie[17], S. Berland[18], P. Willems[19], A. Topa[20], F. Petit[21], A. Destree[22], J. De Backer[1], P.J. Coucke[1], A. De Paepe[1], S. Symoens[1], B. Callewaert[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Clinical Genetics dpt, University Hospital la Paz, Madrid, Spain; 3) Dpt of genetics, University Medical Center Groningen, the Netherlands; 4) West Midlands Clinical Genetics Unit, Birmingham Women's NHS Foundation Trust, Birmingham, United Kingdom; 5) Nottingham Clinical Genetics Service Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom; 6) Clinical Genetic Clinic, Kennedy Center, Copenhagen University Hospital, Denmark; 7) Clinical Genetics Department, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom; 8) Department of clinical genetics, Our Lady's Children & Temple street Hospitals, Dublin, Ireland; 9) Oxford University Hospitals NHS Trust Churchill Hospital, Oxford, United Kingdom; 10) Wessex Clinical Genetics Service University Hospital Southampton NHS Foundation Trust Princess Anne Hospital, Southampton, United Kingdom; 11) Fimlab Laboratories, Tampere, Finland; 12) Genetics clinic, Borough Wing Guy's Hospital Great Maze Pond, London, United Kingdom; 13) 3Clinical Genetics dpt, Great Ormond street hospital, London, United Kingdom; 14) Center for Medical Genetics, Brussels University Hospital, Brussels, Belgium; 15) Clinical genetics dpt, NHS Greater Glasgow & Clyde, Glasgow, United Kingdom; 16) Dpt of Pediatrics, Jawaharlal Institute of Postgraduate Medical Education and Research, Pondicherry, India; 17) Genetic Services of Western Australia, Subiaco, Australia; 18) Center for Medical Genetics and Molecular Medicine Haukeland University Hospital, Bergen, Norway,; 19) Gendia, Antwerp, Belgium; 20) Dpt of Clinical Genetics, Sahlgrenska University Hospital, Göteborg, Sweden; 21) Dpt for clinical Genetics, CHU de Lille, Lille, France; 22) Center for Human Genetics, Institute for Pathology and Genetics, Gosselies, Belgium.

Congenital contractural arachnodactyly (CCA) is an autosomal dominant connective tissue disorder manifesting joint contractures, arachnodactyly, crumpled ears, and scoliosis as main features. Its rarity and substantial overlap with other conditions including Bethlem myopathy, Marfan syndrome and distal arthrogryposes, make the diagnosis challenging, though important for clinical management. CCA is caused by mutations in FBN2. We performed a comprehensive clinical and molecular assessment in a large cohort of CCA patients to delineate clinical diagnostic criteria and guide molecular analyses for FBN2. FBN2 analysis using either Sanger Sequencing or PCR-based next-generation sequencing was performed in 122 clinically well-characterized probands. We collected data on 10 main clinical characteristics, and determined the sensitivity, specificity, and positive and negative predictive value to find an FBN2 mutation for each feature in order to establish a weighted clinical scoring system on 20 points. Forty-seven probands harbored an FBN2 mutation (mutation uptake rate 39%). All but 2 mutations were located in the neonatal region (exons 22-36) with half of them altering or producing cysteines. Two patients carried inframe multi-exon deletions, one had a premature truncating mutation. Logistic regression analysis revealed a significantly higher clinical score in FBN2+ versus FBN2- patients (resp. 11,6 vs 7,7; p<0,001). ROC curve analysis revealed that a clinical score of 11+ yields a sensitivity of 84% and a specificity of 60% to find an FBN2 mutation. A score of 6 or lower is unlikely to be associated with an FBN2 mutation, unless in adult patients. In addition, review of cardiovascular features revealed non- or slowly progressive aortic root dilatation in 5 FBN2+ patients and 6 FBN2- patients. Though aortic root dilatation is a non-discriminative feature, its occurrence in CCA does warrant echocardiographic follow-up. .

2364F

Autosomal dominant adermatoglyphia is attributed to a 50kb deletion impacting *SMARCAD1. X. Chang, L. Tian, Y. Liu, R. Pellegrino, P. Sleiman, H. Hakonarson, L. Castelo-Soccio.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Adermatoglyphia (ADG) is an extremely rare genetic disorder, which causes a person to have no fingerprints. ADG with associated anomalies such as congenital facial milia and acral blisters, digital contractures, and nail abnormalities, is also described as Basan Baird Syndrome. All ADG-causing mutations identified to date are transversions, which disrupt a single conserved donor splice site adjacent to the 3' end of a noncoding exon (the first exon of the short isoform) and result in haploinsufficiency for a skin-specific isoform (short isoform) of *SMARCAD1*. Here we report on a family trio where the father and child were affected with ADG and associated anomalies including bilateral webbing of toes, bilateral contractures. The father and child also exhibited milia on the face and blistering of the feet in the neonatal period consistent with what has been reported as Basan Baird Syndrome. We sequenced the blood DNA samples of this family by whole exome sequencing (WES). No previously reported mutations were found in any members of this family though an additional family members are not affected . However, a heterozygous deletion covering the first 10 exons of the long isoform and the first exon of the short isoform in *SMARCAD1* was detected in both the affected father and child by XHMM. The deletion was further confirmed by PCR validation. In comparison with the four ADG families previously reported, our ADG cases also displayed additional phenotypes, including but not limited to bilateral webbing of the toes, bilateral contractures and milia, suggesting that the disruption of the long isoform may be associated with those phenotypes. In addition, the long isoform of *SMARCAD1* has been shown to regulate the expression of transcriptional factors, histone modifiers and development-associated genes. This is the first report of an ADG family caused by a copy number variation.

2365W

Protein replacement therapy for autosomal recessive congenital ichthyosis (ARCI). *H.C. Hennies[1,2,3], R. Plank[1,2], G. Yealland[4], R. Casper[2], K. Obst[4], M. Hermann[5], E. Miceli[6], M. Calderón[6], S. Hedtrich[4], K.M. Eckl[1,2,7].* 1) Dept. of Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; 2) Div. of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 3) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 4) Inst. of Pharmacy, Freie Universität Berlin, Berlin, Germany; 5) Dept. of Anaestesiology, Innsbruck Medical University, Innsbruck, Austria; 6) Inst. for Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany; 7) Dept. for Biology, Edge Hill University, Ormskirk, Lancashire, UK.

   Autosomal recessive congenital ichthyosis (ARCI) is a rare, genetically heterogeneous cornification disorder characterized by impaired skin barrier function, with no causative therapy available to date. Skin diseases caused by inactivating mutations or lack of protein synthesis are considered favourable targets for local protein substitution. Unpacked therapeutic proteins, however, are usually unable to penetrate the epidermal permeability barrier, owing to their high molecular weight and hydrophilicity. Here we aim for a novel, personalized ARCI therapy in which defective transglutaminase 1 (TGase1) is exogenously substituted with the help of nanotransporters developed for protein delivery into the skin for topical application. We have synthesized mammalian TGase1 in HEK 293 cells. The protein was assessed by flow cytometry and Western blot analysis, and enzyme activity was confirmed by *in vitro* assays. Next we coupled TGase1 to thermoresponsive polyglycerol-based nanogels (dPG-NG) that contain poly(N-isopropylacrylamide) as thermoresponsive linker. The protein-loaded nanogels exhibited a thermal trigger point at 35°C, favourable for cutaneous applications. Staining experiments showed that TGase1 was successfully delivered into keratinocytes. Analysis over time revealed nearly all extrinsic protein was gone 24h after delivery, therefore we defined a treatment schedule with repeated applications of dPG-NG/TGase1. 3D full-thickness skin models, which exhibited functional barrier properties comparable to the epidermal barrier of the skin, were generated as disease models for therapeutic applications with keratinocytes and fibroblasts from ARCI patients or keratinocytes where *TGM1* was knocked-down. As expected, they showed a clearly reduced barrier activity and an increased permeabilitiy. After treatment with dPG-NG/TGase1, TGase1 was present in suprabasal epidermal layers, TGase1 activity was confirmed in the models with *in situ* assays, and permeability assays showed that the barrier dysfunction of treated ARCI models was significantly ameliorated. Thus, our findings revealed an advanced topical drug delivery system suitable for cutaneous protein replacement as a promising approach for causative therapeutic intervention in ARCI. The approach is being optimised for epidermal delivery and protein dosage. After toxicity tests will be accomplished, we want to adapt the system for use with other proteins involved in autosomal recessive cornification disorders.

2366T

Homozygous *ATP6V1E1* and *ATP6V1A* mutations in metabolic cutis laxa syndromes: New clinical and pathogenetic insights. *B. Callewaert[1], T. Van Damme[1], M. Mohamed[2], T. Gardeitchik[2], A. Kariminejad[3], B. Guillemyn[1], W. Steyaert[1], S. Ghaderi-Sohi[3], D. Syx[1], S. Wong[4], D. Lefeber[2], P.J. Coucke[1], S. van Kraaij[2], D. Dalloyaux[2], R. De Rycke[5], A. De Paepe[1], S. Symoens[1], T. Hack[6,7], P. Freisinger[8], E. Morava[4], F. Malfait[1], R. Wevers[2].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Departments of Laboratory medicine, Pediatrics and Genetics, Radboudumc, Nijmegen, The Netherlands; 3) Kariminejad-Najmabadi Pathology & Genetics Center, Tehran, Iran; 4) Hayward Genetics Center, Tulane University School of Medicine, New Orleans, LA; 5) Department of Biomedical Molecular Biology, Inflammation Research Center, Ghent University, Ghent, Belgium; 6) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 7) Institute of Human Genetics, Technische Universität München, Munich, Germany; 8) Department of Pediatrics, Kreisklinken Reutlingen, Reutlingen, Germany.

   Defects of the vacuolar-type H+-ATPase pump (V-ATPase) impair acidification of vesicular compartments and interfere with intracellular trafficking of membrane-bound compartments including secretory granules, endosomes, and lysosomes. Mutations in ATP6V0A2, encoding the a2 subunit of the V0 domain cause autosomal recessive cutis laxa type 2, a congenital disorder of glycosylation with multisystem manifestations including a loose redundant skin, skeletal, and often neurological abnormalities. We report on mutations in two subunits of the V1 domain of the V-ATPase complex in four independent cutis laxa families. In addition to generalized cutis laxa, these patients presented with similar dysmorphic facial features, hypotonia, a marfanoid habitus, and variable cardiopulmonary involvement (ventricular hypertrophy, aortic root dilatation, and pneumothorax). Whole-exome sequencing identified homozygous missense mutations in ATP6V1E1 (c.383T>C, p.(Leu128Pro); c.634C>T, p.(Arg212Trp)) and ATP6V1A (c.1012C>T, p.(Arg338Cys); and c.215G>A, p.(Gly72Asp)), encoding respectively the E1 and A subunitof the V1 domain. Structural modeling using the S. cerevisiae V-ATPase as template indicated that all substitutions affect critical residues, and alter either inter- or intrasubunit interactions. V-ATPase complex profiling using liquid chromatography tandem mass spectrometry showed a reduction in the amount of assembled V1 domain in patient samples, indicating that substitutions in the E1 and A subunits affect either the assembly or the stability of the V1 part of the complex. Transferrin isoelectric focusing showed a type II glycosylation defect in patients with ATP6V1E1 and ATPP6V1A mutations. Vesicular trafficking defects were evidenced by delayed retrograde translocation of Golgi membranes to the endoplasmic reticulum after Brefeldin A treatment, and abnormal swelling and fragmentation of the Golgi apparatus in patient derived cultured dermal fibroblasts on transmission electron microscopy (TEM). In addition, we observed lysosomal storage bodies on a dermal skin sample of one of the affected patients. Finally, TEM of the patient dermis showed severe changes in the amount, structure and organization of elastic and collagen fibers. In conclusion, our study expands the molecular and clinical spectrum of metabolic cutis laxa syndromes, and provides new insights on the cellular processes involved in proper assembly and homeostasis of the extracellular matrix.

2367F

Increased expression of mutant elastin alleles in autosomal dominant cutis laxa is normalized by transforming growth factor beta treatment. *S. Akcay, E.C. Lawrence, Z. Urban.* Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Increased expression of mutant alleles in autosomal dominant diseases is rarely observed, but is generally found to result in more severe presentation, as observed in Huntington's disease and familial hypercholesterolemia. Examining the relative expression of mutant and wild type elastin (*ELN*) alleles in skin fibroblasts from 7 individuals with autosomal dominant cutis laxa (ADCL), with 4 distinct frameshift mutations, we observed 1.5-2-fold increased expression of mutant alleles at the mRNA level. Treatment of mutant fibroblasts with 5 ng/mL transforming growth factor beta 1 (TGFB1) for 10 days yielded an increase in the expression of the wild type alleles but a decrease in the expression of the mutant alleles, abolishing the differences in allelic expression. TGFB1 treatment also normalized the amount of insoluble elastin deposited by ADCL cells. The mechanism of differential regulation of mutant and wild type mRNA levels by TGFB1 is likely related to altered accessibility of the TGFB1-target miR29 to mutant and wild type *ELN* transcripts. Our studies implicate increased mutant allele expression as a contributor to the molecular mechanism of ADCL, and support TGFB1 augmentation as a possible therapeutic approach.

2368W

Identification of a novel mutation in *FBN1* gene in an infant with Marfan syndrome. *B. Lazalde[1,2], G. Aguirre-Hernández[1], R.M. González-Arreola[1], J.L. Venegas-Rodríguez[1], A.E. González-Font[3].* 1) Department of Genetics, Faculty of Medicine and Nutrition, Juárez University of Durango State, Durango, Dgo., Mexico; 2) Biomedical Reserach Unit, Mexican Institute of Social Security. Durango, Dgo., Mexico; 3) Department of Cardiology, Maternal and Child Hospital of Durango. Durango, Dgo., Mexico.

Aim: To report a case of an infant with Marfan syndrome (MFS) (MIM 154700) presenting a novel mutation in *FBN1* gene. **Case report:** A female infant was born at 42 weeks of gestation to a 31 years old (gravida 4, parity 2) healthy mother and 33 years old healthy and not related father. At 18 months of age a clinical survey revealed an IV-VI grade holosystolic heart murmur, and the echocardiogram confirmed a moderate to severe mitral insufficiency with mitral valve prolapsed. The physical examination at two years of age in the genetics service, showed proportionate tall stature, height 94 cm (>97th percentile), weight 16.3 kg (25-50th percentile), upper/lower segment ratio of 1.19 and armspam/height ratio of 1.04, dolicocephaly, long shaped facies, downslanting palpebral fissures, malar hypoplasia, high palate, anteverted auricular pavilions, pectus carinatum, scoliosis, articular hyperelasticity, long fingers and toes, and hindfoot deformity. Ophthalmologic evaluation revealed myopia and excluded ectopia lentis. A fibrillinophaty was suspected and molecular diagnostic tests of the *TGFBR1*, *TGFBR2* and *FBN1* genes were conducted. **Results:** The gene Sanger sequenciation revealed a heterozygous duplication of 4 nucleotides (c.5662_5665dupATGT) in *FBN1* gene, that creates a shift in the reading frame at codon 1889 and the new reading frame ends in a stop codon 6 positions downstream (p.Cys1889Tyrfs*7), which is very likely to result in a truncated protein. This variant was absent in both parents and has not been described previously associated to MFS or other fibrillinopathy. Last echocardiogram performed at the age of 4 years showed a dilated aortic root (24mm, Z value ≥+3) and prolapsed mitral valve with severe regurgitation. **Discussion:** Although several conditions have been recognized which present overlapping clinical characteristics associated to *FBN1* mutations, according to revised Ghent nosology (positivity of two out of four criteria in patients without a family history), the patient fit the criteria for classic MFS because she has a systemic score of 7, aortic dilatation with a Z value ≥+3, and a nonsense mutation in *FBN1* gene. **Conclusion**: The novel mutation c.5662_5665dupATGT in *FBN1* gene is a putative disease causing mutation associated to a MFS. .

2369T

Autosomal Recessive Clouston Syndrome attributable to a missense mutation in *GJB6.* M.O. Cevik[1,2,3], I.H. Bucak[4], Q. Shao[5], M. Karatas[6], Y. Sarikaya[6], F. Dogan[7], S.G. Cevik[8], S. Bilak[9], M. Erzurumluoglu[10], A.Y. Huang[1,2], S. Hann[2], H. Bagis[3], D.W. Laird[5], M. Turgut[4]. 1) Harvard Medical School Department of Genetics 77 Avenue Louis Pasteur NRB 0330 Boston, MA 02115 USA; 2) Boston Children's Hospital Department of Orthopedic Surgery & Research 300 Longwood Ave., EN260 Boston, MA 02115 USA; 3) Adiyaman University School of Medicine Department of Medical Genetics 02040 Adiyaman Turkey; 4) Adiyaman University School of Medicine Department of Pediatrics 02040 Adiyaman Turkey; 5) Department of Anatomy and Cell Biology University of Western Ontario Dental Science Building, Rm 00077 London, Ontario, Canada, N6A-5C1; 6) Adiyaman University School of Medicine Department of Otorhinolaryngology 02040 Adiyaman Turkey; 7) Adiyaman University School of Medicine Department of Plastic and Reconstructive Surgery 02040 Adiyaman Turkey; 8) Şevket Yilmaz Training and Research Hospital, Clinic of Ophthalmology, 16310 Bursa Turkey; 9) Adiyaman University School of Medicine Department of Ophtalmology 02040 Adiyaman Turkey; 10) Department of Health Sciences University of Leicester Adrian Building, University Road Leicester, LE1 7RH UK.

Ectodermal dysplasias (EDs) are a complex and genetically heterogeneous group of conditions that affect the development of tissues derived from early embryonic ectoderm. Clouston syndrome, also known as hidrotic ectodermal dysplasia type-2 (HED2, OMIM#129500), is an autosomal dominant ED characterized by congenital or childhood-onset partial to total hair loss, abnormal nails, and palmar/plantar dyskeratosis. In this paper, a seven year old Turkish girl who were born to a consanguineous family and had Clouston disease with a recessively acting, homozygous and novel *GJB6* mutation. Our data strongly implicate p.G45R in *GJB6* as the cause of an autosomal recessive form of Clouston syndrome. Thus, while most patients with Clouston syndrome have dominantly acting mutations, recessive inheritance should not be precluded especially in the presence of consanguinity.

2370F

Identification of a novel homozygous nonsense mutation of *CAST* **gene in a PLACK family.** S. Temel[1,2], H. Sarcaoğlu[3], B. Türkgenç[4,5], U. Kıran[3], B. Ergüner[6], B. Yücetürk[6], M. Sağıroğlu[6], M.C Yakıcıer[7]. 1) University of Near East, Faculty of Medicine, Department of Histology & Embryology, Lefkosia, Cyprus; 2) University of Uludag, Faculty of Medicine, Department of Histology & Embryology, Bursa, Turkey; 3) University of Uludag, Faculty of Medicine, Department of Dermatology, Bursa, Turkey,; 4) 4Marmara University, Faculty of Medicine, Department of Medical Biology and Genetics, Istanbul, Turkey; 5) University of Acibadem, Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 6) Advanced Genomics and Bioinformatics Research Group (İG-BAM) TÜBİTAK BİLGEM UEKAE, Gebze/Kocaeli, Turkey,; 7) 7University of Acibadem, Faculty of Arts and Sciences, Department of Molecular Biology and Genetics,, Istanbul, Turkey.

Peeling skin syndrome (PSS) is characterized by continuous shedding of the stratum corneum of the epidermis with onset from birth or infancy and lasting throughout life. Skin peeling can be accompanied by erythema, vesicular lesions, or other ectodermal features including fragile hair and nail abnormalities. PSS can be divided into acral and generalized PSS. We report a 5.5 year old boy who initially presented with skin fragility. He was the first-born of IVF twins from a consanguineous marriage at 33+5 week gestational age with hypotonia at birth. The main clinical findings included fragile skin, woolly hair, sparse eyelashes and brows, palmoplantar punctate keratoderma, follicular hyperkeratosis, knuckle pads and cheilitis. Moreover mild cerebral atrophy and mild muscle involvement were observed on his MRI and NCV/EMG, respectively. His aunt also had similar clinical features but in milder form with in addition nail dystrophy. Exome sequencing revealed a homozygous c.544G>T (p.Glu182*) nonsense mutation in the *CAST* gene. The segregation of this rare variant in the family was confirmed by Sanger sequencing. This novel stop-gain *E182X* variant produces a truncated protein lacking inhibitory domains II-IV. CAST is an endogenous specific inhibitor of calpain, a calcium-dependent cysteine protease. Recently, autosomal recessive loss of function mutations in CAST were described in PLACK syndrome characterized by generalized peeling skin, leukonychia, acral punctate keratoses, cheilitis, and knuckle pads. As far as we know our case is the fifth case of PLACK syndrome without leukonychia but with some additional previously unreported associated features including mild cerebral atrophy and muscle involvement. .

2371W

Periodontal Ehlers-Danlos syndrome is caused by gain-of-function mutations in complement 1 subunits C1r or C1s. *J. Zschocke[1], I. Kapfer-er-Seebacher[2], M. Pepin[3], R. Werner[1], T. Aitman[4,19], H. Stoiber[5], N. Thielens[6], C. Gaboriaud[6], A. Amberger[1], A. Schossig[1], R. Gruber[1,7], C. Giunta[8], M. Bamshad[9,10], D. Chitayat[11], M. Dorschner[3], C.J. Hale[3], D. Hanna[3], H.C. Hennies[1], A.L. Mitchell[12], E. Reinstein[13], M. Rohrbach[8], N. Romani[7], M. Schmuth[7], A. Vandersteen[14], J. Vandrovcova[15], R. Weerakkody[16], M. Yang[3], F.M. Pope[17,18], P.H. Byers[3], Molecular Basis of Periodontal EDS Consortium.* 1) Div Human Genetics, Med Univ Innsbruck, Innsbruck, Austria; 2) Department of Operative and Restorative Dentistry, Medical University of Innsbruck, Innsbruck, Austria; 3) Departments of Pathology and Medicine (Medical Genetics), Collagen Diagnostic Laboratory, University of Washington, Seattle, USA; 4) MRC Clinical Sciences Centre and Department of Medicine, Imperial College London; 5) Division of Virology, Medical University of Innsbruck, Austria; 6) Institut de Biologie Structurale (IBS), Univ. Grenoble-Alpes, CEA, CNRS, 38044 Grenoble, France; 7) Department of Dermatology, Venereology and Allergology, Medical University of Innsbruck, Innsbruck, Austria; 8) Connective Tissue Unit, Division of Metabolism and Children´s Research Centre (CRC), University Children's Hospital, Zurich, Switzerland; 9) Department of Pediatrics, University of Washington, Seattle, USA; 10) Department of Genome Sciences, University of Washington, Seattle, USA; 11) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 12) Departments of Genetics and Genome Sciences and Pediatrics, Case Western Reserve; 13) Medical Genetics Institute, Meir Medical Center, Kfar Saba, Israel; 14) Department of Pediatrics, IWK Health Centre, Halifax, Canada; 15) King's College London, Department of Medical & Molecular Genetics, Guy's Hospital, London,; 16) Department of Surgery and Cancer, Imperial College London, London, UK; 17) West Middlesex University Hospital, Isleworth, Middlesex, UK; 18) Hospital of St John & St Elizabeth, London, UK; 19) Institute of Genetics and Molecular Medicine, University of Edinburgh, UK.

Periodontal Ehlers-Danlos syndrome (EDS) is an autosomal dominant disorder characterized by aggressive periodontitis and cutaneous tissue fragility (also called EDS VIII). A locus was mapped to 12p13.1 but no candidate gene identified. Whole exome sequencing in 7 families with the clinical phenotype allowed identification of different mutations in *C1R* (5 families) and *C1S* (2 families). Further analysis by capillary DNA sequencing in an additional 10 families (total of 93 individuals) identified additional mutations in *C1R*. The C1R and C1S genes are similar in structure and sequence and are located in tandem at 12p13.1. We identified heterozygous missense or in-frame insertion/deletion variants in *C1R* or *C1S,* which encode the C1r and C1s subunits of the first component of complement. The variants we identified involve residues essential to maintain the C1r/C1s, C1r/C1r or C1r/C1q interfaces or the hinges between interaction and catalytic domains of C1r and C1s. Overexpression of *C1R* missense mutants in HEK293 cells showed intracellular retention of C1r and mild endoplasmic reticulum enlargement similar to that seen in patient fibroblasts in skin. The clinical features observed in these families included aggressive periodontitis, structural gingival abnormalities, skin fragility, easy bruising, pretibial hyperpigmentation, and musculoskeletal symptoms. None had the clinical features of lupus that are associated with homozygous loss of either or both *C1R* and *C1S* and also did not have the findings of angioedema seen with loss of the C1 esterase inhibitor. These findings point to a gain of function effect. The proteases C1r and C1s form a proenzyme tetramer that associates with the collagen-like stems of C1q, are sequentially activated to form proteases that, in turn, activate the classical complement pathway. *C1R* and *C1S* variants in periodontal EDS likely have a gain-of-function effect, may interact with alternative targets, and/or interfere with local immune functions. Our findings define periodontal EDS as a genetically distinct EDS subtype, caused by mutations in *C1R* and *C1S*, and open new ways to understand relationships between inflammatory pathways and connective tissue homeostasis.

2372T

Clinical and molecular description in a patient with Keratitis-ichthyosis-deafness syndrome presenting a severe phenotype. *V. Cortés-González[1], A. Serrano-Ahumada[2], L. González-Huerta[3], C. Villanueva-Mendoza[1].* 1) Genetics Department, Asociación para Evitar la Ceguera en México, Hospital Dr. Luis Sánchez Bulnes; 2) Ophthalmology Department, Asociación para Evitar la Ceguera en México, Hospital Dr. Luis Sánchez Bulnes; 3) Unit Research of Human Genetics Department, Hospital General de México, Dr. Eduardo Liceaga.

Keratitis-ichthyosis-deafness (OMIM #148210) is a rare autosomal dominant ectodermal dysplasia syndrome. Cardinal features of the phenotype include sensorineural deafness, erythrokeratoderma, vascularizing keratitis, alopecia, and reticulated hyperkeratosis of the palms and soles. This entity is frequently due to mutation in a gene encoding connexin 26 protein affecting certain tissues of ectodermal origin such as epidermis, cochlea, and cornea. The most common pathological genetic variant of this connexin 26 is p.Asp50Asn, leading to substitution of conserved residues in the cytoplasm amino terminus or first extracellular domain which is crucial for voltage gating and connexon-connexon interactions. We present the systemic and ocular findings in a patient with Keratitis-ichthyosis-deafness (KID) syndrome, with contractures, skeletal handicap and an ocular neoplasia. The proband is a 30-year-old male with decreased visual acuity, blurry vision, eye redness, tearing and photophobia. He had history of congenital deafness; his skin was seborrheic at birth and then became erythematous and thick; by 10 years he developed contractures and deambulation was limited, now he is confined to a wheelchair. Ophthalmic evaluation demonstrated best corrected visual acuity of 20/300 in the right eye and 20/250 in the left eye; he had madarosis, hyperkeratotic lesions of the eyelids, bilateral and total corneal neovascularization, and in the right eye an intraepithelial lesion suggestive of an ocular squamous neoplasia. Systemic examination demonstrated alopecia, generalized dry and erythematous skin, severe hyperkeratotic plaques especially at pressure sites as buttocks and talus, kyphoscoliosis, flexion contractures and dystrophic nails. His parents were healthy and there was no relevant family history. With this phenotype it was considered the diagnosis of KID syndrome. Molecular analysis by PCR amplification and automated DNA sequencing of the complete *GJB2* gene (OMIM *121011) coding sequence revealed an apparently the novo heterozygous mutation c.148G>A (p.Asp50Asn). KID patients have an increased risk of infections and the occurrence of squamous cell carcinomas. This syndrome illustrates that gap junction communication plays a role in immune response and epidermal carcinogenesis.

2373F

Functional validation of a novel missense variant in *TGFBR2* **in a patient with aortic dilatation and aneurysms.** *M.A. Cousin[1,2], A.J. Mathison[3], M.T. Zimmermann[1], P.R. Blackburn[1,2], N.J. Boczek[1,2], G.R. Oliver[1], G.A. Lomberk[3], D.R. Deyle[2,4], R.A. Urrutia[2,3], E.W. Klee[1,2,4].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 3) Department of Medicine, Gastroenterology Research Unit, Mayo Clinic, Rochester, MN; 4) Department of Clinic Genomics, Mayo Clinic, Rochester, MN.

Heritable connective tissue disorders characterized by cardiovascular defects including aortic dilatation and aneurysms often involve dysregulated TGF-β signaling, have overlapping symptoms, and have variable penetrance and expressivity. Panel-based testing is a highly utilized approach to identify the disease-causing genetic variant(s) in patients with a suspected genetic disorder. Unfortunately, the variants they identify frequently have uncertain clinical significance thereby obscuring the genetic diagnosis until more information becomes available or appropriate functional studies resolve the pathogenicity of the variant(s) in question. Here we describe a patient with a suspected connective tissue disorder harboring a variant of uncertain significance (VUS) in the *TGFBR2* gene and provide a detailed account of the subsequent functional studies used to establish its pathogenicity. A 197.7 cm tall 43-year-old man presented to Mayo Clinic's Department of Clinical Genomics with a history of ascending aorta dilatation and aneurysms in the common iliac arteries, all requiring surgical repair. Following examination, his Marfan systemic score was 9 indicating a clinical diagnosis of Marfan syndrome. He has no family history of aortic dilatation or aneurysm, but some reports of joint pain, teeth crowding, flat feet, and varicose veins. Clinical gene-panel testing identified a c.1255G>T; p.Val419Leu VUS in *TGFBR2*. We assessed the structural and functional consequence of this VUS using *in silico* protein modeling and *in vitro* cell-based assays. A high-quality homology-based model of TGFβR2 was generated and computational mutagenesis followed by refinement and Molecular Dynamics (MD) simulations was used to assess structural and dynamic changes. Relative to WT, the V419L induced conformational and dynamic changes that may affect ATP binding. Experimentally, we measured the canonical TGF-β signaling pathway activation at two points. The V419L significantly delayed SMAD2 phosphorylation by western blot and reporter assays showed decreased TGF-β-induced gene transcription. Interpreting these functional results in light of our MD simulations, we believe that V419L leads to a dynamic alteration of the ligand-binding site, which results in decreased enzymatic activity, decreased TGF-β signaling, and decreased downstream gene transcription. The aberrant TGF-β signaling leads to Loeys-Dietz syndrome 2 (OMIM:610168) and the clinical presentation in our patient.

2374W

Types of albinism in southern Africa and genetic investigation of pigment-related genes. *R.A. Kerr, J.G.R. Kromberg.* Division of Human Genetics, University of the Witwatersrand and NHLS, Johannesburg, Gauteng, South Africa.

The first aim of this study was to describe and classify the different types of albinism found in the local southern African black population. Secondly, a mutation screen was undertaken to identify disease-causing mutations in a group of black OCA subjects. All subjects were initially screened for the common black *OCA2* [MIM 611409] mutation: the 2.7 kb deletion mutation. Subjects where one or two mutations remained to be identified were included in an *OCA2* mutation screen (N=63). Certain individuals with "unclassified" OCA were investigated further. Classification of types involved a descriptive survey in which 96 affected individuals underwent a clinical and/or dermatological examination. The *OCA2* genetic screen involved Sanger sequencing of all 25 exons of the gene. Certain individuals were also investigated at the *Tyrosinase* (*TYR* [MIM 606933]) or *Tyrosinase-Related Protein 1* (*TYRP1* [115501]) locus. Most (79/96; 82%) subjects had classic OCA2 [MIM 203200] and a further 11/96 (12%) had the brown oculocutaneous albinism (BOCA [MIM 203200]) subtype. The remainder of the subjects (6/96; 6%) had rufous oculocutaneous albinism (OCA3 or ROCA [MIM 203290]), caused by mutations in the *TYRP1* gene. In the black OCA2 patients, four mutations which are likely to be pathogenic were found. No further mutations (besides the 2.7kb deletion) were identified in the BOCA sample. In conclusion, the most common types of OCA found locally are OCA2 and OCA3. No patients with OCA1 were described. Besides the 2.7kb deletion, no other common *OCA2* mutation has been found. Approximately 20% of mutations remain unidentified. Genotyping of particular unusual cases suggests that variation at more than one pigment locus is responsible for the resulting phenotype.

2375T

Males with Incontinentia Pigmenti: Somatic and germ line mosaicism for the NEMO/IKBKG mutation underline the risk of recurrence. *F. Fusco[1], A. Diociauti[2], S. Bigoni[3], R. Sabbatella[1], A. Ferlini[3], M. El Hachem[2], M.B. Lioi[4], M.V. Ursini[1].* 1) Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso', IGB-CNR,, Naples, Italy; 2) Dermatology Unit, Bambino Gesù Children's Hospital- IRCCS, Rome, Italy; 3) UOL of Medical Genetics, Ferrara University Hospital, Ferrara, Italy; 4) University of Basilicata, Potenza, Italy.

The Incontinentia Pigmenti (IP; OMIM#308300) is an X-linked dominant disease lethal in males and therefore the only expected type of transmission is from the IP mother to the daughter. However, some IP male cases have been reported. The discovery of *NEMO/IKBKG* as the causative gene has enabled the molecular genetic confirmation of IP in mosaic male. All the IP-males were isolated cases presenting a postzygotic genetic mosaicism that produces a cellular condition comparable to that of heterozygous female, who are functional mosaics due to random X-inactivation. Amongst the 421 IP cases that comprises our cohort (IPGB, http://www.igb.cnr.it/ipgb) there are currently 40 IP male, 38 are teenagers and two presented an unexpected father-to-daughter transmission for an X-linked male lethal disease. We demonstrated that the underlining cause of such transmission is the somatic and germ line mosaicism for the NEMO/IKBKG mutation in the IP fathers of the two families. The father of Family 1 presented classical IP, with typical cutaneous rush and epilepsy and has one IP affected daughter. They both carried the new p.Gln132X mutation. The father of Family 2 with a mild form of IP has two IP affected daughters. They all carry the common exon4-10 deletion of *NEMO*. The levels of mosaicism was 20% of *NEMO* mutated cells in the affected skin, 8.3% in the urine and 16.7% in germ line in father of Family 1 and 25% of *NEMO* mutated cells in the urine and 35% in germ line in father of Family 2. Those are the first molecularly documented evidences of somatic and germline mosaicism for the NEMO/IKBKG mutation in IP cases that implicate a risk for the disease to be transmitted from the father to multiple offspring.

2376F

Postzygotic *KITLG* mutation in a congenital non-progressive linear nevoid hyperpigmentation. *A. Sorlin[1], A. Maruani[2], J.-B. Rivière[1,3,4], Y. Duffourd[1,3], P. Kuentz[1,3,4], J. St-Onge[1], M. Chevarin[1], T. Jouan[1], C. Thauvin-Robinet[1,3,5], J. Thevenon[1,3,5], L. Faivre[1,3,5], P. Vabres[1,3,5,6].* 1) Laboratoire Génétique des Anomalies du Développement, Dijon, France; 2) Service de Dermatologie, Tours University Hospital, Université François Rabelais, France; 3) FHU TRANSLAD, Dijon University Hospital and Université Bourgogne-Franche Comté University, Dijon, France; 4) Molecular Biology, Dijon University Hospital, Dijon, France; 5) Centre de référence Anomalies du Développement et Syndromes Malformatifs, Dijon University Hospital, Dijon, France; 6) Dermatology, Dijon University Hospital, Dijon, France.

Linear and whorled naevoid hypermelanosis (LWNH) - hyperpigmented macular swirls and streaks following Blaschko's lines - has often been considered as a nonspecific manifestation of mosaicism. It has sometimes been mistaken with the pigmentary stage of incontinentia pigmenti. In a few patients, various X chromosome rearrangements have been reported, but the molecular basis of LWNH has remained unknown. We performed deep exome sequencing on skin DNA from a 6 year-old patient with congenital non-progressive linear naevoid hyperpigmentation following Blaschko's lines on his trunk and limbs, without other cutaneous or neurosensory symptoms. We identified a postzygotic heterozygous *KITLG* c.329A>G (p.Asp110Gly) mutation, confirmed by targeted deep sequencing in skin fibroblasts (28% of reads) and blood (18%). *KITLG* (c-KIT Ligand, also known as Stem Cell Factor) regulates skin pigmentation through control of melanocyte migration, proliferation and survival, and melanin synthesis. Germline *KITLG* mutations have been reported in patients with Familial Progressive Hyper-and Hypopigmentation (FPHH), who exhibit early-onset hyperpigmented macules, increasing in size and number until adulthood, and hypopigmented macules. *KITLG* mutations have also recently been found in patients with isolated hearing loss or Waardenburg syndrome type 2A (congenital hearing loss, pigmentary abnormalities of the hair, skin, and eyes). This is the first report of a genetic basis for LWNH, which can now - at least in this patient - be considered as a mosaic presentation of a Mendelian disorder, FPHH The p.Asp110Gly mutation is also the first reported mosaic mutation of *KITLG*, in a patient with normal hearing, extending the spectrum of clinical manifestations associated with mutations in *KITLG*.

2377W

Postzygotic *MTOR* activating mutations in hypomelanosis of Ito with brain overgrowth. *P. Vabres[1], V. Parker[2], J. St-Onge[1], Y. Duffourd[3], D. Rodriguez[4], C. Mignot[4], R. Knox[2], A. Boland[5], R. Olaso[5], M. Delepine[5], V. Darmency[1], C. Vincent-Delorme[6], B. Catteau[6], L. Guibaud[7], A. Arzimanoglou[7], M. Keddar[1], P. Callier[1], D. Bessis[8], D. Geneviève[8], J.F. Deleuze[5], R. Semple[2], L. Faivre[1], J.B. Rivière[2].* 1) CHU Dijon Bourgogne, DIJON, France; 2) Institute of Metabolic Science, Cambridge, UK; 3) Université de Bourgogne Franche-Comté, Dijon, France; 4) Assistance Publique – Hôpitaux de Paris, Paris, France; 5) 9Centre National de Génotypage, Evry, France; 6) Centre Hospitalo-Universitaire, Lille, France; 7) Hospices Civils de Lyon, Bron, France; 8) Centre Hospitalo-Universitaire, Montpellier, France.

   Hypomelanosis of Ito consists of hypopigmentation along Blaschko's lines (naevus depigmentosus) associated with neurological features. Except for inconstant mosaic chromosomal anomalies, its genetic basis had remained elusive. Recently, postzygotic *MTOR* mutations have been identified. We performed whole-exome and targeted deep sequencing from skin and blood-derived DNA in four unrelated patients with extensive nevus depigmentosus or patchy depigmentation, brain overgrowth, intellectual disability and epilepsy. We identified mosaic activating missense *MTOR* mutations on skin biopsy, usually absent in the blood in all patients. One had severe cognitive impairment, intractable seizures, and carried an *MTOR* p.Glu2419Lys variant. Analysis of her dermal fibroblasts showed upregulation of mTOR signaling, with increased phosphorylation of AKT and p70 ribosomal protein S6 kinase. Since this mTOR mutation was sensitive to rapamycin *in vitro* and *in vivo*, we initiated everolimus therapy, which however failed to improve seizures. Hypopigmentation may specifically result from mTOR activation, as in tuberous sclerosis complex. Despite failure to improve neurological symptoms in one patient, rapamycin remains a candidate drug for treatment neurological symptoms in patients carrying an MTOR mutation.

2378T

A novel *PLEC1* nonsense homozygous mutation (c.7159G>T; p.Glu2387*) causes epidermolysis bullosa simplex with muscular dystrophy and diffuse alopecia: A case report. *L. Mota-Vieira[1], Z. Argyropoulou[1], L. Liu[2], L. Ozoemena[2], J.A. McGrath[2], A. Reis-Rego[3], R. Senra[3].* 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo of Ponta Delgada, Azores Islands, Portugal; 2) The National Diagnostic EB Lab, St Thomas' Hospital, London, UK; 3) Internal Medicine Department, Hospital of Divino Espírito Santo of Ponta Delgada, Azores Islands, Portugal.

   Epidermolysis bullosa simplex with muscular dystrophy (EBS-MD; OMIM #226670) is a autosomal recessive disease, characterized mainly by skin blistering at birth or shortly thereafter, the primary clinical manifestation, by progressive muscle weakness with highly variable age of onset (from infancy to the fourth decade life), and rarely by alopecia. EBS-MD is caused by mutations in the *PLEC1* gene (OMIM *601282), which encodes the cytolinker protein plectin expressed in many tissues (*i.e*., skin, mucous membranes, gut, muscle and heart). Here, we described a patient affected with EBS-MD and diffuse alopecia in which we identify a novel pathogenic variant in the *PLEC1* gene. The patient was a 28 year old girl and the only child of consanguineous (first degree cousins) healthy parents from the Azorean island of São Miguel (Portugal), born after an uneventful pregnancy. From the second day after birth, she developed blistering of the skin and oral mucosa, accompanied by hoarseness of the voice. In addition, at early age, she suffered from recurrent upper tract respiratory infections, and showed extensive nail dystrophy and hair weakness. All developmental milestones were normal, with exception of weight (below the 5th percentile). Muscle weakness was first noted in adolescence and gradually progressed, resulting in the inability to perform activities of daily living. On physical examination, we observed sparse, tense blisters and hemorrhagic crusts distributed over the patient's body. Diffuse non-scarring alopecia was evident on the scalp, and onychodystrophy (pachyonychia) were found to affect all 20 nails; she also presented decayed teeth, mild dysphonia, and severe muscle atrophy in the extremities. On the neurological examination, reflexes were profoundly diminished. To identify the pathogenic mutation, we performed PCR amplification of all coding exons and exon–intron boundaries of the *PLEC1* gene (isoforms 1a, 1b, and 1c), followed by bidirectional Sanger sequencing. This strategy revealed the patient to be homozygous for the novel nonsense mutation c.7159G>T (p.Glu2387*) in exon 31 (according to the isoform 1c: NM_00445/NP_00336). Presumably, this mutation predicts the expression of shortened plectin polypeptides (*i.e.,* the rodless truncated forms of plectin without the full-length isoform). To our knowledge, the present case corresponds to the second reported mutation with EBS-MD and diffuse alopecia in the literature. (luisa.mq.vieira@azores.gov.pt).

2379F

Reverse phenotyping of a patient with *CRIPT* gene mutation and further delineation of the associated phenotype. *B. Demeer[1,2], A. Dadban[3], P. Vabres[4,5], A. Varenterghem[7], J. Thevenon[4,8], D. Bremond-Gignac[9], J.B Riviere[4,6], C. Thauvin[4,8], L. Faivre[4,8].* 1) genetics department, Amiens University hospital, Amiens, France; 2) EA 4666, Université de Picardie Jules Verne, Amiens, France; 3) dermatology department, CHU Amiens-Picardie, Amiens, France; 4) Equipe d'Accueil 4271 Génétique des Anomalies du Développement, Fédération Hospitalo-Universitaire, Université de Bourgogne, Dijon, France; 5) Service de Dermatologie, CHU Le Bocage, Dijon, France; 6) Laboratoire de Génétique moléculaire, FHU-TRANSLAD, Plateau technique de Biologie, CHU, Dijon, France; 7) Pediatric gastroenterology department, CHU Amiens-Picardie, Amiens, France; 8) Centre de Référence Maladies Rares "Anomalies du Développement et Syndromes Malformatifs" de l'Est, Centre de Génétique et Pédiatrie 1, Hôpital d'Enfants, Dijon, France; 9) Ophtalmology department, CHU Amiens-Picardie, Amiens, France.

We report on a 3 ½ year old boy, with prenatal onset growth deficiency (height : -4SD), microcephaly (OFC :-3.5 SD) , transient neonatal pancyto-penia, facial dysmorphism, feeding difficulties, developmental and speech delay, global hyperlaxity, significant sleep disturbance, and genital, ocular and extremities anomalies. He also presented with generalized pigmentation anomalies, and signs of ectodermal dysplasia. When last seen for follow-up at 5 years old, he had developed epilepsy. Array CGH (Agilent 60k), cytogenetic diagnosis of chromosomal breakage syndrome and metabolic screening were negative. Whole exome sequencing performed found a homozygous frame-shift mutation of the *CRIPT* gene, recently described as a novel primordial dwarfism gene (Shaheen et al, 2014). The mutation (c.132delA), described as "probably pathogenic", was confirmed by sanger sequencing. Both healthy consanguineous parents were proved to be carrier in the heterozygous state. Few available clinical data of the 2 described patients shows very similar clinical appearance with strikingly facial dysmorphism, growth deficiency, microcephaly, psychomotor delay, and ocular and extremities anomalies. Mottled hypopigmentation is noted in the older patient described. This description is an example of "reverse phenotyping". The first description of the *CRIPT* gene by Shahen et al, helped us to reach a diagnosis in our patient. Nevertheless considering this new case the term of "primordial dwarfism" seem to be too restrictive. For instance, cutaneous signs are specific in our patient. Reports of additional patients will help to further delineate the associated phenotype of this unique syndrome.

2380W

Pain related polymorphisms in COMT and SCN9A are significantly higher in patients with Ehlers-Danlos Syndrome. *P. Coucke, S. Symoens, A. De Paepe, D. Syx, F. Malfait.* Medical Genetics, University Ghent, Ghent, Oost-Vlaanderen, Belgium.

Acute and chronic pain are common, though variable, manifestations observed in individuals suffering from Ehlers-Danlos syndrome (EDS). This clinically and genetically heterogeneous group of heritable connective tissue disorders is characterized by joint hypermobility, skin fragility and generalized soft connective tissue fragility. Although pain is a frequent cause for seeking medical help, it is often inadequately controlled and represents an unmet medical need. Currently, the pathogenic basis of EDS-related pain is poorly understood and virtually unexplored. Intriguingly, pain is frequently observed in the classical and hypermobility type of EDS, but less so in the vascular type. As genetic polymorphisms related to pain sensitivity in genes such as *COMT* and *SCN9A* have been described, we investigated the occurrence of these variants for the 3 EDS subtypes.   A group of 59 and 46 patients with molecu-larly proven classical (*COL5A1/COL5A2)* or vascular (*COL3A1)* EDS, respec-tively, and 64 clinically diagnosed familial cases with hypermobile EDS, were analysed. A random group of 123 individuals from the same ethnic population was used as control. We investigated two polymorphisms in *COMT* (rs4680; rs4818) and two polymorphisms in *SCN9A* (rs6746030; rs6754031).   In the cohort of classical EDS patients the presence of the A-allele (rs4680) in *COMT* was significantly higher ($p<0.05$) compared to controls whereas no differences were observed for all 3 other *COMT* or *SCN9A* polymorphisms. In the cohort of hypermobile EDS patients, we observed a significant difference ($p<0.05$) of the G-allele (rs6754031) in *SCN9A* whereas no significant differences were observed for the 3 other polymorphisms. The A-allele in *COMT* and the G-allele in *SCN9A* are reported to be associated with a higher pain sensitivity. No statistical differences were observed for any of these polymorphisms in the vascular EDS cohort.   Although further investigations are necessary, these observations suggest a role for polymorphisms in genes related to pain sensi-tivity in the pain pathology of classical and hypermobile EDS.

2381T

Growth characteristics in osteogenesis imperfecta – Results from an observational study from the Linked Clinical Research Centers. *A. Tam[1], M. Jain[1], J.R. Shapiro[2], R.D. Steiner[3], P.A. Smith[4], M.B. Bober[5], T. Hart[6], D. Cuthbertson[7], J. Krischer[7], E.L. Germain-Lee[2], M. Mullins[1], S. Bellur[1], G. Harris[4,8], P.H. Byers[9], M. Pepin[9], D. Eyre[10], D. Krakow[11,12,13], L. Tosi[14], C. Raggio[15], E. Orwoll[16], E. Rush[17], M. Durigova[18], F.H. Glorieux[18], F. Rauch[18], B.H. Lee[1,19], V.R. Sutton[1,19], S.C. Nagamani[1,19].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Bone and Osteogenesis Imperfecta, Kennedy Krieger Institute, Baltimore, MD; 3) Departments of Pediatrics and Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA; 4) Department of Orthopaedic Surgery, Shriners Hospitals for Children, Chicago, IL, USA; 5) Division of Medical Genetics, Alfred I. duPont Hospital for Children, Wilmington, DE; 6) Osteogenesis Imperfecta Foundation, Gaithersburg, MD, USA; 7) College of Medicine, University of South Florida, Tampa, FL; 8) Orthopaedic and Rehabilitation Engineering Center, Marquette University and Medical College of Wisconsin, Milwaukee, WI; 9) Departments of Medicine and Pathology, Division of Medical Genetics, University of Washington, Seattle, WA; 10) Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA; 11) Department of Orthopedic Surgery, University of California, Los Angeles, California, USA; 12) Department of Human Genetics, University of California, Los Angeles, California, USA; 13) Department of Obstetrics and Gynecology, University of California, Los Angeles, California, USA; 14) Children's National Health System, George Washington University School of Medicine and Health Sciences, Washington, DC; 15) Hospital for Special Surgery, New York, NY; 16) Bone and Mineral Unit, Oregon Health & Science University, Portland, OR; 17) University of Nebraska Medical Center, Munroe-Meyer Institute for Genetics and Rehabilitation, Omaha, NE; 18) Department of Orthopedic Surgery, Shriners Hospital for Children and McGill University, Montreal, QC; 19) Texas Children's Hospital, Houston, TX.

Osteogenesis imperfecta (OI) predisposes to recurrent fractures and bone deformities, which are factors that affect growth. There are no large-scale studies that have systematically investigated growth in patients with OI. The Linked Clinical Research Centers (LCRC) is a network of five clinical centers across North America established to advance clinical research and care for OI. From the LCRC data, we compared height, weight, BMI and arm-span to height ratio in 553 individuals with OI (types I n=244; III n=110; IV n=150; V n=16; VI n=12; VII n=5; unclassified n=16). The classification of OI was predominantly based on clinical features; however, genotypic information was used to reclassify patients whenever available. The large number of individuals with OI types I, III and IV enrolled in the study allowed for detailed statistical comparisons. In the pediatric population (age < 20 years), the median (IQT) of Z-scores for height in OI types I, III, and IV were -0.66(-1.43,-0.02), -6.91(-8.41,-5.17) and -2.79 (-3.95,-1.69) respectively. Furthermore, in OI type I adults without an affected parent, 48% were below 2 standard deviations from the expected mid-parental height. Although the median Z-score for weight in children with OI type III was -4.55, the odds of them being classified as overweight (>85th percentile for BMI) and obese (>95th percentile) were 2.48 and 5.23, respectively, as compared to OI type I. The median (IQT) of Z-scores for BMI in the pediatric population in OI types I, III and IV were 0.10(-0.58-0.94), 0.91 (0.42-1.61) and 0.67 (-0.18-1.33). Hence, the standard method for calculation of BMI in the general population may not be accurate in OI. The arm-span/height was significantly different in adults with OI types III and IV, as compared to OI type I, and thus demonstrates axial and lower limb growth restriction/deformation compared to the upper extremities. Generalized linear model analyses demonstrated that reduced height Z-score positively correlates with severity of OI subtype (p<0.001), age, bisphosphonate use and rodding (p<0.05). From this cross-sectional analysis of the largest cohort of individuals with OI, we provide median values for height, weight and BMI Z-scores which can aid in evaluation of overall growth in the clinical setting. With further collection of longitudinal data through the Brittle Bone Disorders Consortium, we plan to develop standardized longitudinal age and sex-based growth curves specific to OI.

2382F

Insights in the clinical features and molecular pathogenesis of Spinocerebellar Ataxia 38 (SCA38). *E. Di Gregorio[1,2], B. Borroni[3], M. Ferrero[2], L. Boccone[4], C. Costanzi[3], L. Orsi[5], S. Cavalieri[2], E. Giorgio[2], C. Mancini[2], E. Pozzi[2], N. Mitro[6], D. Caruso[6], A. Brusco[1,2].* 1) Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Neurology Unit, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy; 4) Ospedale Regionale Microcitemie, ASL 8, Cagliari, Italy; 5) Neurologic Division 1, Department of Neuroscience and Mental Health, Città della Salute e della Scienza University Hospital, Turin, Italy; 6) Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy.

SCA38 is a rare form of autosomal dominant ataxia caused by mutations in the elongase gene *ELOVL5*. At present, only two missense mutations have been reported in the gene, c.689G>T (p.Gly230Val; three Italian families) and c.214C>G (p.Leu72Val; one French family). In twenty-one SCA38 patients, from three families, with the *ELOVL5* c.689G>T mutation, we characterized clinical features and disease course. SCA38 showed an onset in the fourth-fifth decade of life with slow progression, and complete penetrance by age 50 yrs. The clinical picture corresponded to a pure cerebellar ataxia and included gait ataxia, nystagmus and dysarthria as the foremost manifestations. Among features helping the diagnosis, *pes cavus* without paraparesis (82% of cases) and early hyposmia/anosmia (73% of cases) were annotated. Disease progression was characterized by cerebellar symptoms (limb ataxia, dysarthria, dysphagia, and ophtalmoparesis) followed, in the later stages, by ophtalmoplegia. Pathogenetic mechanisms underlying the disease are still poorly defined. We previously demonstrated a subcellular mislocalization of p.Gly230Val ELOVL5 protein in different cellular models. The wild type protein localized as expected in the endoplasmic reticulum (ER) compartment. In contrast, aberrant ELOVL5 showed a perinuclear and polarized subcellular localization. Based on these data, we hypothesized p.Gly230Val-ELOVL5 is a misfolded protein able to activate the cellular unfolded protein response (UPR). We showed ELOVL5 protein expression was significantly increased in SCA38 fibroblasts after a treatment with the proteasome inhibitor MG-132, supporting ELOVL5 folding was affected by p.Gly230Val. By western blot analysis of the UPR markers CHOP, ATF-4 and XBP1, all significantly increased in COS7 cells stably expressing p.Gly230Val ELOVL5, we demonstrated activation of ER-stress response. Based on these data, we reasoned that the use of the chemical chaperone PBA might act on misfolded ELOVL5 reverting to physiological ER distribution of the protein and to cellular rescue from ER-stress. Our preliminary data on COS7 cells, stably expressing p.Gly230Val ELOVL5, demonstrated that PBA treatment relocalized the protein into ER. In conclusion, our results support a role for altered ER-stress response in SCA38 pathogenesis, suggesting chemical chaperones might be useful in the treatment, as already proposed for other neurodegenerative diseases. .

2383W

Clinical characteristics and detailed haplotype analysis of patients with SCA36 in Japan. *K. Koh[1], H. Ishiura[2], Y. Ichikawa[3], T. Matsukawa[2], J. Goto[4], J. Mitsui[2], Y. Takahashi[5], M. Kawabe[2], K. Doi[6], J. Yoshimura[6], S. Morishita[6], M. Namekawa[7], T. Ogawa[8], Y. Sunada[9], T. Itoh[9], T. Inoue[9], H. Kurizaki[10], K. Hasegawa[11], S. Tsuji[2], Y. Takiyama[1].* 1) Neurology, University of Yamanashi, Chuo, Yamanashi, Japan; 2) Department of Neurology, Graduate School of Medicine, The University of Tokyo; 3) Department of Neurology, Kyorin University; 4) Department of Neurology, International University of Health and Welfare, Mita Hospital; 5) Department of Neurology, National Center Hospital, National center of Neurology and Psychiatry; 6) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo; 7) Department of Neurology, Jichi Medical University, Saitama Medical Center; 8) Department of Neurology, International University of Health and Welfare Hospital; 9) Department of Neurology, Kawasaki Medical School; 10) Department of Neurology, National Hospital Organization, Tokyo National Hospital; 11) Department of Neurology, National Hospital Organization, Sagamihara National Hospital.

   Purpose Spinocerebellar ataxia type 36 (SCA36) [MIM 614153] was originally identified in Japan. Thereafter, SCA36 has been also reported from, Spain, France, Portugal, and China. SCA36 is caused by a large hexanucleotide GGGCCTG repeat expansion in the first intron of *NOP56* [MIM 614154] gene. We screened SCA36 in our spinocerebellar ataxia patients and tried to identify disease-relevant haplotype. Methods Eighty six probands of SCA families with unknown genetic cause after mutation analyses of SCA1, SCA2, MJD/SCA3, SCA6, SCA7, SCA8, SCA12, SCA17, and DRPLA were enrolled for screening of SCA36. A large expansion ;in the first intron of *NOP56* had been screened by repeat-primed PCR. To determine the disease causing haplotypes, we genotyped all the patients and family members available using Genome-wide SNP array 6.0. Haplotypes were reconstructed manually by minimizing the number of recombination events. When the genotypes of the other family members were unavailable, we determined haplotypes using homozygosity mapping method. We also performed whole genome sequence analysis for detecting rare variants near the mutation. Result We found 17 SCA36 patients from nine families. Our patients developed symptoms in their 40s to 80s. All 17 patients revealed truncal ataxia, and 15 patients revealed limb ataxia. Muscle atrophy, tongue atrophy, and tongue fasciculation were observed in only one patient. Brisk tendon reflexes were revealed in all the patients, though only one patient showed positive Babinski sign. We found a founder haplotype around *NOP56* spanning more than 2.1 Mb in 6 families originated from the western part of Japan. On the other hand, three families originated from the eastern part of Japan shared another haplotype around *NOP56* spanning 2.5 Mb. Using rare variants detected by whole genome sequence analysis, the common region of the two disease-relevant haplotypes were less than 22 kb, raising a possibility that the repeat expansion originated independently. Conclusion In this study, we found 17 SCA36 patients from nine families. Detailed haplotype analysis revealed two haplotypes. Minimum shared haplotype around *NOP56* was less than 22 kb, raising the possibility that the repeat expansion originated independently, although the notion that all the families were originated from a very far ancestor could not be completely excluded.

2384T

The c.1529C>T (p.Ala510Val) *SPG7* missense mutation in sporadic cerebellar ataxias in Italy. *C. Mancini[1], E. Giorgio[1], S. Cavalieri[1], E. Pozzi[1], M. Ferrero[1], E. Di Gregorio[2], M. Ferrone[1], S. Bagnoli[3], S. Piacentini[3], L. Orsi[4], M. Zibetti[4], L. Pradotto[5], A. Brusco[1,2].* 1) Dep. Medical Sciences, University of Torino, Torino, Italy; 2) Medical Genetics Unit, "Città della Salute e della Scienza" Hospital, Torino, Italy; 3) Dep. of Neuroscience, Psychology, Drug Research and Child's Health, University of Florence, Florence, Italy; 4) Dep. of Neuroscience and Mental Health, Città della Salute e della Scienza University Hospital, Torino; 5) Division of Neurology and Neurorehabilitation, San Giuseppe Hospital, IRCCS-Istituto Auxologico Italiano, Piancavallo (VB), Italy.

   Ataxias are a clinically, genetically, and etiologically heterogeneous group of neurodegenerative disorders, characterized by a cerebellar syndrome associated with various other manifestations. More than 150 genes have been reported so far, with overlapping pathogenic mechanisms (e.g., cellular aggregates, mitochondrial dysfunction, defects in autophagic functions). However, the causative mutation is unknown in >40% of patients. A recent report suggested that *SPG7* mutations, causing the most common form of autosomal recessive spastic paraplegia (MIM#607259), account for approx. 18.6% of cases with unexplained ataxia in UK. All were homozygotes or compound heterozygotes for the c.1529C>T (p.Ala510Val) missense mutation (Pfeffer et al., 2015). We set up a rapid assay for c.1529C>T/p.Ala510Val using restriction enzyme analysis after PCR amplification, and screened 600 ataxic patients of Italian origin. We identified three homozygotes and 14 heterozygotes. We further extended mutation analysis by sequencing the entire *SPG7* coding region in 10 available heterozygotes. We found three cases with a second mutation: two were reported mutation (c.1617delC; c.2102A>C) and one was novel (c.1553-2delAG splicing mutation). Thus, *SPG7* mutants account for approx. 1% of sporadic cerebellar ataxias in Italy. In seven carriers of p.Ala510Val the pathogenicity of *SPG7* gene needs further studies to be confirmed. Clinical data collection is in progress. All patients were recruited for ataxia as initial presentation, and in most cases a pure cerebellar ataxia, without pyramidal signs was reported. The relatively high frequency of the *SPG7* allele c.1529C>T in the ataxic patients suggests it should be considered as priority test in the presence of pure ataxia even with absent or subtle pyramidal signs. Moreover, these findings further highlight the heterogeneity of disease spectrum linked to mutations in components of the mitochondrial *m*-AAA protease.

2385F

De novo missense variants in *CACNA1A* can cause congenital cerebellar ataxia with global developmental delay. *J.A. Rosenfeld[1], T. Harel[1], X. Luo[1], S. Yamamoto[1], M. Hall[2], K. Wierenga[2], M. Pastore[3], D. Bartholomew[3], M. Delgado[4], J. Rotenberg[5], R.A. Lewis[1], M. Almannai[1], L. Emrick[1], T. Lotze[1], M. Ummat[1], C.A. Bacino[1], M. Eldomery[1], Z. Coban Akdemir[1], F. Xia[1], H. Bellen[1], J. Lupski[1], Y. Yang[1], B. Lee[1], S.R. Lalani[1], M. Wangler[1], Members of the UDN.* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Nationwide Children's Hospital & The Ohio State University, Columbus, OH; 4) Texas Scottish Rite Hospital, Dallas, TX; 5) Houston Specialty Clinic, Houston, TX.

Variants in *CACNA1A* [MIM 601011], encoding the α-1A subunit of the neuronal P/Q type voltage-dependent Ca^{2+} channel, are known causes of ataxia: triplet repeat expansions extending a polyglutamine tract in the protein cause adult-onset spinocerebellar ataxia, type 6 [MIM 183086], and heterozygous variants (typically loss-of-function) cause childhood-onset episodic ataxia, type 2 [MIM 108500]. Additionally, missense variants in the gene cause autosomal dominant familial hemiplegic migraines [MIM 141500]. We report five individuals with congenital, non-fluctuating ataxia, hypotonia, ophthalmologic abnormalities, and global developmental delay. Four individuals had a recurrent, *de novo* c.4991G>A/p.R1664Q variant, which has been described previously in another individual with early-onset, persistent limb and trunk ataxia. The fifth individual has a *de novo* missense variant, c.5018G>C/p. R1673P. Both missense variants alter arginine residues within the fourth transmembrane domain, changing the pattern of positive charges within the voltage sensor. This pattern is also altered by other severe pathogenic alleles, suggesting that more severe clinical phenotypes could result from disruption of this specific domain. A sixth individual with a severe presentation of neonatal stroke and subsequent refractory epilepsy had a *de novo* variant within this transmembrane domain, c.5075T>A/p.L1692Q. To explore the functional consequences of these variants, we generated a loss-of-function (LOF) allele in the homologous gene (*cac*) in *Drosophila* that will allow expression of these human variant forms. This LOF allele recapitulates the lethal phenotypes previously observed in *cac* mutant flies. Human variants will be assessed through their ability to rescue the loss of *cac* in *Drosophila* photoreceptors, where conditional knockout causes neurodegeneration, synaptic transmission deficits and neuronal accumulation of autophagic vesicles. This cohort, combined with previous reports, shows that variants in *CACNA1A* can cause congenital non-fluctuating cerebellar ataxia and other severe neonatal presentations, thus expanding the spectrum of ataxia and other features associated with this important neuronal calcium channel gene. Functional studies in model organisms will provide further insight into this disease spectrum.

2386W

Rapamycin and bafilomycin A1 regulated autophagy against reactive oxygen species -induced apoptosis in transformed lymphoblastoid cells from patients with Machado-Joseph disease. *H. Tsai[1], L. Chen[2], C. Wu[1].* 1) Sch Med Tech, Chung Shan Med Univ, Taichung, Taiwan; 2) Department of Nursing, National Taichung University of Science and Technology.

Machado–Joseph disease (MJD) or spinocerebellar ataxia type 3 (SCA3), it is the commonest autosomal dominant spinocerebellar ataxias worldwide. Alleles of the *ATXN3* gene on chromosome 14q24.3-q31 that carry 13-36 CAG-trinucleotide repeats are present in normal individuals. Contrariwise, alleles with a CAG triplet 51-79 repeats are present in patients with MJD. Although the detail mechanism of pathogenesis is yet to be defined, neurotoxin, especially reactive oxygen species (ROS), release from aggregated mutant proteins, may play a role in the pathogenic process. We proposed that enhancement of autophagy lysosome pathway may be beneficial for some neurodegenerative disorders, such as MJD, in which the accumulation of misfolded/aggregated proteins and the dysfunction of mitochondria are the two major pathogenesis. Rapamycin is an inducer of autophagy, as inhibition of mTOR mimics cellular starvation by blocking signals required for cell growth and proliferation. Bafilomycin A1 is a known inhibitor of the late phase of autophagy. We investigated rapamycin and bafilomycin A1 regulated autophagy against ROS, a toxin created by oxidative stress and implicated in neurodegenerative disease, in mutant ataxin-3 cells. Results found the autophagy marker protein, Atg8 (LC3 class II) were increased in mutant ataxin-3 cells, which treated with rapamycin or bafilomycin A1. Moreover, rapamycin or bafilomycin A1 increased the levels of the anti-apoptotic protein Bcl-2, and decreased pro-apoptotic protein Bax. These results show that rapamycin and lowly concentration bafilomycin A1 may be a potential therapeutic candidate for neurodegenerative diseases involving glutamate cytotoxicity such as MJD.

2387T

De novo alterations in *KLF7* are a novel cause of intellectual disability, psychiatric and neuromuscular issues. *D.N. Shinde[1], Z. Powis[1], J.S. Cohen[2], D. Escolar[2], J. Burton[3], C.M.A. van Ravenswaaij-Arts[4], A.P.A. Stegmann[5,6], R. Chikarmane[7], A. Begtrup[7], S. Tang[1].* 1) Clinical Genomics, Ambry Genetics, Aliso Viejo, CA 92656, USA; 2) Kennedy Krieger Institute, Baltimore, MD 21205, USA; 3) University of Illinois College of Medicine at Peoria, Peoria, IL 61603, USA; 4) University Medical Center Groningen, The Netherlands; 5) Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; 6) Radboud University Medical Center, Nijmegen, The Netherlands; 7) GeneDx, Gaithersburg, MD 20877, USA.

The *KLF7* gene (MIM 604865) encodes the Krüppel-like factor 7 (KLF7), a transcriptional activator belonging to the KLF family with roles in several developmental processes including apoptosis, cell proliferation, differentiation and energy metabolism. Mouse and *in vitro* studies implicate *KLF7* in neurogenesis (axon guidance in hippocampus, olfactory bulbs and cortex), differentiation and quiescence, and demonstrate high expression in skeletal muscle. *KLF7* is also a candidate gene for the 2q33.3-q34 deletion phenotype in patients with Autism Spectrum Disorder (ASD) and Rett-like features, microcephaly, hypotonia, psychomotor retardation and mild dysmorphic features. Here we report three unrelated affected individuals with *de novo* missense alterations in *KLF7* detected by Diagnostic Exome Sequencing (DES). The first individual was 2 years old at the time of evaluation with hypotonia, gross motor delay, mild language and cognitive delay, swallowing issues, mildly elevated creatine kinase levels and later ASD. DES identified a *de novo KLF7* c.410C>T (p.T137M) alteration predicted by *in silico* methods to be a part of the F-box motif (137-TPPSSP-142), involved in protein-protein interactions and ubiquitin-mediated proteolysis, an important part of cell cycle progression. The second individual is 15 years old and reported to have intellectual disability (ID), tonus dysregulation, motor dyspraxia, coordination issues and is described as anxious and shy. She was previously reported to have neonatal feeding issues and hypertonia. DES revealed a *de novo KLF7* c.415C>T (p.P139S) alteration, also part of the F-box motif and located two amino acids away from the alteration identified in the first individual. The third individual is 16 years old and reported to have severe ID, ADD/anxiety and self-injuring behavior, feeding difficulties, decreased tone in upper extremities and increased tone in lower extremities. DES identified a *de novo KLF7* c.790G>A (p.D264N) alteration located in the zinc finger-containing DNA binding domain shown to bind to the minimal enhancer of the *NTRK1* gene (MIM 191315) encoding a receptor for nerve growth factor. These findings suggest a phenotype for individuals with *KLF7* alterations including developmental delay/ID, abnormal muscle tone, feeding/swallowing issues, psychiatric features and neuromuscular issues. Additional patients with *KLF7* alterations are needed to delineate the phenotypic spectrum and make genotype-phenotype correlations.

2388F

HARS-related Charcot-Marie-Tooth disease mimicking adult polyglucosan body disease. *B. Royer-Bertrand[1,2], L. Mittaz Crettol[1], J.A. Lobrinus[3], P. Tsouni[4], J. Ghika[5], C. Wider[5], A. Superti-Furga[1,6,7], C. Rivolta[2], T. Kuntzer[4], C. Tran[1,8].* 1) Center for Molecular Diseases, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 2) Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 3) Department of Pathology, University Hospital Geneva, Geneva, Switzerland; 4) Nerve-Muscle Unit, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 5) Department of Clinical Neurosciences, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 6) Medical Genetics Service, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland; 7) Department of Pediatrics, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland; 8) Service of Endocrinology, Diabetes and Metabolism, Lausanne University Hospital (CHUV), Lausanne, Switzerland.

Pathogenic monoallelic variants in the histidyl-tRNA synthetase gene (*HARS*) were recently associated with inherited Charcot-Marie-Tooth (CMT) disease, characterized by axonal or demyelinating sensorimotor neuropathy. We investigated a 51-y old man with adult-onset pes cavus, hammer toes, stuttering, dysarthria, and mild cognitive amnestic deficit with emotional blunting. Electrodiagnostic studies were consistent with a late onset demyelinating peripheral neuropathy; a brain MRI study showed slight cerebellar atrophy; and sural nerve biopsy showed presence of polyglucosan bodies suggesting a diagnosis of adult polyglucosan body disease. Whole-exome sequencing (WES) was done with appropriate informed consent, and in house bioinformatics pipelines were used to extract rare, non-synonymous, high quality variants that were not present among sizeable healthy control cohorts. While known genes associated with polyglucosan storage myopathy (*GYS1*, *GYG1*, *GBE1*, *RBCK1*, *PFKM*, *EPM2A*, *EPM2B*, *PRDM8* and *PRKAG2*) were negative, WES identified a novel heterozygote mutation in *HARS*: c.397G>T (p.Val133Phe). The amino acid substitution is in close proximity of other substitutions associated with CMT type 2. The variant was confirmed by Sanger sequencing in the patient, and shown to be absent in maternal DNA (paternal DNA unavailable). While the newly identified *HARS* gene variant is a likely explanation for the peripheral neuropathy, a cognitive deficit has not been reported so far in patients with HARS-related CMT mutations. It is possible that this phenotypic trait is only a chance occurrence. Yet, it seems likely that HARS mutations are more pleiotropic than hitherto assumed, and that their clinical expression may comprise polyglucosan body disease. .

2389W

Huntington Disease: Case report, clinical and genetic molecular diagnosis. *M.A. Aceves-Aceves[1], S.A. Alonso-Barragán[1,2,3], I.M. Salazar-Dávalos[1], A. Ochoa-Morales[4], L. Martínez-Ruano[4], M.G. González-Mercado[2], N.O. Dávalos[1,2], D. García-Cruz[1,2], I.P. Dávalos[1,2,3].* 1) CUCS, Instituto de Genética, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) DGH, CUCS, Universidad de Guadalajara, México; 3) CIBO, HE, IMSS, México; 4) Instituto Nacional de Neurología y Neurocirugía, México.

INTRODUCTION Huntington Disease (HD) MIM #143100, is an autosomal dominant, neurodegenerative disorder with a worldwide prevalence of 5 - 10 per 100,000, typically diagnosed between 30 to 50 years of age. HD is clinically characterized by progressive motor, cognitive and behavioral manifestations. HD is caused by trinucleotide repeat expansion (CAG) in the huntingtin gene (HTT) on chromosome 4p16.3. OBJECTIVE To present a female patient with clinical diagnosis of Huntington Disease, confirmed by genetic molecular analysis. CASE REPORT A 54 years old female, with a history of involuntary movements of head, neck and hands; and onset at 48 years old. At present emotional instability is evident; abnormal movements have been progressed and she presented speech disorder. Family history: the mother and four brothers are similarly affected. The patient has three sons, apparently healthy, without symptoms. At physical examination she presented dysarthria, abnormal movements, normal tendon reflexes in limbs, and no muscular weakness neither dementia. Brain CT scans showed cortical atrophy. Clinical features, family history and CT findings were suggestive of HD. Molecular analysis was performed (informed consent). The genetic molecular test detected an allele with a trinucleotide repeat expansion CAG (n) in the huntingtin gene (HTT). Number of CAG repeats 15/48. CONCLUSIONS The clinical course of the HD is progressive and incapacitating, the survival since the onset is approximately of 20 years. The clinical features, family history and cortical atrophy revealed in brain CT scans lead us to consider the diagnosis of Huntington Disease in this patient, confirmed by genetic molecular analysis.

2390T

Cranial irradiation in childhood mimicking neurofibromatosis type II. *S. Ben-Shachar[1,3], T. Dubov[3], H. Toledano-Alhadef[1,3], R. Bernstein-Molho[2,3], S. Constantini[1,3], D.G. Evans[4].* 1) The Gilbert Israeli Neurofibromatosis Center, Tel-Aviv Medical Center, Tel-Aviv, Israel; 2) Department of Oncology, Tel-Aviv Medical Center, Tel-Aviv, Israel; 3) Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel; 4) 4Genomic Medicine, University of Manchester, Manchester Academic Health Science Center, Institute of Human Development, Center Manchester NHS Foundation Trust, Manchester Royal Infirmary, Manchester, UK.

Background Neurofibromatosis type II (NF2) is a genetic disease characterized by bilateral vestibular schwannomas (VS) and other nerve system tumors. However, such tumors may be associated with environmental, rather than a genetic, etiology. **Methods** Individuals fulfilling the clinical criteria of NF2 who had been treated by head ionized irradiation at a young age were compared for disease characteristics and molecular analysis with non-irradiated sporadic NF2 cases. **Results** Three of 33 sporadic adult cases fulfilling NF2 diagnostic criteria had a history of early age cranial irradiation exposure. None of the irradiated patients had bilateral VS compared with 73.3% of the non-irradiated individuals. One of the irradiated patients had no VS, while none of the non-irradiated NF2 cases had absence of VS. All of the irradiated individuals had brain meningiomas and thyroid tumors compared with 47% and 0%, respectively, of the non-irradiated individuals. Molecular analyses for *NF2* mutations in blood of the irradiated individuals failed to detect disease-causing mutations. **Conclusions**: Environmental factors may mimic NF2. Identifying such non-genetic cases fulfilling clinical criteria of the genetic disease may be crucial for the purposes of genetic counseling and patient management.

2391F

Ribosomal RNA-processing protein 7 homolog A (RRP7A) is associated with primary microcephaly and localizes to the centrosome and primary cilia. *M. Farooq[1], L. Lindbæk[2], V.S. Nielsen[2], M. Mönnich[1], Y. Mang[1], L.B. Pedersen[2], K. Møllgaard[1], L. Hansen[1], K.W. Kjær[1], S.M. Baig[3], N. Tommerup[1], S.T. Christensen[2], L.A. Larsen[1].* 1) Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3 DK-2200 Copenhagen, Denmark; 2) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 3) Human Molecular Genetics Laboratory; Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE, PIEAS), Faisalabad, Pakistan.

Primary microcephaly (MCPH) is a rare autosomal recessive neurodevelopmental disorder in which the cerebral cortex is reduced in size leading to mild to severe mental retardation. MCPH is genetically heterogeneous, and most of the MCPH proteins are associated with the centrosome and play an important role in neuronal progenitor cell proliferation. In a search for novel centrosomal genes associated with MCPH, we performed linkage analysis and whole exome sequencing of a large consanguineous family and identified a homozygous missense mutation affecting a highly conserved amino acid residue in the ribosomal RNA-processing protein 7 homolog A (*RRP7A*). *RRP7A* encodes a 280 amino acids long protein which contains a RNA recognition motif. The missense mutation, predicted to be pathogenic by SIFT and Polyphen2 prediction tools, was present neither in the exome variant databases nor in 300 healthy controls of Pakistani origin. Both endogenous and GFP-tagged RRP7A localize to the nucleoli, centrosomes and spindle poles during all stages of mitosis as well as to primary cilia in various cell types, including patient-derived primary skin fibroblast. Primary patient fibroblasts show increased frequency of di-nucleated cells, mis-shaped nuclei and have decreased proliferation. RNA *in situ* hybridization in zebrafish embryos revealed that *rrp7a* is widely expressed in different brain compartments at 1-2 days post fertilization. Zebrafish *rrp7a* mutants have impaired craniofacial development, small heads and eyes, cell proliferation defects and increased apoptosis. Protein mass spec analysis and Co-IP assays show that RRP7A interacts with CEP164 and DNA damage-binding protein 1 (DDB1). Further studies are underway to characterize the molecular mechanisms of RRP7A function in the centrosome, primary cilia, neuronal development and MCPH.

2392W

De novo mutations of *RERE* cause a genetic syndrome with features that overlap those associated with proximal 1p36 deletions. *B.M. Fregeau[1], B.J. Kim[2], A. Hernández-García[2], V.K. Jordan[2], M.T Cho[4], R.E. Schnur[4], K.G. Monaghan[4], J. Juusola[4], J.A. Rosenfeld[2], E. Bhoj[5], E.H. Zackai[5], S. Sacharow[6], K. Barañano[7], D.G.M. Bosch[8,10,11], B.B.A. de Vries[8,11], K. Lindstrom[9], A. Schroeder[14], P. James[9], P. Kulch[9], S. Lalani[2], M.M. van Haelst[12], K.L.I. van Gassen[12], E. van Binsbergen[12], A.J. Barkovich[13], D.A. Scott[2,3], E.H. Sherr[1].* 1) Department of Neurology, University of California, San Francisco, San Francisco, CA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030, USA; 3) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX, 77030, USA; 4) GeneDx, Gaithersburg, MD, 20877, USA; 5) Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 6) Division of Medical Genetics, Boston Children's Hospital, Boston, MA, 02115, USA; 7) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, 21287, USA; 8) Department of Human Genetics, Radboud university medical center, 6525 GA Nijmegen, the Netherlands; 9) Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, AZ 85006, USA; 10) Bartiméus, Institute for the Visually Impaired, 3702 AD Zeist, the Netherlands; 11) Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud university medical center, 6525 GA Nijmegen, the Netherlands; 12) Department of Genetics, University Medical Center Utrecht, 3508 GA Utrecht, the Netherlands; 13) Department of Radiology, University of California, San Francisco, San Francisco, CA 94158, USA; 14) Division of Genetics, University of Rochester Medical Center, Rochester, NY 14642, USA.

Deletions of chromosome 1p36 affect approximately 1 in 5,000 newborns and are associated with developmental delay, intellectual disability, and defects involving the brain, eye, ear, heart, and kidney. Arginine-glutamic acid dipeptide repeats (*RERE*) is located in the proximal 1p36 critical region. RERE is a widely-expressed nuclear receptor coregulator that positively regulates retinoic acid signaling. Animal models suggest that RERE deficiency might contribute to many of the structural and developmental birth defects and medical problems seen in individuals with 1p36 deletion syndrome, although human evidence supporting this role has been lacking. In this report, we describe ten individuals with intellectual disability, developmental delay, and/or autism spectrum disorder who carry rare and putatively damaging changes in *RERE*. In all cases in which both parental DNA samples were available, these changes were found to be de novo. Associated features that were recurrently seen in these individuals included hypotonia, seizures, behavioral problems, structural CNS anomalies, ophthalmologic anomalies, congenital heart defects, and genitourinary abnormalities. The spectrum of defects documented in these individuals is similar to that of a cohort of 31 individuals with isolated 1p36 deletions that include *RERE* and are recapitulated in RERE-deficient zebrafish and mice. Taken together, our findings suggest that mutations in *RERE* cause a genetic syndrome and that haploinsufficiency of *RERE* might be sufficient to cause many of the phenotypes associated with proximal 1p36 deletions.

2393T

Absence of the autophagy adaptor SQSTM1/p62 causes childhood-onset neurodegeneration with ataxia. *T.B. Haack[1,2], E. Ignatius[3,4], J. Calvo-Garrido[5], A. Iuso[1,2], P. Isohanni[3,4], C. Maffezzini[6], T. Lönnqvist[3,4], A. Suomalainen[3], L. Kremer[1,2], E. Graf[2], R. Berutti[2], H. Stranneheim[5,8], G. Brandberg[9], A. Wedell[5,8], M. Kurian[10,11], S.A. Hayflick[12,13,14], P. Venco[15], V. Tiranti[15], M. Dichgans[16,17,18], R. Horvath[19,20], E. Holinski-Feder[19], T.M. Strom[1,2], C. Freyer[6,8], T. Meitinger[1,2,17], H. Prokisch[1,2], J. Senderek[21], A. Wredenberg[6,8], C.J. Carroll[3], T. Klopstock[17,18,21].* 1) Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany; 2) Institute of Human Genetics, Helmholtz Zentrum München, 85764, Neuherberg, Germany; 3) Research Programs Unit, Molecular Neurology, University of Helsinki, 00290 Helsinki, Finland; 4) Department of Child Neurology, Children's 18 Hospital, University of Helsinki and Helsinki University Hospital, 00029 HUS, Finland; 5) Department of Molecular Medicine and Surgery, Science for Life Laboratory, Karolinska Institutet, Stockholm 17176, Sweden;; 6) Department of Medical, Biochemistry and Biophysics, Karolinska Institutet, Stockholm 17177, Sweden; 7) Department of Clinical Neuroscience, Karolinska Institutet, Stockholm 17176, Sweden; 8) Centre for Inherited Metabolic Diseases, Karolinska University Hospital, Stockholm 17176, Sweden;; 9) Department of Pediatrics, Falu lasarett, 79182 Falun, Sweden; 10) Neurosciences Unit, Institute of Child Health, University College London, London WC1N 3BG, UK; 11) Department of Paediatric Neurology, Great Ormond Street Hospital, London WC1N 3BG, UK; 12) Department 27 of Pediatrics, Oregon Health & Science University, Portland, OR 97239, USA; 13) Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR 97239, USA; 14) Department of Neurology, Oregon Health & Science University, Portland, OR 97239, USA; 15) Unit of Molecular Neurogenetics – Pierfranco and Luisa MarianiCenter for the study of Mitochondrial Disorders in Children, IRCCS Foundation Neurological Institute "C. Besta," 20126 Milan, Italy; 16) Institute for Stroke and Dementia Research, Ludwig-Maximilians-University of Munich, 81377 Munich, Germany; 17) Munich Cluster for Systems Neurology (SyNergy), 80336, Munich, Germany; 18) DZNE – German Center for Neurodegenerative Diseases, 80336, Munich, Germany; 19) MGZ - Medical Genetics Center, 80335 Munich, Germany; 20) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK; 21) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilians-University, 80336 Munich, Germany.

SQSTM1 (sequestosome 1; also known as *p62*) encodes a multidomain scaffolding protein involved in various key cellular processes, including the removal of damaged mitochondria by its function as a selective autophagy receptor. Heterozygous variants in *SQSTM1* have been associated with Paget disease of the bone and might contribute to neurodegeneration in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Using exome sequencing, we identified three different bi-allelic loss-of-function variants in *SQSTM1* in nine affected individuals from four families with a childhood- or adolescence-onset neurodegenerative disorder characterized by gait abnormalities, ataxia, dysarthria, dystonia, vertical gaze palsy, and cognitive decline. We confirmed absence of the SQSTM1/p62 protein in affected individuals` fibroblasts and found evidence of a defect in the early response to mitochondrial depolarization and autophagosome formation. Our findings expand the *SQSTM1*-associated phenotypic spectrum, and lend further support to the concept of disturbed selective autophagy pathways in neurodegenerative diseases.

2394F

A suppressor screen in *Mecp2* mice reveals pathways for Rett syndrome pathogenesis. *M.J. Justice[1,2,3], S.M. Kyle[1,3], J. Ruston[1], R. Zeng[1].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Dept. of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas USA.

Mutations in methyl CpG binding protein 2 (*MeCP2*) cause Rett syndrome, an X-linked neurological disorder with autistic features and developmental regression. Mutation of *MeCP2* causes widespread gene dysregulation, making the syndrome difficult to treat. *Mecp2* mouse mutants provide an excellent animal model to identify molecules that are important for disease pathogenesis. Studies of the mouse model show that the symptoms can be reversed by re-introduction of *Mecp2*, providing evidence that therapeutic intervention is possible in humans. *MeCP2* mutation impacts many biological pathways, but it is unclear which are relevant to symptom onset and progression. A modifier screen is a forward-genetic approach to find mutations that suppress or enhance a phenotype of interest, allowing the organism to reveal important pathways for morbidity. Here, random unbiased mutagenesis with *N*-ethyl-*N*-nitrosourea (ENU) was employed to isolate mutations that suppress clinical signs and improve overall health in a *Mecp2* mouse model. Strikingly, the lesions identified by whole exome sequencing in 46 trait-suppressing lines reveal 32 candidate genes that predominantly lie in four pathways. A modifying mutation in Squalene epoxidase (*Sqle*) showed that cholesterol synthesis was abnormal in *Mecp2* mice, and revealed the importance of brain lipid homeostasis to neurological function (1). Subsequent data show that MECP2 is required to link a repressor complex that regulates lipid homeostasis to DNA, resulting in metabolic syndrome (2). Of the pathways that contribute to Rett syndrome morbidity, it is clear that metabolism plays a key role in disease, and metabolic modulators such as statin drugs improve health in mice. Lipid homeostasis thus represents a potential new therapeutic target for Rett syndrome treatment and statin clinical trials are underway. Altogether, our data suggest that multiple factors will be required to reverse disease entirely, requiring combination therapies. A similar genetic approach could be exploited to identify targetable pathways involved in other "untreatable" diseases, opening a new field for translational discovery. (1) Buckovecky CM, et al. (2013) A suppressor screen in *Mecp2* mutant mice implicates cholesterol metabolism in Rett syndrome. *Nature Genetics* 45(9): 1013-1020. (2) Kyle SM, Saha PK, Brown HM, Chan LC, and Justice MJ (2016) MeCP2 co-ordinates liver lipid metabolism with the NCoR1/HDAC3 corepressor complex. *Human Molecular Genetics*, in press.

2395W

Screen for *de novo* mutations in epileptic encephalopathy genes in a cohort of patients with brain malformations. *A.M. Muir[1], N.T. Nguyen[1], G.L. Carvill[1], H. Crabtree[1], S. Hermans-Beijnsberger[1], W.B. Dobyns[2], I.E. Scheffer[3,4,5], H.C. Mefford[1].* 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, Washington 98195, USA; 2) Pediatrics and Neurology, University of Washington, Seattle, WA, 98195, USA; 3) Epilepsy Research Centre, Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Victoria, 3084, Australia; 4) Florey Institute of Neuroscience and Mental Health, The University of Melbourne, VIC 3010, Australia; 5) Department of Paediatrics, Royal Children's Hospital, The University of Melbourne, Parkville, Victoria, 3050, Australia.

Malformations of cortical development (MCDs) are brain abnormalities affecting the cerebral cortex that arise from disruption in the proliferation, differentiation, migration, or post-migrational organization of neurons in the developing cortical plate. Many MCDs appear to have an underlying genetic etiology, and mutations in a number of genes involved in cortical development and neuron migration have been shown to cause MCDs. The disruption of normal neuronal circuitry from MCDs can lead to a number of downstream clinical consequences including epileptic seizures. Epileptic encephalopathies (EEs) are severe, early-onset epilepsy disorders associated with intractable seizures and poor developmental outcome. Typically, epilepsy associated with MCDs and EE, which typically occurs without MCD, are thought to have distinct etiologies. However, as more patients undergo testing by whole exome sequencing or large gene panels, it is clear that there is some shared genetic etiology. *De novo* mutations in genes involved in cortical development and associated with brain malformations such as *FLNA* and *ARX* have been found in patients with EE, and mutations in *DEPDC5* have been found to cause focal epilepsy with and without focal cortical dysplasia in the same family. Together, this suggests that overlapping molecular mechanisms exist between MCDs and epilepsy. While *DEPDC5, FLNA,* and *ARX,* illustrate specific examples of shared etiologies between MCDs and epilepsy, a more comprehensive approach is needed to determine the extent to which common mechanisms are contributing to these two disorders. To systematically investigate the role of EE genes in MCD, we are screening 53 known EE and focal epilepsy genes in a cohort of 425 probands with MCDs including polymicrogyria (n=319), subcortical band heterotopia (n=43), and cortical dysplasia (n=29). We are currently using Molecular Inversion Probe (MIP) technology to sequence the coding regions from the epilepsy gene panel using a targeted capture sequencing approach. SNVs and indels will be filtered to identify rare, protein-changing variants. Since the cohort is comprised primarily of sporadic cases of MCDs, segregation testing will also be preformed to identify *de novo* variants, as these are most likely to be disease causing. *De novo* mutations in this screen would not only identify additional MCD genes but also provide further evidence towards the common underpinnings of MCD and epilepsy.

2396T

***GPAA1* mutations cause a new GPI-anchor biosynthesis defect with developmental delay, epilepsy, cerebellar hypoplasia and osteopenia.** *T.T.M. Nguyen[1], Y. Murakami[3], E. Sheridan[4], J. Chai[8], L. Fairbrother[5], A. Bateman[7], T. Reimschisel[6], E. Berry-Kravis[5], F. Xia[9], J. Tardif[1], J.A. Rosenfeld[9], C. Johnson[4], J.G. Gleeson[8], T. Kinoshita[3], P.M. Campeau[1,2].* 1) CHU-Sainte Justine Research Centre, University of Montreal, Montreal, QC, Canada; 2) Department of Pediatrics, University of Montreal, Montreal, QC, Canada; 3) Immunology Frontier Research Center, and Department of Immunoregulation, Research Institute for Microbial Diseases, Osaka University, Japan; 4) Faculty of Medicine and Health University of Leeds, UK; 5) Rush University Medical Center, Rush University, Chicago, IL , USA; 6) Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, USA; 7) Vanderbilt University Medical Center, Monroe Carell Children's Hospital, Nashville, TN, USA; 8) Department of Neurosciences, University of California, San Diego, CA, USA; 9) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA.

Approximately one in every 200 mammalian proteins are anchored to cell membranes through a glycosylphosphatidylinositol (GPI) anchor, and these proteins play important roles in neurological development and function, as well as in most other systems. GPAA1 together with PIGK, PIGS, PIGT, and PIGU form the GPI anchor transamidase complex which is responsible for the attachment of GPI anchors to protein precursors. We found bi-allelic *GPAA1* mutations in seven patients from four unrelated families by exome sequencing. Features present in most individuals include global developmental delay, hypotonia, early-onset seizures, cerebellar hypoplasia, and osteopenia. Plasma alkaline phosphatase was normal. These features overlap with those of other known GPI biosynthesis defects. Mutations identified include frameshift (NM_003801.3:p.G307Afs*11, p.Q327Hfs*102), splicing (c.1164+5C>T) and missense mutations (p.S51L, p.W176S, p.L291P, p.A389P). Flow cytometry studies on five available patient samples showed that several GPI-anchored proteins had decreased cell surface expression in leukocytes (FLAER, CD16, CD55) and fibroblasts (FLAER, CD73, CD87, CD109). Lentiviral rescue studies are underway. These findings thus allow us to include GPAA1 deficiency in the expanding group of GPI-anchor biosynthesis defects, and highlight the role of the transamidase complex in the formation and function of the cerebellum and bone. TTMN, YM and ES contributed equally to the work.

2397F

THAP11 mutations in a patient with a *cblX*-like phenotype implicates the HCFC1/THAP11 complex in regulation of cobalamin metabolism and early vertebrate development. *A.M. Quintana[1], H. Yu[2], A. Brebner[3], M. Pupavac[3], E.A. Geiger[2], W. Cheung[3], S. Chen[3], D. Watkins[3], T. Pastinen[3], F. Skovby[4], B. Appel[5], D.S. Rosenblatt[3], T.H. Shaikh[2].* 1) Biological Sciences, University of Texas El Paso, El Paso, TX; 2) Pediatrics, Section of Genetics, University of Colorado School of Medicine, Aurora, CO; 3) Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada; 4) Clinical Genetics, Rigshospitalet, Copenhagen, Denmark; 5) Pediatrics, Section of Developmental Biology, University of Colorado School of Medicine, Aurora, CO.

Mutations in *HCFC1* are associated with *cblX* (MIM309541), an X-linked recessive disorder, characterized by defects in cobalamin metabolism, nervous system development, neurological impairment, failure to thrive and mild to moderate facial dysmorphism. Functional analyses have demonstrated that *HCFC1* mutations disrupt cobalamin metabolism via the down-regulation of *MMACHC*. HCFC1 is a transcriptional co-regulator which requires interaction with transcription factors in order to regulate the expression of a myriad of genes. Notably, HCFC1 appears to regulate the expression of *MMACHC* through its interaction with THAP11, a THAP domain-containing transcription factor. In addition, the HCFC1/THAP11 complex has been shown to bind to regulatory regions of numerous genes involved in diverse cellular functions including cell cycle, proliferation and transcription. Thus, it is likely that mutations in *THAP11* may also result in biochemical and other phenotypes similar to those observed in patients with *cblX*. We report a patient who presented with phenotypic features that overlap *cblX*, but did not have any mutations in either *MMACHC* or *HCFC1*. We sequenced *THAP11* by Sanger sequencing and discovered a potentially pathogenic, homozygous variant in *THAP11*, c.240C>G (p.Phe80Leu). Functional analysis in the developing zebrafish embryo, demonstrate that both Thap11 and Hcfc1 regulate the proliferation and differentiation of neural precursors, suggesting important roles in normal brain development. Further, the loss of *thap11* in zebrafish embryos, results in craniofacial abnormalities including the complete loss of Meckel's cartilage, the ceratohyal, and all of the ceratobranchial cartilages. These data are consistent with our previous work which demonstrated a role for HCFC1 in vertebrate craniofacial development. Furthermore, our high throughput RNA-sequencing analysis reveals several overlapping gene targets of HCFC1 and THAP11 that likely contribute to the biochemical as well as other phenotypes associated with *cblX*. Taken together, our results provide evidence that the HCFC1/THAP11 complex plays an important role in the regulation of cobalamin metabolism as well as other pathways involved in early vertebrate development.

2398W

Non-optic gliomas in adults and children with neurofibromatosis 1 (NF1). *L. Sellmer[1], M. Marangoni[2], M. Heran[2], P. Birch[1], S. Farschtschi[3], V.-F. Mautner[3], J. Friedman[1].* 1) Child and Family Research Institute, University of British Columbia, Vancouver, Canada; 2) Department of Radiology, University of British Columbia, Vancouver, Canada; 3) University Hospital Hamburg-Eppendorf, Hamburg, Germany.

Background: Non-optic gliomas are one of the most common tumours in patients with NF1. The natural history and frequency of these tumours in children and especially in adults are incompletely understood. Methods: 1722 head MRI scans were obtained from 562 unselected individuals with NF1 at the NF outpatient department of the University Hospital Hamburg-Eppendorf between 2003 and 2015. We analyzed the clinical features of non-optic gliomas and focused on the differences between adults and children. Results: The number of scans per patient ranged from 1 to 12; patients were followed for a median of 3.7 years. We identified 23 patients (4.1%) with non-optic gliomas, a prevalence about twice that reported in most previous studies. Median age of these patients at first scan was 22.4 years, much higher than previously reported in the literature. 13 tumours were located in the cerebellum, 8 in the brain stem, 4 in the corpus callosum, 3 in the temporal lobe, 1 in the frontal lobe, 1 in the thalamus and 1 in the internal capsule. Only 7 of the 23 glioma patients were symptomatic. Histology was available on 9 tumours and showed 8 pilocytic astrocytomas and one dysembryoplastic neuroepithelial tumour. 4 individuals developed a tumour after having at least one tumour-free scan, and 4 cases showed progression. The risk of new tumour development was 0.19% (95% confidence interval 0.06% to 0.52%) per patient year of follow-up for all patients older than 10 years of age. There were no newly appearing tumours in children under the age of 10, and the prevalence of non-optic gliomas in children under 10 years of age was only 1.1%. Non-optic tumours in this study only progressed in the first 4.9 years after diagnosis and only in patients younger than 22.3 years of age. The rate of progressing non-optic gliomas per patient years of follow-up in the first 5 years after tumour diagnosis was 4.9% (95% confidence interval 1.6% to 13%). Conclusion: Non-optic gliomas are twice as common in an unselected cohort of NF1 patients as previously reported. This is most likely due to increased frequency of diagnosis of asymptomatic tumours when routine head MRIs are performed and a higher prevalence in older individuals, who are poorly represented in previously reported series.

2399T

GeneMatcher facilitates the characterization of a novel autosomal-recessive early-onset neurodegenerative encephalopathy caused by biallelic mutation in *TBCD*, **encoding the tubulin folding cofactor D.** *I. Thiffault[1,2,3], E. Flex[4], M. Niceta[5], S. Cecchetti[6], M.G. Au[7], A. Capuano[5], E. Piermarini[5], A.A Ivanova[8], J.W. Francis[8], G. Chillemi[9], B. Chandramouli[10], G. Carpentieri[11], C.A Haaxma[12], A. Ciolfi[5,13], S. Pizzi[5], D. Ganka[14], K. Levine[14], A. Sferra[5], R.R Pfundt[15], J.B. Lepichon[16], E.G Farrow[1], C. Saunders[1,2,3], F. Baas[17], F. Piemonte[5], B. Dallapiccola[5], J.M. Graham Jr[7], B. Bertini[5], R.A Kahn[8], D.A Koolen[12], M. Tartaglia[5].*
1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO 64108, USA; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospitals, Kansas City, MO 64108, USA; 3) University of Missouri-Kansas City School of Medicine, Kansas City, MO 64108, USA; 4) Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome 00161, Italy; 5) Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, Rome 00146, Italy; 6) Department of Cell Biology and Neurosciences, Istituto Superiore di Sanità, Rome 00161, Italy; 7) Department of Pediatrics, Medical Genetics, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; 8) Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA; 9) CINECA, SCAI-SuperComputing Applications and Innovation Department, Rome 00185, Italy; 10) Scuola Normale Superiore, 56126 Pisa, Italy; 11) Dipartimento di Medicina Sperimentale, Sapienza Università di Roma, Rome 00161, Italy; 12) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen 1105, The Netherlands; 13) Centro di Ricerca per gli alimenti e la nutrizione, CREA, Rome 00178, Italy; 14) GeneDx, Gaithersburg, MD 20877, USA; 15) Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen 1105, The Netherlands; 16) Department of Pediatrics, Children's Mercy Hospitals, Kansas City, MO 64108, USA; 17) Department of Genome Analysis, Academic Medical Center, Amsterdam 1105, The Netherlands.

The clinical usefulness of performing NGS in the clinical setting varies for different disorders. The identification of the underlying disease etiology can be challenging, given the clinical heterogeneity, sporadic status and the increasing number of causative genes. Many still remain without a molecular diagnosis, despite extensive investigation, exposing the need to study rare genetic variants in genes not yet linked to human diseases. Alterations of microtubule dynamics can lead to neurodevelopmental defects. Five conserved tubulin cofactors/chaperones (including *TBCD*, *TBCE*, and *ARL2*) regulate α- and β-tubulin assembly into heterodimers and maintain the soluble tubulin pool in the cytoplasm, but their physical mechanisms are unknown. The Tubulin Folding Cofactor D (*TBCD*) modulates microtubule dynamics by capturing GTP-bound beta-tubulin (TUBB). It also acts as a GTPase-activating protein (GAP) for ADP-Ribosylation Factor-Like 2 (ARL2). Here we describe a novel autosomal-recessive disease in six individuals from 4 unrelated kindred; biallelic mutations in *TBCD*, encoding one of the five co-chaperones required for assembly and disassembly of the α-β-tubulin heterodimer, the structural unit of microtubules, underlie a neurodegenerative disease characterized by early-onset cortical atrophy with microcephaly, thin corpus callosum, developmental delay and intellectual disability, seizures, optic atrophy, and spastic quadriplegia. Biochemical analyses documented variably reduced levels of TBCD, indicating relative instability or insolubility of mutant proteins, and defective β-tubulin binding in a subset of the tested mutants. Reduced/defective TBCD function resulted in decreased soluble α/β-tubulin levels and accelerated microtubule polymerization in fibroblasts from affected subjects, demonstrating an overall shift towards a more rapidly growing and stable microtubule population. These cells also displayed an aberrant mitotic spindle with disorganized, tangle-shaped microtubules and markedly reduced aster formation, which however did not alter appreciably the rate of cell proliferation. Our findings establish that defective TBCD function underlies a recognizable neurodegenerative disorder and causes accelerated microtubule polymerization and enhanced microtubule stability, underscoring a tight link between aberrant microtubule dynamics and neurodegeneration.

2400F

Searching for mutations in focal cortical dysplasia using next generation sequencing. *F.R. Torres[1], V.S. de Almeida[1], S.H. Avansini[1], M.K.M. Alvim[3], M. Borges[1], B.S. Carvalho[2], H. Tedeschi[3], F. Cendes[3], I. Lopes-Cendes[1].*
1) Medical Genetics Dept, UNICAMP, Campinas, Brazil; 2) Statistics Dept, UNICAMP, Campinas, Brazil; 3) Neurology Dept, UNICAMP, Campinas, Brazil.

Focal Cortical Dysplasia (FCD) is a brain malformation often associated with the occurrence of refractory seizures and it has been considered a mild phenotype within the hemimegalencephaly-tuberous sclerosis (TS) spectrum. The goal of our work is to search for germline and mosaic mutations in patients with FCD. Deep-Whole Exome Sequencing (WES) was performed in genomic DNA extracted from brain tissue resected by surgery (BTRS) and blood leucocytes of four patients. We performed capture and enrichment with Nextera® Expanded Kit. Samples were sequenced following a 200bp paired-end protocol in a Hiseq2500 to achieve at least 200x of average coverage. Realignment around indels and SNP clusters were performed using the Genome Analysis Toolkit (GATK). Variants were filtered prioritizing frameshift, missense, nonsense and splicing site mutations that were localized in coding regions or exon-intron boundaries. In addition, mosaicism was evaluated using Somatic Sniper and Mutect. Variants were classified as mosaic mutations when less than 50% of reads do not align to the sequence reference and are present only in BTRS. WES sequencing revealed a c2026T>A variation in the *TSC1* gene present both in BTRS and lymphocytes of a patient with FCD. This variant results in a missense mutation (W676R, Q92754) in *TSC1*. Furthermore, mosaicism analysis performed in additional patients identified several somatic mutations in BTRS, including nine variants on genes belonging to the mTOR pathway. This is the first description of a pathogenic germline variant in *TSC1* leading to FCD and supports the concept that FCD may be indeed a *forme fruste* of TS. Moreover, since mosaic mutations in genes associated with the mTor pathway have been reported in patients with FCD, we believe that the somatic variants identified in our cohort of patients are strong candidates to be involved with FCD etiology.

2401W

Severe growth retardation, microcephaly, intellectual disability, and characteristic facial features due to homozygous mutation in the QARS gene: Expanding the phenotype. *C. Vinkler[1,2], E. Leshinsky-Silver[2,6,7], J. Ling[3], C. Jalas[5], T. Lerman-Sagie[2,4,6], D. Lev[1,2,6].* 1) Inst Med Gen, Wolfson Med Ctr, Holon, Israel; 2) Metabolic-Neurogenetic Clinic; Wolfson Medical Center, Holon, Israel; 3) Department of Microbiology and Molecular Genetics, The University of Texas, Health Science Center at Houston, Houston, TX 77030, USA; 4) Pediatric Neurology Unit; Wolfson Medical Center, Holon, Israel; 5) Bonei Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, New York, USA; 6) Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, IsraelI; 7) Molecular Genetics Laboratory, Wolfson Medical Center, Holon, Israel.

Aminoacyl-tRNA synthetases (the aaRS family) are enzymes responsible for the proper pairing of codons on mRNA with amino acids and essential for interpreting the genetic code. They also form a multisynthetase complex (MSC) which participates in various other cell functions. Glutaminyl tRNA synthase is a member of class I aaRS enzymes. It is responsible for synthesis of the cytosolic pool of glutaminyl-tRNA^{Gln} (Gln-tRNA^{Gln}). Human GlnRS is encoded by QARS gene and is highly expressed in the developing fetal human brain. Mutations in *QARS* gene have been previously reported in patients from three unrelated families presenting with progressive microcephaly, cerebral-cerebellar atrophy and intractable seizures. We describe two sisters who are homozygous for a yet undescribed mutation in QARS. These girls have been recently reported by us presenting a new recessive syndrome. They were born to healthy unrelated parents of Ashkenazi-Jewish origin. They have six healthy siblings. The girls have severe linear growth retardation, poor weight gain, microcephaly, characteristic facial features, cutaneous syndactyly of the toes, high myopia and severe intellectual disability. Homozygosity mapping together with exome sequencing revealed homozygous missense (V476I) mutation in the QARS gene. The mutation is located in the catalytic domain. The parents and three of the healthy siblings are heterozygote carriers. Indeed, the patient's fibroblasts demonstrated a disrupted QARS aminoacylation activity in vitro. The clinical presentation in our patients differs from the original QARS-associated syndrome in the severe postnatal growth failure, non-progressive microcephaly, dysmorphic features and the absence of seizures. Mutations in QARS causing reduction in hGlnRS activity, may manifest not only as neurological disorder of brain development but should be considered also in patients with severe linear growth, microcephaly and intellectual disability.

2402T

Variants spectrum in 31 Chinese newborns with idiopathic seizures revealed by next-generation sequencing. *L. Yang[1], Z. Li[2,3], B. Sun[1], M. Mei[1], W. Tian[2], H. Wang[3], W. Zhou[1,3,4].* 1) Children's Hospital of Fudan University, Shanghai, Shanghai, China; 2) Department of Biostatistics and Computational Biology, Life Science, Fudan University; 3) Key Laboratory of Birth Defects, Children's Hospital of Fudan University, Shanghai, China; 4) Key Laboratory of Neonatal Diseases, Ministry of Health, Children's Hospital of Fudan University.

Background: The etiology of the idiopathic neonatal seizures is important to identify as it may serve a significant effect on diagnosis, prognosis, treatment and outcome. **Objectives:** The purpose of this study is to get the whole spectrum of mutation spectrum in 31 Chinese unrelated patients with idiopathic neonatal seizures. **Methods:** We performed *KCNQ2* gene Sanger sequencing and disease-targeted multi-gene next-generation sequencing across 741 nervous system-related genes in 31 unrelated neonates with idiopathic seizures. *Results:* A total of 8 variants in *KCNQ2* gene, including 7 possible pathogenic mutations in 7 patients (7/31, 22.6%) were identified by Sanger sequencing. The patient with R213W mutation suffered from epileptic encephalopathy, which updates the genotype-phenotype correlation of *KCNQ2*. Expanded to 741-gene panel, more 8 possible disease-causing mutations (in 5 genes) in 8 probands (8/31, 25.8%) were detected. In addition, the results of this study extend mutation spectrum in neonatal seizures in several related genes such as *KCNQ2*, *DEPEC5* and *SCN2A*. **Conclusions:** For first-tier clinical screening, a focused panel including just about 6 important genes with about 50% contribution is more cost effective, and *KCNQ2* should be a priority during genetic screening. This study enhances the current knowledge of genotypic and phenotypic heterogeneity of idiopathic neonatal seizures, which will assist both clinical diagnoses and personalized treatments of Chinese newborns with idiopathic seizures.

2403F

Loss-of-function mutations in *DPP6* increases risk for frontotemporal dementia. *R. Cacace[1,2], T. Van den Bossche[1,2,3,4], J. Lippiat[5], S. Van Mosselvelde[1,2,3,4], S. Engelborghs[2,4], P. De Rijk[6], G. De Baets[7], J. Janssens[1,2,8], A. Sieben[1,2,9], C. Robberecht[1,2], L. Dillen[1,2], N. Geerts[1,2], K. Peeters[1,2], M. Mattheijssens[1,2], S. Maudsley[6,8], J. Schymkowitz[7], E. Salmon[10], P. Santens[9], P. Cras[2,3], J.J. Martin[2], R. Vandenberghe[11,12], P.P. De Deyn[2,4], M. Cruts[1,2], J. van der Zee[1,2], C.M. van Duijn[13], K. Sleegers[1,2], C. Van Broeckhoven[1,2] on behalf of the Belgium Neurology (BELNEU) Consortium.* 1) Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Belgium; 2) Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; 3) Department of Neurology, Antwerp University Hospital (UZA), Edegem, Belgium; 4) Department of Neurology and Memory Clinic, Hospital Network Antwerp (ZNA) Middelheim and Hoge Beuken, Antwerp, Belgium; 5) School of Biomedical Sciences, Faculty of Biological Sciences, University of Leeds, Leeds, UK; 6) Bioinformatics Unit, Department of Molecular Genetics, VIB, Belgium; 7) VIB Switch Laboratory, Department of Cellular and Molecular Medicine, Katholieke Universiteit Leuven, Leuven, Belgium and Vrije Universiteit Brussel, Brussels, Belgium; 8) Translational Neurobiology Group, Department of Molecular Genetics, VIB, Antwerp, Belgium; 9) Department of Neurology, University Hospital Ghent and University of Ghent, Ghent, Belgium; 10) Department of Neurology, Centre Hospitalier Universitaire de Liège, Liège, Belgium; 11) Department of Neurosciences, Faculty of Medicine, University of Leuven, Leuven, Belgium; 12) Laboratory of Cognitive Neurology, Department of Neurology, University Hospitals Leuven Gasthuisberg, Leuven, Belgium; 13) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Centre, Rotterdam, The Netherlands.

In a previous genome-wide linkage study on an autosomal dominant Alzheimer disease (AD) family, we identified a genetic locus of 5.44 Mb at chromosome 7q36, but the underlying pathogenic mutation remained elusive (Rademakers et al. 2005). To reinvestigate the unresolved linkage, we performed whole genome sequencing on 4 patients from the pedigree. Bioinformatics revealed within the 7q36 locus, the presence of low copy repeats (LCRs), known to be prone to genomic instability. In the 4 sequenced individuals, we detected a chromosomal inversion of *c.a.* 4 Mb on the disease haplotype. The breakpoints were located in the LCRs in intron 1 of the dipeptidyl peptidase 6 (*DPP6*) gene and in an intergenic region outside the linked locus. Due to i) the heterogeneous clinical presentation - without definite pathological confirmation - of the linkage family, with an AD patient presenting with frontal signs and ii) reduced DPP6 protein expression in the brain of AD patients (Hondius et al., 2016), we assessed the genetic contribution of rare *DPP6* mutations to the dementia spectrum. We performed massive parallel sequencing of *DPP6* in Dutch AD (n=82), Belgian AD (n=335) and frontotemporal dementia (FTD, n=435) patients cohorts as well as control individuals (n=755). This revealed a significant enrichment of rare variants in FTD patients, when known pathogenic mutation carriers were excluded (19/359, 5.3%) compared to control individuals (13/755, 1.72%; SKAT-O p-value=0.001, relative risk=3.1, 95% confidence interval (CI)=1.54–6.37). We also identified two premature stop codon mutations in patients supporting the hypothesis of loss-of-function (LoF) as the underlying molecular mechanism. DPP6 regulates synaptic excitability by modulating the expression and function of the potassium channel Kv4.2 in brain. Expression studies on brain tissue of patient carriers of missense mutations showed a decreased expression of DPP6 transcript (p-value=0.0096), confirmed on protein level. Kv4.2 protein expression was also decreased in these patients. Electrophysiology studies in *Xenopus laevis* oocytes showed a significant negative effect (p-value <0.05) on Kv4.2 gating properties when the channel was co-expressed with two premature stop codon mutations. Our data support a risk association of *DPP6* with frontotemporal dementia that may involve brain network activity alterations.

2404W

First report of a Mendelian disease caused by a mutation in the *HTR2C* gene. *C. Vural[1], R. Scott[1], L. Cantor[2], M. Banikazemi[1].* 1) Regional Medical Genetics Center, Boston Children's Health Physicians, Valhalla, NY; 2) CWPW Neurology, Boston Children's Health Physicians, Sleepy Hollow, NY.

HTR2C, a clinically novel gene on Xq23, encodes the 5-HT2CR receptor, a member of a family of G-protein-coupled-receptors that control physiological responses to serotonin. Animal studies have implicated mutations in this gene to be associated with seizures, appetite abnormalities, hyperinsulinemia, and other pre-diabetic symptoms. It has also been hypothesized that polymorphisms in *HTR2C* gene are associated with symptoms including psychiatric disorders, hallucinations, and increased BMI, but no Mendelian disease in humans has yet been shown to be caused by mutations in *HTR2C*. We present the first report of a 6 year old boy with seizure disorder, tremors, and psychiatric symptoms including hallucinations, paranoia psychosis, and disruptive behavior due to mutation of the *HTR2C* gene. We have treated this patient since initial presentation at 6 months of age when he presented with infantile spasm. His physical growth has remained appropriate, but his seizures and psychiatric symptoms have become more difficult to control with conventional anti-epileptic and anti-psychotic medications. He has demonstrated delay in his cognitive development and progressive decline in his psychosocial development expanding to self-injurious behavior, prolonged and frequent crying, agitation, restlessness, trouble sleeping, claustrophobia, visual hallucinations, acute episodes of psychotic symptoms, speech delay, and autistic behsviors. He has required several hospitalizations for worsening of his symptoms. He has shown improved behavior, mood and severity of tremors after recent addition of Seroquel to his treatment. Whole Exome Sequencing (WES) was completed through Ambry Genetics and revealed a deleterious mutation (HTR2C c.1219_1222del) on the proband's only copy of the HTR2C gene. The proband's mother, who is heterozygous for this deletion, remains asymptomatic. Despite the clinical ambiguity surrounding the *HTR2C* gene, the patient has sufficient overlapping clinical symptoms with the *HTR2C* gene family for us to declare that the proband's deletion in *HTR2C* is responsible for his presentation. This is the first description of a pathogenic mutation in the *HTR2C* gene in a patient with infantile spasms, seizures, psychiatric symptoms, tremors, autism, and pre-diabetes. Our findings suggest that WES is an important diagnostic tool for detecting the genetic etiology of psychiatric symptoms and seizure disorders as well as for the discovery of novel causative gene mutations.

2405T

Compound heterozygous *TRPV4* mutations causing severe global developmental delay and overlapping phenotype of neuropathy, myopathy and skeletal involvement. *M. Thibodeau[1], C.H. Peters[2], K. Townsend[1,3], Y. Shen[4], G. Hendson[5], S. Adam[1], K. Selby[6], P.M. Macleod[7], C. Gershome[2], P. Ruben[2], S. Jones[1,4], J.M Friedman[1], W. Gibson[1], G. Horvath[8], the FORGE Canada Consortium.* 1) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 2) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada; 3) Child and Family Research Institute, Vancouver, Canada; 4) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, Canada; 5) Department of Anatomic Pathology, University of British Columbia, Vancouver, Canada; 6) Department of Pediatrics, Division of Pediatric Neurology, University of British Columbia, Vancouver, Canada; 7) Victoria General Hospital, Division of Medical Genetics, Victoria, Canada; 8) Department of Pediatrics, Division of Biochemical Diseases, University of British Columbia, Vancouver, Canada.

   Background: Several genes of the transient receptor potential channel family are known to cause hereditary channelopathies when mutated. *TRPV4* gene encodes a polymodal calcium-permeable plasma membrane channel. Dominant pathogenic mutations in the *TRPV4* gene lead to a wide spectrum of human phenotypes.We investigated the molecular mechanism underlying the complex neuromuscular, skeletal and neurodevelopmental phenotypes of two affected siblings in relation to their compound heterozygous state for *TRPV4* missense variants p.E840K (c.2518G>A, NM_021625.4) and p.N833S (c.2498A>G, NM_021625.4). **The Family**: A brother and sister presented with severe intellectual disability, progressive peripheral neuropathy and myopathy, severe sensorineural hearing loss, severe kyphoscoliosis, joint contractures, and small hands and feet. The brother also had retinopathy. Both sibs manifested signs within the first year of life. No other relatives had any of these conditions. Muscle biopsies, nerve biopsies and nerve conduction studies showed peripheral axonal neuropathy and myopathy in both patients. **Sequencing**: Whole exome sequencing was performed in both children and their unaffected mother. Compound heterozygous candidate *TRPV4* missense variants p.E840K and p.N833S were identified. Both variants are rare and p.E840K is predicted to be damaging. These variants were identified in both affected children; unaffected relatives had either variant but not both. The boy was the only family member with compound heterozygous *ABCA4* missense variants p.G863A (c.2588G>C, NM_000350.2), previously described as pathogenic in the literature, and p.R152Q (c.455G>A, NM_000350.2), which probably explains his retinopathy. **Methods and Results**: Functional assays of the p.E840K and p.N833S variants transfected into CHOk1 cells revealed decreased gated current densities, despite normal protein immunostaining localization assay. **Conclusions**: We report here the first two individuals with compound heterozygous *TRPV4* variants and classical TRPV4-related neuromuscular phenotypes. In the light of previously well-described dominant inheritance for this group of disorders, our study provides supporting evidence for compound heterozygosity leading to a far more severe neuromuscular phenotype than expected with a single *TRPV4* pathogenic variant. Unexplained severe developmental delay was noted in both affected children and future studies might implicate *TRPV4* gene in cortical neurogenesis.

2406F

Relationship between various clinical features in a genotyped population investigated for tuberous sclerosis complex. *N. Ishihara[1], H. Sasaki[2], T. Kato[3], H. Inagaki[3], M. Tsutsumi[3], R. Shiroki[2], H. Kurahashi[3].* 1) Dept Pediatrics, Fujita Health University, Toyoake, Aichi, Japan; 2) Dept Urology, Fujita Health University, Toyoake, Aichi, Japan; 3) Div Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, Japan.

   Introduction: Tuberous Sclerosis Complex (TSC) is a multi-system disorder, typically involving severe neurological symptoms, such as epilepsy, cognitive deficit and autism, severe cutaneous manifestation, and various tumors, such as subependymal giant cell astrocytoma (SEGA), renal angiomyolipoma (rAML) and lung lymphangioleiomyomatosis (LAM). TSC is inherited in an autosomal dominant pattern and is caused by mutations in either the TSC1 or TSC2 genes, which enhance activation of the mammalian target of rapamycin (mTOR) signaling pathway. As recent studies have suggested that mTOR inhibitors such as rapamycin can reverse TSC-associated deficits, the advent of pharmacological therapies has made early diagnosis important in patients with TSC. Although there is a substantial overlap in the clinical phenotype produced by TSC1 and TSC2 mutations, accumulating evidence indicates that TSC2 mutations cause more severe manifestations than TSC1 mutations. In this study, we investigated patients with various TSC manifestations to reveal the difference with genotype. Patients and Methods: We retrospectively investigated in a cohort of TSC outpatients of Fujita Health University Hospital, clinical manifestations and genetic mutations. Thirteen patients who were diagnosed as TSC clinically underwent genetic testing, and causal gene mutations were found in 12 patients. Genetic testing was performed with target exome sequencing, and validated with Sanger sequencing. Result: Clinical manifestations of 12 patients were following; epilepsy in 9 patients, rAML in 9 patients, LAM in 3 patients, and SEGA in 2 patients. Eleven patients had mutations in TSC2 gene, and one had in TSC1 gene. Mutations were following; nonsense mutation in 2 patients, frameshift mutation in 7 patients, missense mutation, one amino acid deletion and exonal deletion in one patient, respectively. The patient with exonal deletion had concomitant PDK1 deletion which led to contiguous gene syndrome. Conclusion: In our cohort, many of them have epilepsy and rAML, and have TSC2 truncated mutations, indicating that the TSC2 truncated mutation is linked to much severe clinical manifestations. Early pharmacotherapy interventions based on genotype may be available for our TSC patients.

2407W

Wieacker-Wolff syndrome and *ZC4H2*: Case report and discussion of proposed neural mechanism. *J. Moeschler, J. Ozmore.* Pediatrics, Medical Genetics, Dartmouth-Hitchcock Med Ctr, Lebanon, NH.

Wieacker-Wolff syndrome (OMIM 314580) is an X-linked recessive neuro-developmental disorder characterized by fetal akinesia sequence. Affected boys present with severe distal arthrogryposis, delayed motor development, facial and bulbar weakness, characteristic dysmorphic facial features, and skeletal abnormalities, such as hip dislocation, scoliosis, and equinovarus. Wieacker et al (1985) described this apparently novel X-linked syndrome in 6 males in 4 sibships of 3 generations of a family. Hirata et al (2013) identified a family with affected individuals who presented with neonatal respiratory distress, arthrogryposis multiplex congenita, muscle weakness, and ptosis. Hirata et al also identified the *ZC4H2* gene associated with this syndrome in 2013. Miles and Carpenter (1991) reported on a single family of 3 brothers and a male cousin with intellectual disability in association with exotropia, microcephaly, distal muscle wasting, and 10 low digital arches. Until 2015, no gene was noted associated with this syndrome. Miles-Carpenter syndrome (OMIM 309605) was reported in one family. The case reported herein is of an infant boy who presented at birth with distal arthrogryopsis, weakness, hypotonia, and poor feeding. Later he demonstrated cortical vision impairment and severely delayed development. He also had sinus bradycardia requiring pacemaker placement at age 5 months; and, myoclonic seizures. Abnormal brain MRI was noted: thin corpus callosum, mega cisterna magnum, non-specific signal abnormality in globus pallidus, left lateral genticulate, hypothalamus, and central tegmental tracts; irregular formation of folia of cerebellum. Whole exome sequencing identified in this patient a pathogenic variant in *ZC4H2* gene: c.593G>A, p.Arg198Gly. This variant was maternally inherited and has been reported (family 2) by Hirata et al. Hirata et al identified pathogenic variants in *ZC4H2* in four families with Wieacker-Wolff syndrome. May et al (2015) identified four mutations in *ZC4H2*, three missense and one in-frame insertion, in the original family with Miles-Carpenter syndrome and three other families with a constellation of microcephaly, short stature, contractures, spasticity and ID. Wieacker Wolff, Miles Carpenter and XLID syndromes associated with *ZC4H2* will be discussed together with published information regarding a proposed mechanism for the clinical phenotype.

2408T

Sleep movement disorders in adults with 22q11.2 deletion: A new dopamine-related manifestation? A case report. *G.M. Repetto[1], R. Fritsch[2], R. Fadic[3], P. Chana[4], C. Ornstein[2], C. Juri[3], V. Kramer[5], R. Pruzzo[5], H. Amaral[5], A. Ocampo[6].* 1) Center for Genetics and Genomics, Clin Alemana- Univ Desarrollo; 2) Departamento de Psiquiatria y Salud Mental Hospital, Clinico Universidad de Chile; 3) Departamento de Neurologia, P Universidad Catolica de Chile; 4) Centro de Trastornos del Movimiento (CETRAM); 5) Nuclear Medicine and PET/CT Center PositronMed-FALP; 6) Facultad de Medicina, Universidad de Chile, Santiago, Chile.

Adults with 22q11 microdeletion (22q11DS) are at high risk of developing early-onset movement disorders such as Parkinson disease (PD) or parkinsonism, either primarily as part of the syndrome or as a side effect of antipsychotic medication. A role for dopaminergic dysregulation has been suggested. Certain parasomnias (sleep disorders that involve abnormal movements, behaviors and/or perceptions) are frequent in PD. Patients can have non-specific non-REM sleep disorders such as periodic limb movements (PLM). On the other hand, REM behavior disorder is considered a prodromal and specific feature of PD. We report a 45 year old female with 22qDS with psychosis, who developed tremor, bradykinesia, rigidity and insomnia while treated with haloperidol. Her abnormal movements improved after its withdrawal and addition of trihexyphenidyl. A cerebral PET CT scan with the dopamine transporter ligand 18F-PR04.MZ showed bilateral symmetric increased uptake of the radioligand in putamen and caudate nuclei (130% of control). A sleep study showed intense and frequent periodic limb movements during REM sleep, consistent with Periodic Limb Movement Disorder (PLMD) and associated with a high number of microawakening episodes. The condition manifests intermittently and the movements may cause poor sleep and subsequent daytime somnolence. Treatment involves either dopaminergic medication in an attempt to modify activity of the subcortical motor system or, more commonly, sedative medications to allow uninterrupted sleep. The etiology of PLMD is uncertain. Suprasegmental disinhibition of the descending inhibitory pathways may be a factor. Vetrugno et al (2007) proposed neuronal hyperexcitability with involvement of the central pattern generator for gait as the pathophysiology of PLMD. This leads to decreased dopamine transmission, potentially supporting the use of dopaminergic therapy to treat the condition. This combination of symptoms due to elevated (psychosis) and low (Parkinson) dopamine signaling constitute a relevant therapeutic challenge for patients with 22qDS. We propose that sleep movement disorders are another manifestation of altered dopamine signaling of 22q11DS phenotype in adults. It is relevant to explore their frequency and impact on daily life in a large patient cohort. Funded by Fondecyt-Chile grant #1130392.

2409F

Huntington disease reduced penetrance alleles occur at high frequency in the general population. *C. Kay[1], J.A. Collins[1], Z. Miedzybrodzka[2], G.E. Wright[1], S.J. Madore[3], E.S. Gordon[3], N. Gerry[3], E.R. Fisher[1], M. Davidson[2], R.A. Slama[1], M.R. Hayden[1].* 1) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 2) Medical Genetics Group, School of Medicine and Dentistry, University of Aberdeen, Aberdeen, UK; 3) Molecular Biology Group, Coriell Institute for Medical Research, Camden, New Jersey.

Huntington disease (HD) is a fatal neurodegenerative disorder caused by an expanded CAG trinucleotide repeat in the Huntingtin (HTT) gene. HD is diagnosed in 10.6-13.7 individuals per 100,000 in Western populations (1 in 7300), but the general population frequency and penetrance of the causative CAG repeat expansion is unknown.   To directly estimate the frequency of CAG repeat alleles associated with Huntington disease (HD), CAG repeat length was evaluated in 7315 individuals from three population-based cohorts in British Columbia, the United States, and Scotland. The allelic frequency of ≥36 CAG alleles was assessed out of a total of 14630 alleles, and the genotypic frequency assessed out of a total of 7315 individuals. 18 of 7315 individuals had ≥36 CAG, revealing that approximately 1 in 400 individuals in the general population have an expanded CAG repeat associated with HD (0.246%). Individuals with CAG 36-37 genotypes are the most common (36, 0.096%; 37, 0.082%; 38, 0.027%; 39, 0.000%; ≥40, 0.041%).   To directly estimate the penetrance of the most common HD-associated alleles in our general population sample, the frequency of CAG 36-38 repeat alleles was compared to the prevalence of HD patients with genetically confirmed CAG 36-38 in a multisource clinical ascertainment across British Columbia, Canada. The penetrance of 36-38 CAG repeat alleles for HD was directly estimated for individuals ≥65 years of age and compared against previous inferred clinical penetrance estimates for this age class. General population CAG 36-38 penetrance rates are lower than inferred clinical penetrance rates, suggesting that a large proportion of individuals with 36-38 CAG repeat alleles are not clinically diagnosed with HD over a typical lifespan. The relatively infrequent diagnosis of HD at 36-38 CAG repeats suggests low penetrance in this range. However, given that 36-38 CAG repeats result in the latest average age of onset in clinical cohorts, and another contributing factor may be reduced ascertainment of HD in those of older age. Our data imply that an aging population will lead to higher prevalence rates of HD, and that improved ascertainment of late-onset HD, particularly with 36-38 CAG repeat alleles, may lead to further increases in the future.

2410W

A 78 kb interchromosomal insertion from 8q24.3 into a mutation hotspot palindrome at Xq27.1 causes X-linked Charcot-Marie-Tooth neuropathy (CMTX3). *M.H. Brewer[1,2], R. Chaudhry[1,2], J. Qi[1,3], A. Kidambi[1], A.P. Drew[1], M.M. Menezes[4,5,6], M.M. Ryan[7,8,9], M.A. Farrar[10,11], D. Mowat[11,12], G.M. Subramanian[13], H.K. Young[14,15,16], S. Zuchner[17,18], S.W. Reddel[19], G.A. Nicholson[1,2,20], M.L. Kennerson[1,2,20].* 1) Northcott Neuroscience Laboratory, ANZAC Research Institute, Concord, NSW, Australia; 2) Sydney Medical School, University of Sydney, Camperdown, NSW, Australia; 3) Discipline of Pathology, University of Sydney, Camperdown, NSW, Australia; 4) The Institute for Neuroscience and Muscle Research, The Children's Hospital at Westmead, Westmead, NSW, Australia; 5) T.Y. Nelson Department of Neurology and Neurosurgery, The Children's Hospital at Westmead, Westmead, NSW, Australia; 6) Paediatrics and Child Health, University of Sydney, Camperdown, NSW, Australia; 7) Department of Neurology, Royal Children's Hospital, Parkville, VIC, Australia; 8) Murdoch Childrens Research Institute, Parkville, VIC, Australia; 9) Department of Paediatrics, University of Melbourne, Parkville, VIC, AU Australia; 10) Department of Neurology, Sydney Children's Hospital, Randwick, NSW, Australia; 11) School of Women's and Children's Health, UNSW Medicine, University of New South Wales, Kensington, NSW, Australia; 12) Department of Medical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia; 13) Department of Paediatrics, John Hunter Children's Hospital, Newcastle, NSW, Australia; 14) Department of Paediatrics, Royal North Shore Hospital, St Leonards, NSW, Australia; 15) Northern Clinical School, Sydney Medical School, University of Sydney, St Leonards, NSW, Australia; 16) Department of Neurogenetics, The Children's Hospital at Westmead, Westmead, NSW, Australia; 17) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 18) Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA; 19) Department of Neurology, Concord Repatriation General Hospital, Concord, NSW, Australia; 20) Molecular Medicine, Concord Repatriation General Hospital, Concord, NSW, Australia.

Rationale: Charcot-Marie-Tooth (CMT) disease is an inherited peripheral neuropathy affecting both the motor and sensory neurons. An X-linked subtype, CMTX3, was previously mapped to Xq26.3-q27.3 in two large, multigenerational families. Both families had the same disease haplotype suggesting they shared a disease mutation inherited from a common founder. Extensive mutation scanning of the coding exons, splice-sites and untranslated regions of all annotated genes within the disease locus did not reveal a causative mutation. Therefore we hypothesised that a large structural variation within the CMTX3 locus is the cause of disease. **Objective:** To identify the genetic mutation causing CMTX3 in our two large families using whole genome sequencing (WGS). **Methods and results:** Six individuals (two affected males and one unaffected male from each family) underwent WGS to identify all non-coding point mutations and structural variants in the CMTX3 locus. Analysis of WGS data identified a 78 kb interchromosomal insertion from 8q24.3 into a 180 bp palindrome sequence at Xq27.1 in all four patient samples that was not in the two unaffected controls. The insertion break points were confirmed by Sanger sequencing and the novel mutation segregated with the disease phenotype in the all available members of both families (25 affected males, 30 carrier females and 50 normal family members). The 8q24.3 interchromosomal insertion physically interrupts a gene desert between the genes *LOC389895* and *SOX3*. Preliminary data from patients with CMTX3 showed that mRNA expression of the *FGF13* gene (located ~300 kb from the insertion) is elevated in patient lymphoblast cells. This suggests that the interchromosomal insertion can dysregulate the transcription of nearby genes. Interestingly, large unique interchromosomal insertions that disrupt the same palindrome sequence on chromosome X have been associated with hypoparathyroidism, hypertrichosis, XX sex reversal and ptosis. Comparison of the sequence breakpoints for these different diseases suggests that these Xq27.1 interchromosomal insertions are mediated by double-stranded DNA breaks occurring within the chromosome X palindrome followed by microhomology-mediated template switching repair. **Conclusion:** Our data suggests this large structural variant [der(X)dir ins(X;8)(q27.1;q24.3)] is the underlying genetic cause of CMTX3 in these two families. This is the first report of an interchromosomal insertion causing CMT disease.

2411T

Omics-analysis of an iPSC derived model for Alzheimer's disease (APP p.V717I) exposed to low-energy diet reveals disease specific metabolic changes. *J. Klar[1], J. Schuster[1], L. Laan[1], L.P. Lorenzo[1], J. Nordlund[2], M. Shah-savani[3], A. Falk[3], N. Dahl[1].* 1) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden; 2) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden.

   Background: Early-onset Alzheimer's disease is a complex neurodegenerative disorder that can be caused by pathogenic variants in the gene encoding amyloid precursor protein (APP) in a subset of cases. Recent studies have indicated that ketogenic diet (KD) may have beneficial effects on cognitive functions in AD patients. The diet is believed to have a neuroprotective effect but the underlying mechanisms are poorly understood. **Aim:** We have used patient derived neural cells to investigate the effect of low glucose and ketogenic diet on Alzheimer cells. **Methods:** We have used induced pluripotent stem (iPS) cell derived neural cells to study the effects of KD and reduced glucose levels. Neuroepithelial stem (NES) cells derived from an AD patient who carries the familial p.V717I "London" mutation and a healthy control were differentiated towards neurons in low glucose (2 mM), in KD conditions (2 mM glucose + 6 mM beta-hydroxybuturate) and in standard glucose concentrations (20 mM), respectively. After four weeks of differentiation, we analyzed transcriptome, metabolome, parameters for differentiation and cell death. **Results**: We detected 1255 differentially expressed genes when comparing healthy control neurons and AD neurons cultured in low glucose (2 mM) conditions, and 3178 genes when comparing control vs. AD under KD conditions. Analysis for GO categories revealed extra-cellular matrix structural constituents (GO:0005201) and lipid transporter activity (GO:0005319) as the most are dysregulated functional modules in AD cells. Metabolome analysis confirmed strong differences in intracellular lipid levels when comparing AD neurons with healthy control neurons. **Conclusion:** Our data indicate that low glucose and/or KD result in dramatic changes in neuronal cell metabolism with specific characteristics for cells with the APP p.V717I variant. The specific metabolic responses in neuronal cells with a genetic predisposition to AD may reveal pathways and factors for further therapeutic considerations.

2412F

Somatic mosaicism in Nemaline myopathy detected by next generation sequencing. *T. Yokoi[1], M. Minatogawa[1], C. Abe[1], Y. Tsurusaki[1], Y. Enomoto[1], K. Watanabe[2], T. Naruto[3], K. Kurosawa[1].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Kanagawa, Japan; 2) Division of Neurology, Kanagawa Children's Medical Center, Kanagawa, Japan; 3) Department of Stress Science, Tokushima University Faculty of Medicine, Tokushima, Japan.

   <Background>Nemaline myopathy (NM) represents a group of clinically and genetically heterogeneous neuromuscular disorders. Different clinical-genetic entities have been characterized in recent years, with implications for diagnostics and genetic counseling. In this report we experienced a case of NM with somatic mosaicism. <Case report>He was born at 42 weeks' gestation, weighing 3018 g. He was admitted to neonatal intensive care unit because of apnea and feeding difficulty. He showed hypotonia, contraction of hand and knee, high arch, micrognathia and long thin finger. Electromyogram of extremities showed low amplitude. Muscle biopsy of quadriceps were performed. The result showed Nemaline bodies observed on Gomori trichrome stain and electron microscopic examination which are typical findings for NM. <Method>Using the patient's DNA from peripheral blood, we performed Mendelian disorders related gene-based Next Generation Sequencing (TruSight One, 4813 genes) on the Illumina Miseq. <Results> We identified missense mutation in *ACTA1*; c.478G>A (read depths 10/29: 33%), p.G160S. The pathogenicity of G160S is severely deleterious by ANNOVAR. Sanger sequencing demonstrated that mutation was *de novo*. <Discussion> This is the first report of the patient of NM with somatic mosaicism. Generally, mosaicism of genetic disorder shows milder manifestation. Similarly, He shows milder manifestation than expected. Motor development of this patient is gradually improved as he grew up. Even though the pathological findings are typical for NM, genetic identification for mosaicism in sample of other tissue can be useful for prediction the prognosis. NM is inherited in an autosomal dominant or autosomal recessive manner. Although genotype-phenotype correlation remains poorly defined in NM, genetic analysis is crucial to identify the inheritance and useful for genetic counseling. In conclusion, next generation sequencing is useful for the diagnostics and the detection of mosaicism for NM.

2413W

Initial, 24 week results of heparan sulfate levels in cerebrospinal fluid, brain structural MRI and neurocognitive evaluations in an open label, phase I/II, first-in-human clinical trial of intravenous SBC-103 in mucopolysaccharidosis IIIB. *C.B. Whitley[1], M.L. Escolar[2], S. Vijay[3], G. Parker[4], X. Zhang[5], A. Cinar[5], G. Bubb[5], K.C. Patki[5], S. Rojas-Caro[5].* 1) Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA; 2) Children's Hospital of Pittsburgh, University of Pittsburgh Medical Center, Pittsburgh, PA, USA; 3) Birmingham Children's Hospital, Birmingham, United Kingdom; 4) Bioxydyn Limited, Manchester, United Kingdom; 5) Alexion Pharmaceuticals, Inc., New Haven, CT, USA.

   Purpose: Mucopolysaccharidosis IIIB (MPS IIIB; Sanfilippo syndrome type B) is caused by a marked decrease in alpha-N-acetyl-glucosaminidase (NAGLU) enzyme activity which leads to the accumulation of heparan sulfate (HS) in the brain and other organs, progressive brain atrophy, neurocognitive decline and behavioral disturbances. **Methodology**: This was an open label study with 3 parallel dose groups. SBC-103 [0.3 mg/kg (n=3), 1 mg/kg (n=4), and 3 mg/kg (n=4)] was administered as an IV infusion every other week in patients aged 2-12 years (developmental age ≥1 year) for 24 weeks. Primary objective: assessment of safety and tolerability of IV SBC-103. Secondary objectives included effect of SBC-103 on total HS levels in cerebrospinal fluid (CSF), brain structures (MRI) and neurocognitive status, and pharmacokinetic (PK) profile of SBC-103. **Results**: Eleven patients were enrolled (median age 4 years; range 2-10 years) and received IV SBC-103. During the 24 weeks, there were 3 treatment-emergent serious adverse events (SAEs) in a single patient and 6 infusion-associated reactions in 3 patients. No SAEs were considered related to SBC-103. At week 24, total HS percent change from baseline (mean [SD]) in the CSF was 10.9 [14.7], -0.43 [11.9] and -26.2 [20.9] for 0.3 mg/kg, 1 mg/kg, and 3 mg/kg groups, respectively. HS reduction in CSF was linearly correlated with SBC-103 serum PK exposures. Preliminary volumetric MRI (cortical gray matter volume) and neurocognitive outcomes (age equivalent/developmental quotient) suggested potential for disease stabilization in subjects receiving the 3 mg/kg QOW for 24 weeks compared to their baseline. In the 3 mg/kg group, 3 out of 4 (75%) subjects had an increase or no change in cortical gray matter volume and 3 out of 4 (75%) subjects had an increase in age equivalent/developmental quotient compared to their baseline. Overall, response profiles among the 3 mg/kg treatment groups suggest a potential dose effect as compared to the 0.3 mg/kg and 1 mg/kg groups. **Conclusions**: These initial observations suggest that IV SBC-103 was biologically active and well tolerated in MPS IIIB patients. Potential blood brain barrier penetration of IV SBC-103 was illustrated by PK/PD correlation for HS reduction in CSF. Preliminary evidence of disease stabilization in pharmacodynamics (PD)/efficacy was suggested after 24 weeks of IV SBC-103 in some patients. These data support plans to test higher doses in the study.

2414T

Investigation of genes of the mTORC1 pathway in familial focal epilepsies. *S. Baulac[1], E. Marsan[1], E. Leguern[1], F. Picard[3], S. Ishida[1], M. Chipaux[2], M. Pende[4], S. Weckhuysen[1].* 1) ICM INSERM U1127, Paris, France; 2) Department of Pediatric Neurosurgery, Fondation Rothschild, Paris, France; 3) Department of Neurology, University Hospitals of Geneva (HUG), CH-1211 Geneva 14, Switzerland; 4) Institut Necker-Enfants Malades, CS 61431, Paris, France.

   Abstract 150-200 words   Epilepsy is characterized by spontaneous and recurrent seizures and affects ~1% of the worldwide population. Combining genetics and functional studies, we aim to unravel the molecular basis of familial forms of focal epilepsies. We use family-based whole-exome sequencing and gene panels to search for new genes. In a field so-far dominated by mutations in genes encoding ion channels and neurotransmitter receptors, we identified mutations in genes (*DEPDC5*, *NPRL2*, *NPRL3*,) encoding repressors of the mechanistic target of rapamycin (mTOR) pathway, which controls protein synthesis, cell growth and proliferation. Mutations are found in a range of phenotypes from non-lesional focal epilepsies to epilepsies associated with brain cortical malformations such as focal cortical dysplasia. Concomitantly, we confirmed mTOR hyperactivation by immunohistochemistry in resected brain tissue from patients carrying mutations. Besides, we generated and characterized a constitutive knockout rat model of *Depdc5*. While the *Depdc5*-/- rats are lethal due to mTOR upregulation, the *Depdc5*+/- rats reproduce the disease at the histological level, which is rescued by rapamycin, an mTOR inhibitor. Altogether, *Depdc5* knockout rats exhibit multiple features of rodent models of mTORopathies, and thus, stand as a relevant model to study their underlying pathogenic mechanisms. .

2415F

Wnt signalling holds the molecular key to unlocking a 70 year old Aicardi Syndrome mystery. *T.T. Ha[1], R. Burgess[2], M. Newman[3], M. Li[4], A. Gardner[5], D. Pham[5], R. Kumar[5], C. Reid[6], S. Petrou[4], N. Smith[5,7], S. Berkovic[6], C. van Eyk[9], S. Calvert[10], S. Malone[10], M. Ryan[6,7], V. Shashi[11], R. Leventer[6], S. Mandelstam[6], S. Petrovski[8], M. Lardelli[3], I.E. Scheffer[2,6], J. Gecz[1,5,9,12], M. Corbett[5,9].* 1) School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia; 2) Epilepsy Research Centre, University of Melbourne, Heidelberg, Victoria, Australia; 3) Alzheimer's Disease Genetics Laboratory, University of Adelaide, Adelaide, South Australia,, Australia; 4) Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia; 5) School of Medicine, University of Adelaide, Adelaide, South Australia, Australia; 6) Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia; 7) Department of Neurology, Women's and Children's Hospital, North Adelaide, South Australia, Australia; 8) Department of Medicine, Royal Melbourne Hospital, Parkville, Victoria, Australia; 9) Australian Collaborative Cerebral Palsy Research Group, University of Adelaide, Adelaide South Australia, Australia; 10) Department of Neurosciences, Lady Cilento Children's Hospital, South Brisbane, Queensland, Australia; 11) School of Medicine, Duke University, Durham, North Carolina, United States; 12) South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia.

   Aicardi Syndrome (OMIM 304050; AIC) is a rare neurodevelopmental disorder defined by chorioretinal lacunae, infantile spasms and a spectrum of brain malformations including agenesis of the corpus callosum. AIC has traditionally been considered an X-linked male lethal disorder. The molecular mechanism(s) underlying the disease are not well understood despite considerable research. We aimed to identify the genetic and molecular cause(s) of AIC. We performed whole exome sequencing (WES) on nine proband-parent trios and five additional probands with AIC. We enriched for likely causative variants based on: predicted pathogenicity, amino acid conservation, variant allele frequency and clinical significance. We identified a *de novo* variant in four individuals, each in a different gene: *SLF1, HCN1, SZT2* and *WNT8B*. In one female we found a paternally inherited variant in *TCF4,* a gene implicated in Pitt-Hopkins Syndrome. Using *in vitro* assays we showed that the *HCN1* variant produced a left-shift in the voltage-dependence of activation consistent with a loss of function. The *WNT8B* variant had a dominant negative effect leading to a 1.5-fold reduction in luciferase activity (p-value < 0.05). The effects of *SZT2* and *SLF1* variants are as yet unknown. From expression databases, we found all five genes are expressed in the mouse brain; *Slf1* and *Wnt8b* are tightly coexpressed in the forebrain. Finally, we investigated morpholino knockdown of the genes, and a gene (*TEAD1*) previously implicated in AIC in zebrafish studies. In embryos at 72 hours post fertilisation, we observed a unifying morphant phenotype of AIC-like eye and brain defects among 3/5 genes tested. Our data demonstrate that AIC is a genetically heterogenous malformation-eye-epileptic encephalopathy syndrome and challenges the dogma that AIC is X-linked. We show that the Wnt signaling pathway, which is fundamental to eye and brain development, is likely a major contributor to the molecular pathogenesis of AIC. Collectively, the genes that others and we implicated in AIC show convergence on TCF/LEF mediated transcription. Both the Wnt (*WNT8B, TCF4*) and Hippo (*TEAD1*) signaling pathways can act through this family of transcription factors. Furthermore, SZT2 has been shown to interact with TCF4 by high throughput yeast two-hybrid assay. Our results likely hold the key to unlocking the 70 year old mystery behind the molecular pathogenesis of AIC.

2416W

The FHF1 epileptic encephalopathy. *F. Hamdan[1], S. Al-Mehmadi[2], M. Splitt[3], V. Ramesh[4], S. DeBrosse[5], K. Dessoffy[5], F. Xia[6,7], Y. Yang[6,7], J. Rosenfeld[6], The DDD study[8], P. Cossette[9], J. Michaud[1,10,11], P. Campeau[1,10,11], B. Minassian[2].* 1) Sainte Justine Hospital Research Center, Montreal, Quebec, Canada; 2) Program in Genetics and Genome Biology and Division of Neurology, Department of Pediatrics, The Hospital for Sick Children, and University of Toronto, Ontario, Canada; 3) Institute of Genetic Medicine, International Centre for Life, Newcastle upon Tyne, UK; 4) Pediatric neurology, Newcastle General Hospital, Newcastle upon Tyne, UK; 5) Center for Human Genetics, UH Case Medical Center, Cleveland, OH, USA; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 7) Baylor Miraca Genetics Laboratories, Houston, TX 77030, USA; 8) The Deciphering Developmental Disorders study, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 9) Division of Neurology, CHUM Notre-Dame, Hospital University of Montreal, Montreal, Quebec, Canada; 10) Department of Pediatrics, University of Montréal, Quebec Canada; 11) Department of Neurosciences, University of Montreal, Quebec, Canada.

   Voltage-gated sodium channels (Nav) are mainstays of neuronal function and mutations in the genes encoding CNS Navs [Nav1.1 (SCN1A), Nav1.2 (SCN2A), Nav1.3 (SCN3A), Nav1.6 (SCN8A)] are causes of some of the most common and most severe genetic epilepsies and epileptic encephalopathies (EE). This suggests that other genes encoding Nav–binding proteins, such as fibroblast growth factor homologous factors (FHFs), may be also involved in these pathologies. Recently, Siekierska et al. (2016) reported the identification of a germ-line mosaic heterozygous de novo mutation (NM_004113.5:c.155G>A, p.R52H) in *FGF12*, which encodes FHF1, in 2 siblings with early onset EE. Functional studies showed that the p.R52H missense confers a gain-of-function, increasing the depolarizing shifts in Nav1.6 voltage-dependent fast inactivation. These findings indicated that the p.R52H mutation was likely causal of the EE phenotype; however, the identification of additional similarly affected unrelated cases was still required to confirm the role of FHF1 in the disease. Here, using whole genome and exome sequencing, we report the identification of 3 additional unrelated EE cases each of which carries the de novo FHF1 p.R52H mutation. Phenotype comparison of our 3 cases and the 2 siblings described by Siekierska et al. indicates that the core FHF1 p.R52H encephalopathy comprises neonatal-onset persistent intractable epilepsy and moderate-to-severe intellectual disability. Radiologically, neurodegeneration, especially cerebellar, is present, which, beyond a mutational consequence, appears to be aggravated by the severity and frequency of status epilepticus (SE). It may also be exacerbated by treatment of SE; indeed, all four patients with cerebellar atrophy, including the siblings reported by Siekierska et al., were on Phenytoin, an anti-epileptic drug whose toxicity was previously reported to cause cerebellar abnormalities. In conclusion, our findings confirm the role of FHF1 in early onset EE and allow a better profiling, from infancy to adulthood, of the phenotypes caused by its recurrent p.R52H mutation.

2417T

Recessive *AP3B2* mutations cause severe developmental delay, epilepsy, cortical visual impairment and hyperkinetic motor disorder. *D. Josifova[1], A. Onoufriadis[2], R. Williams[3], A. Barnicoat[4], R. Robinson[5], A. Siddiqui[6], M. Simpson[2]*. 1) Clinical Genetics, Guy's Hospital, 7th Floor, Borough Wing, London, United Kingdom; 2) Department of Medical & Molecular Genetics, Division of Genetics and Molecular Medicine, Faculty of Life Sciences & Medicine, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond London SE1 9RT, United Kingdom; 3) Department of Neurology, Evelina Childrens' Hospital, Guy's and St. Thomas NHS Trust, Westminster Bridge Road, London SE1 7EH, United Kingdom; 4) Department of Clinical Genetics, Great Ormond Street Hospital, Great Ormond Street, London WC1N 3JH, United Kingdom; 5) Department of Neurology, Great Ormond Street Hospital, Great Ormond Street, London WC1N 3JH, United Kingdom; 6) Department of Radiology, St. Thomas Hospital, Guy's and St. Thomas NHS Trust, Westminster Bridge Road, London SE1 7EH, United Kingdom.

Next generation sequencing technology has revolutionised the study of the human genome and led to rapid aetiological resolution of many rare genetic conditions. By using this approach, we have identified a homozygous mutation in *AP3B2* [OMIM 602166] in three individuals from two branches of a consanguineous Turkish family, affected with severe global developmental delay, epilepsy with onset in infancy, cortical visual impairment, and progressive hyperkinetic motor disorder. *AP3B2* codes for a subunit of the clathrin-associated adaptor protein, AP-3. Structurally a hetero-tetramer, AP-3 plays a role in protein sorting in the late-Golgi/trans-Golgi network and endosomes. Unlike other subunits, *AP3B2* is expressed exclusively in neurons. By inference, mutated AP3B2 protein was anticipated to give rise to a neurological phenotype, and this is the first report of clinical cases associated with mutations in this gene in humans. Given the significant recurrence risk for carrier parents, it is an important gene to consider in appropriate cases, to establish an aetiological diagnosis, to provide counselling and facilitate reproductive options. Identification of further patients will help understand the breadth of the clinical phenotype.

2418F

Coexistence and coinheritance of two noble mutations, G325A of *KDM6A* causing Kabuki syndrome and G1877V of *SCN1A* gene causing GEFS+ in a family, confused with Dravet syndrome. *C. Lee*. Division of Pediatric Neurology, Department of Ped, College of Medicine, Eulji Unive, Seoul, Nowon-gu, South Korea.

Objective Co-existence of two rare genetic syndromes is extremely rare. Generalized epilepsy with febrile seizures plus (GEFS+) is a genetic epileptic syndrome with a variable clinical phenotype spectrum for which a variety of causative mutations have been identified. Kabuki syndrome (KS) is a rare, multiple congenital anomaly syndrome with two molecular subtypes of *MLL2* (KS1) and *KDM6A* (KS2). **Study design** We performed a genetic analysis of a family with GEFS+ and suspicious KS. First, a cytogenetic study including G-band karyotyping and SNP microarray was performed. Next, molecular analysis using targeted-exome sequencing and Sanger sequencing were conducted for molecular confirmation and differential diagnosis in the family. **Results** The proband, a 7-year-old male was initially misdiagnosed with Dravet syndrome, and presented with GEFS+, dysmorphic facial features, short status, developmental delay, and intellectual disability. His mother and her elder sister had FS+ and dysmorphic facial features. G-band karyotyping and SNP microarray of the proband were normal. Targeted exome sequencing identified two noble missense mutations: G-to-C transition at single nucleotide 974 (p.G325A) in *KDM6A* responsible for KS and GA-to-TT at two nucleotides 5630-5631 (p.G1877V) in *SCN1A* responsible for GEFS+. **Conclusion** This family is the first in the literature with the co-existence and dominant co-segregation of GEFS+ and KD2 to be confirmed molecularly. Furthermore, two affected female patients with X-linked KS2 showed a partial escape X-inactivation pattern of *KDM6A* with milder phenotypes than the male affected proband in this study. .

2419W

New *SMARCA2* mutation in Nicolaides Baraitser syndrome and Myoclonic astatic epilepsy. *S. Tang[1], E. Hughes[2], K. Lascelles[2], M.A. Simpson[1], D.K. Pal[1], EuroEPINOMICS RES Myoclonic astatic epilepsy working group.* 1) King's College London, London, London, United Kingdom; 2) King's Health Partners, London, UK.

  Myoclonic astatic epilepsy is a genetically heterogenous rare childhood epilepsy syndrome characterised by the onset of myoclonic, myoclonic-atonic or atonic seizures between the ages of 6 months to 6 years. Prognosis is variable from a complete seizure remission to a drug resistant epileptic encephalopathy. Approximately 10% of cases have an identifiable genetic aetiology but the majority remain unexplained. We report a *de novo SMARCA2* missense mutation c.3721C>G, p.Gln1241Glu (ENST00000349721 NM_0030703), on whole exome sequencing and confirmed by Sanger sequencing in a patient with myoclonic astatic epilepsy. This novel variant is not reported in 1000G, ExAC or EVS and in the ExAC database there are only very rare missense and stop gain variants in the region (MAF<0.0005). Annotation from insilico prediction tools was SIFT 0.184 (tolerated), PolyPhen2 0.900 (possibly damaging), mutation taster 0.999 (disease causing) and CADD 22.5 (top 1% most deleterious). A specific role for *SMARCA2* mutations has been implicated in Nicolaides Baraitser syndrome. Nicolaides Baraitser syndrome is a well delinated phenotype with sparse hair, microcephaly, typical facial morphology, brachydactyly, prominent interphalangeal joints, intellectual disability with marked language impairment and seizures. Following the identification of the mutation, the patient was reassessed and observerved to have features consistent with Nioclaides Baraitser syndrome: sparse hair, skin wrinking, slightly thick distal phalanges in the hands, specific craniofacial features, reduced speech, intellectual disability with a verbal IQ 51 (0.1th centile), performance IQ 57 (0.2th centile) and full scale IQ 47 (<0.1th centile) and autism.   This *de novo SMARCA2* missense mutation c.3721C>G, p.Gln1241Glu is the only reported mutation on exon 26 and the second mutation outside the ATPase domain of SMARCA2 to be associated with Nicolaides Baraitser syndrome. We hypothesize that this mutation similarly affect ATP hydrolysis and chromatin remodeling. We report that the epilepsy phenotype in Nicolaides Baraitser syndrome can be consistent with myoclonic astatic epilepsy syndrome. This study also underlines the fact that, in rare genetically heterogenous condidtions like myoclonic astatic epilepsy, whole exome sequencing can lead to unexpected clinical diagnosis.

2420T

***PIGG* is associated with developmental delay, early onset seizures and cerebellar ataxia.** *A.C. Thuresson[1], J. Zhao[1], J. Halvardson[1], A. Knaus[2], P. Krawitz[2], P. Georgii Hemming[1,3], P. Baeck[4], L. Feuk[1].* 1) Dept. of Immunology, genetics & Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Institute for Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany; 3) Dept of Molecular Medicine and Surgery, Karolinska Institute, Karolinska University Hospital Solna, Stockholm, Sweden; 4) Childrens Clinic, County Hospital, Kalmar, Sweden.

  Glycosylphosphatidylinositol (GPI) is a glycolipid that tethers more than 150 different proteins to the cell surface. Aberrations in biosynthesis pathway of GPI anchors cause congenital disorders of glycosylation with clinical features including moderate to severe developmental delay, seizures and facial dysmorphism. To date mutations in 12 genes of the GPI-anchor synthesis pathway have been reported in association with multi-system disease. Here we present two siblings from a consanguineous family with symptoms including development delay, cerebellar hypoplasia, cerebellar ataxia, early onset seizures and minor facial dysmorphology. Microarray analysis indicated several large homozygous regions across the genome, making up approximately 4% of the genome sequence. Using exome sequencing, we identified a novel homozygous nonsense mutation, p.Trp539*, in the gene Phosphatidylinositol Glycan Anchor Biosynthesis, Class G (PIGG), located within a stretch of homozygous sequence. Both siblings have a lower expression of PIGG compared to their parents, and the parent show a clear reduction in expression of the mutated allele, indicative of nonsense-mediated decay. Mutations in several other GPI anchor synthesis genes lead to a reduced expression of GPI-APs that can be measured by flow cytometry. Here, the patients' fibroblasts and granulocytes were tested for GPI-APs CD55 and CD59. Surface expression of the GPI anchor itself was measured using fluorochrome conjugated aerolysin (FLAER). No significant differences could be detected in the patients' blood, however, in fibroblasts FLAER test showed a generally reduced level of GPI anchor on the patient cells. Based on genetic and functional evidence we propose that mutations in PIGG are affecting the GPI biosynthesis pathway and cause a congenital disorder of glycosylation.

2421F

Diagnostic exome sequencing of Danish families with rare genetic diseases. *J. Ek, L. Risom, E. Østergaard, M. Dunø.* Department of Clinical Genetics, Copenhagen University Hospital, Copenhagen, Denmark.

Background: The advancement in DNA sequencing has in recent years highly increased the likelihood of identifying causative variants among patients with various disorders. Whole Exome Sequencing (WES) facilitates an unbiased screening for genetic defects. We have implemented WES for the genetic work-up of patients primarily under suspicion for neurological and mitochondrial disorders, or complex syndromes involving intellectual disability. A precise molecular diagnosis is important for treatment, follow-up, prognosis, and a prerequisite for prenatal testing and preimplantation genetic testing. The aim of this study was to increase the number of patients receiving a genetic diagnosis and further to evaluate in which groups of patients WES gives the highest diagnostic yield. Materials and methods: This is a retrospective study of *in-house* whole exome sequencing conducted at the Department of Clinical Genetics at Copenhagen University Hospital in the period from first of January 2014 and forth, and includes more than 60 probands and their families. The majority of the surveyed probands were included on the suspicion of neurological disorders and mitochondrial disorders or on the basis of intellectual disability. All the patients/families were prior to analysis thoroughly assessed by medical specialists and counselled by a clinical geneticist, in order to provide detailed and complete clinical information, aiding variant classification. Results: Causative mutations were identified in approx. 35% of the patients. One fifth of the mutations had arisen de novo or was found in mosaic form in one parent. Approx. 75% of the identified mutations were absent from the literature and common mutation databases, i.e. novel mutations. Four variants were initially classified as incidental findings of which two were reported back to the patient/parents. Interpretation: In this study, WES facilitated a genetic diagnosis in approx. 35% of patient with complex etiologies such as neurological disorders, intellectual disability, or mitochondrial disorders. The tight collaboration and discussions between clinicians and laboratory scientists is instrumental in reaching a conclusive diagnosis. WES will move to the forefront of genetic work-up and become the first tier analysis for disorders of complex aetiology. Most likely, this will result in an even higher diagnostic yield of WES, when it is used in the initial screening in patients with suspected genetic disorders.

2422W

Novel *de novo* mosaic triallelic mutation in *SPAST* may result in autosomal dominant spastic paraplegia-4. *A.M. Matthews[1,2,3], M. Tarailo-Graovac[1,2,3], E.M. Price[2,3], I. Blydt-Hansen[4], W.P. Robinson[2,3], C.J. Ross[1,2,3], W.W. Wasserman[1,2,3], H. Siden[3,5,6], C.D. van Karnebeek[1,3,4].* 1) Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 3) Child & Family Research Institute, Vancouver, BC, Canada; 4) Division of Biochemical Diseases, Department of Pediatrics, BC Children's Hospital, Vancouver, BC, Canada; 5) Division of General Pediatrics, Department of Pediatrics, BC Children's Hospital, Vancouver, BC, Canada; 6) Canuck Place Children's Hospice, Vancouver, BC, Canada.

Case study: The index is a 12 year-old male, born to healthy, non-consanguineous parents. At 12 months he presented with speech and motor delay and was noted to be dysmorphic. Spastic paralysis was first observed at 1 year, initially as spastic lower extremity diplegia with normal upper-extremity function. His condition progressed to a state of ascending spastic tetraplegia with significant accompanying contractures. Family history was unremarkable for spastic paralysis and neurological disorders, but remarkable for a paternal great-grandfather's sister with cerebral palsy-like symptoms post-fever. Extensive metabolic testing, karyotype (46,XY) and multiple single gene tests were normal. **Background**: Hereditary Spastic Paraplegia (HSP) is caused by mutations in over 60 genes and is characterized by progressive weakness and spasticity. 45% of HSP cases have been linked to autosomal dominant spastic paraplegia-4 due to heterozygous mutations in Spastin (*SPAST*), generally affecting the lower extremeties. **Methods**: Exome sequencing was performed on DNA extracted from blood for index and parents with data analysis via the TI-DEX semi-automated bioinformatics pipeline. Pyrosequencing was performed to determine the degree of mosaicism in multiple tissues (n=6) from the index and parents. **Results**: Exome sequencing followed by Sanger revealed a triallelic (G/C/T) variant in *SPAST* at chr2:32362241 (hg19); parents were homozygous reference (G/G). Both alternative alleles (NM_014946.3:c.1477G>C p.Asp493His and c.1477G>T p.Asp493Tyr) were predicted to be highly damaging by multiple *in silico* tools. Chromosome micro-array was normal in this region and no alternative alleles were detected in parental blood, suggesting that both alternative alleles arose as *de novo* variants in the index. Based on Pyrosequencing, the percentage of reference alleles in the index ranged from 44% in urine to 58% in saliva and blood. The reference to alternative ratio across tissues suggested the presence of three cell populations, two different heterozygous reference/alternative cells (G/C and G/T) in addition to double mutant cells (C/T). **Conclusion**: This is the first report of an individual with the HSP phenotype due to more than one damaging *SPAST* variant, in the same or different cells. Bi-allelic variant or double mutant cells may explain the extreme phenotype observed in the index; mosaicism on the other hand (with heterozygous variants present as well) potentially explains viability.

2423T

A new form of epileptic encephalopathy caused by mutations of the Arf4 guanine-nucleotide exchange factor *DENND5A.* *A.G. Minassian[1], C. Han[2], R. Alkhater[3], T. Froukh[4], M. Galatti[2], R. Han Lui[2], M. Fotouhi[2], J. Sommerfeld[5], A.J. Alfrook[6], C.R. Marshall[1], S. Walker[1], P. Bauer[5], S.W. Scherer[1], O. Riess[5], R. Buchert[5], B.A. Minassian[3], P.S. McPherson[2].* 1) The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, 686 Bay Street, Toronto, Ontario M5G 0A4, Canada; 2) Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec H3A 2B4, Canada; 3) Program in Genetics and Genome Biology, Department of Pediatrics (Neurology), Hospital for Sick Children, University of Toronto, 686 Bay Street, M5G 0A4, Toronto, Ontario, Canada; 4) Department of Biotechnology and Genetic Engineering, Faculty of Science, Philadelphia University, Jarash Road, Amman, Jordan; 5) Institute of Medical Genetics and Applied Genomics, Rare Disease Center, University of Tübingen, Calwestr. 7, 72076, Tübingen, Germany; 6) Private Clinic, Amman, Jordan.

With a lifetime incidence of 3%, epilepsy is a common neurological disorder. Epileptic encephalopathies are a catastrophic group of epilepsies characterized by refractory seizures and cognitive arrest resulting from abnormal brain development. Here we identify a new epileptic encephalopathy that additionally features cerebral calcifications and coarse facial features, caused by recessive loss-of-function mutations in *DENND5A*. We performed Whole Exome Sequencing (WES) on two sisters affected with epileptic encephalopathy and born of consanguineous parents. We discovered a shared homozygous 2bp frameshift deletion in exon 4 that results in a premature stop codon at amino acid 173 (p.Asp173Profs*8). We then performed functional experiments on DENND5A, which we discovered is expressed predominantly in developing neuronal tissues in rats. Further, we found that loss of function in this gene leads to disrupted trafficking of neurotrophin receptors, enhanced receptor levels and signaling, and massive alterations in neuronal development. Identification of loss-of-function mutations in *DENND5A* in a second family with two affected children, and of predicted damaging compound heterozygous variants absent from any control databases in a third family, all with features closely similar to our original family, further shows that we have identified a novel epileptic encephalopathy gene. Ongoing work on the functions of *DENND5A* will elucidate a heretofore unknown facet of neurodevelopment. Note: The first two authors contributed equally.

2424F

Heterozygous missense mutation in *SEC24A* **encoding a coat protein complex II vesicle associated with autosomal dominant spinocerebellar ataxia.** *T. Morikawa[1], S. Miura[1,2], K. Kosaka[1], R. Fujioka[3], K. Sano[2], A. Yorita[2], K. Aoki[1], Y. Uchiyama[4], T. Taniwaki[2], H. Shibata[1].* 1) Division of Human Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 2) Division of Respirology, Neurology and Rheumatology, Department of Medicine, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 3) Department of Food and Nutrition, Beppu University Junior College, Beppu, Oita, Japan; 4) Department of Radiology, Kurume University School of Medicine, Kurume, Fukuoka, Japan.

Hereditary spinocerebellar ataxia (SCA) comprises a group of clinically and genetically heterogeneous inherited disorders mainly characterized by slowly progressive cerebellar ataxia. We ascertained a Japanese pedigree with ADCA consisting of four family members including three patients. The cardinal clinical features of this family were as follows: i) slow progression; ii) around 40 at onset; iii) gait and limb ataxia; iv) cerebellar dysarthria ; v) gaze evoked nystagmus; vi) saccadic pursuit ; vii) ocular dysmetria ; viii) upward gaze palsy; and iv) hyperreflexia, spasticity, and impaired vibration sense with dominancy in the lower limbs. We first confirmed the absence of repeat expansion in known genes responsible for SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA12, SCA17, and dentatorubral-pallidoluysian atrophy. We performed the whole exome sequencing for the three patients with the average of sequencing depth in the target regions to be 28.7x. We identified 1,143 functional single nucleotide variants (SNVs) shared by the three patients. We identified only two candidate SNVs, P361L in *SEC24A* and N588S in *SGOL2* of which allele frequencies are less than 0.002 in the public variants databases (1000G Project, Ensemble, DBexome) and in 11 in-house exome data, By Sanger sequencing we confirmed that both SNVs were absent in the unaffected individual in the pedigree. We also confirmed the absence of both of the SNVs in 520 healthy Japanese individuals by Sanger sequencing. Only one of the two SNVs, P361L in *SEC24A* was predicted to be damaging by PolyPhen-2 while the other SNV, N588S in *SGOL2* was predicted to be benign. *SEC24A* encodes a component of coat protein II (COPII)-coated vesicles that mediate protein transport from the endoplasmic reticulum. The SNV is located in the highly conserved region in *SEC24A* among vertebrates (SIFT score = 0.02). Since the pathological variants have not been reported in *SEC24A*, this can be the first case of human disease caused by a variant in *SEC24A*. It is notable that SEC24A protein is known to interact with p125 that harbors the DDHD domain. Recently several genes with the DDHD domain have been reported to be responsible for spastic paraplegia (SPG), such as *DDHD1* and *DDHD2*. Therefore, SPG-like symptoms such as hyperreflexia and spasticity observed in the present family can be explained by the DDHD dysfunction caused by the variant of *SEC24A*.

2425W

SLC13A5 is the second gene associated with Kohlschütter-Tönz syndrome. *S. Raskin[1], A. Schossiga[2], A. Bloch-Zupanb[3], N.I. Wolf[4], M. Cohen[5], F. Giuliano[6], J. Jurgens[7], B. Krabichler[2], D. Koolen[8], N.L.M. Sobreira[7], C. Souza[1], C. Joseph[9], S. Perelman[10], E. von Hülsen[11], A. Kohlschütter[12], O. Tönz[13], E. Maurer[2], M. Muller-Bollai[14], J. Penzien[15], J. Zschocke[2], I. Kapferer-Seebacherl[2]. 1) Pontifícia Universidade Católica do Paraná (PUCPR) Curitiba (S.R.), PR; 2) Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria; 3) Université de Strasbourg, Faculté de Chirurgie Dentaire, Strasbourg, France; Hôpitaux Universitaires de Strasbourg (HUS), Pôle de Médecine et Chirurgie Bucco-dentaires, Centre de Référence des Manifestations Odontologiques des Maladies Rares, Strasbourg;; 4) Department of Child Neurology, VU University Medical Center and Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; 5) kbo-Kinderzentrum München gGmbH, Mwunich, Germany; 6) Centre de Référence Anomalies du Développement et Syndromes malformatifs PACA, Service de Génétique Médicale, CHU Nice, Nice, France; 7) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 8) Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands; 9) Département d'Odontologie Pédiatrique, Université de Nice Sophia-Antipolis, CHU Nice, Nice, France; 10) Hôpitaux pédiatriques de Nice CHU-Lenval, Nice; 11) kbo-Kinderzentrum Mwunich, Germany; 12) Department of Pediatrics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 13) Children's Hospital, Lucerne, Switzerland; 14) UFR Odontologie, Département d'Odontologie Pédiatrique, Université de Nice Sophia-Antipolis, UCA France CHU de Nice. Pôle Odontologie, UF soins pour enfants; Laboratory URB2i - EA 4462, Paris Descartes, France; 15) Department of Neuropaediatrics, Klinikum Augsburg, Germany.*

Kohlschütter-Tönz syndrome (KTZS) is a rare autosomal recessive disease characterised by epilepsy, intellectual disability, and yellowish discolouration of the teeth. It is frequently caused by biallelic mutations in *ROGDI*. Here we report that the majority of individuals with *ROGDI*-negative KTZS carry biallelic *SLC13A5* mutations. We studied seven families by clinical and dental evaluation, parametric linkage analysis (one family), and exome and/or Sanger sequencing. Consistent dental findings in *SLC13A5*-KTZS include small and widely spaced teeth with a smooth enamel surface, suggesting smooth-hypoplastic amelogenesis imperfecta. Discolouration of the teeth is variable and may be mild. Dental findings differ from individuals with homozygous *ROGDI* mutations, where the enamel is usually rough and discolouration is more intense due to hypomineralized-hypoplastic amelogenesis imperfecta. Epileptic encephalopathy in *SLC13A5*-KTZS usually presents in the neonatal and (less frequently) early infantile period. We conclude that *SLC13A5* is the second major gene associated with Kohlschütter-Tönz syndrome; careful clinical delineation provides clues for the likely candidate gene. .

2426T

Diagnostic exome sequencing reveals disruptive splice-site variants in VPS13C in a sporadic case of early-onset Parkinson disease. *B. Schormair[1,2], D. Kemlink[3], O. Fiala[3,4], E. Graf[5], T.-M. Strom[5], J. Roth[3], E. Ruzicka[3], J. Winkelmann[1,2,6,7]. 1) Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany; 2) Institute of Human Genetics, Technische Universität München, Munich, Germany; 3) Department of Neurology and Center of Clinical Neuroscience, Charles University, 1st Faculty of Medicine and General University Hospital, Prague, Czech Republic; 4) Institute of Neuropsychiatric Care (INEP), Prague, Czech Republic; 5) Institute of Human Genetics, Helmholtz Zentrum München, Munich, Germany; 6) Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 7) Munich Cluster for Systems Neurology, SyNergy, Munich, Germany.*

Recently, truncating variants in vacuolar protein sorting 13C (*VPS13C*) have been described as a cause of autosomal-recessive early-onset Parkinson disease (EOPD). A total of three unrelated subjects with homozygous or compound heterozygous truncating mutations in *VPS13C* were reported, all with rapid and severe disease progression and early cognitive decline. The identification of additional carriers of such variants in independent EOPD cases is needed to confirm the causal role of *VPS13C* in EOPD. We have performed diagnostic whole exome sequencing in individuals with sporadic EOPD (age of onset ≤ 40 years) of Caucasian ancestry. Exomes were generated with SureSelect XT 50Mbv5 kit and sequenced on a HiSeq2500 system. Alignment was done with BWA and SNVs and small indels were called with SAMTools and Pindel. Annotation was performed with SAMTools, custom Perl scripts and Variant Effect Predictor. For analysis we kept rare variants (minor allele frequency < 0.01 in public databases) which disrupt protein function (splice, nonsense, frameshift). One individual was found to carry two variants affecting canonical splice sites in *VPS13C*: c.[2029+2T>G];[3215-1G>T]. Variant c.2029+2T>G is predicted to alter the splice acceptor site in intron 31, variant c. 3215-1G>T the splice donor site in intron 21 of *VPS13C* (NM_020821.2). Both variants were not present in 7,500 control exomes sequenced on the same platform, the 1000 Genomes Project and the EVS dataset. One heterozygous carrier of the c.2029+2T>G variant was found in the non-Finnish European ExAc population. For the affected individual, 98.2% of the exome target regions were covered at least 20x. The individual does not carry pathogenic or likely pathogenic variants in other PD-associated genes. Age of onset of PD symptoms was 39 years with rigidity and micrographia as the initial symptoms, followed by mild tremor, speech impairment, and mild dystonia. Cognitive changes include attention difficulties, delusions and memory impairment. Due to severe complications, treatment was switched to deep brain stimulation after five years. We are currently performing segregation analysis of the variants in the family and RT-PCR-based analysis of the splicing defects in *VPS13C*. Positive results will provide additional evidence that biallelic disruptive variants in *VPS13C* are a rare cause of autosomal-recessive early-onset Parkinson disease.

2427F

Identification of new genes in consanguineous Tunisian Arab-Berber families with Parkinson disease. *J. Trinh[1], E. Gustavsson[1], I. Guella[1], M. McKenzie[1], S. Bortnick[1], B. Chen[1], S.B. Sassi[2], E. Hentati[2], F. Nabli[2], E. Farhat[2], R. Amouri[2], F. Hentati[2], M.J. Farrer[1]. 1) Medical Genetics, University of British Columbia, Vancouver, BC, BC, Canada; 2) Mongi Ben Hamida National Institute of Neurology, La Rabta, Tunis, 1007, Tunisia.*

Parkinson disease (PD) is an insidious neurodegenerative disorder. LRRK2 p.G2019S is responsible for 30-40% of familial patients from Tunisia. Recessive PINK1 mutations account for 10-15% of familial PD. However, there are many families affected with PD in Tunisia with unknown genetic cause. Eighteen Tunisian Arab Berber families multi-incident for parkinsonism (30 patients) without known pathogenic mutations were investigated. Exome sequencing was performed using the using the Ion AmpliSeq exome kit (57.7Mb) and sequenced on an Ion Proton sequencer (Life Technologies, Carlsbad, CA, USA). Single nucleotide polymorphisms (SNP) were genotyped using the Illumina Multi-Ethnic Genome Arrays (MEGA). Further genotyping was performed on larger patient-control series (2650 patients and 2264 control subjects). We prioritized homozygous variants in the consanguineous families predicted to produce frameshifts, stop codons, missense or splice defects that would disrupt protein function and were rare or novel. We report two consanguineous families that presented juvenile parkinsonism in a Tunisian consanguineous family with a homozygous recessive mutation in sepiapterin reductase (SPR) and hexokinase 3 (HK3). We have also identified multiple low frequency DNAJC6 variants segregating with PD in three families. Subsequent genotyping and sequencing has identified higher frequency of rare variants in these genes compared to controls. Furthermore, a shortlist of candidate genes has been identified from these families and validation is in progress. Identification of novel candidates in consanguineous families will help in developing targeted therapeutic treatment and understanding the basic pathophysiology of the disease.

2428W

Accelerating the novel *NTRK1* variants in patients with HSAN-IV phenotype via whole exome sequencing. *S. Majid[2], R. Al Tassan[1], H. Al Saud[2], T. Masoodi[2], H. Al Dosssari[2], B. Meyer[2], O. Mohammad[1]. 1) Dept. of Genetics, KFSHRC, RIYADH; 2) Dept. of Medical Genetics, KFSHRC, RIYADH.*

Genomic technologies have emerged as powerful tools delving and dissecting human genome that allows the rapid evaluation of the underlying genetic causes of diseases in various motor and sensory neuropathies. Hereditary sensory autonomic neuropathy type IV (HSAN-IV) is a rare autosomal recessive disorder that usually begins in infancy and is characterized by anhidrosis, insensitivity to noxious stimuli leading to self-mutilating behavior, and intellectual disability. HSAN-IV is caused by mutations in the Neurotrophic Tyrosine Kinase, Receptor, Type 1 gene, *NTRK1*, encoding for the high-affinity receptor of nerve growth factor (NGF). Patients with HSAN-IV lack all NGF-dependent neurons, the primary afferents and sympathetic postganglionic neurons responsible for pain sensation and the presence of anhidrosis, respectively. Herein, we report nine patients from nine unrelated families with HSAN-IV due to various mutations in *NTRK1*, five of which are novel. The affected patients had variable intellectual deficits, and some had delay in the diagnosis of HSAN-IV. In addition to being the first report of HSAN-IV from the Arabian Peninsula, this report expands the mutational spectrum of patients with *NTRK1* mutations.

2429T

Clinical implication of the *FMR1* **intermediate alleles in the Spanish population.** *M. Mila[1,2], M.I. Alvarez-Mora[1,2], S. Izquierdo[3], A. Feliu[1], A. Vilumara[1], I. Madrigal[1,2], L. Rodriguez-Revenga[1,2].* 1) Biochemistry and Molecular Genetics, Hospital Clinic, Barcelona, Barcelona, Spain; 2) CIBERER and IDIBAPS; 3) Servicio de Bioquímica Clínica, Genética Clínica. Hospital Miguel Servet, Zaragoza. Spain.

The *FMR1* premutation (55-200 CGG) has mainly been associated with Fragile X-associated primary ovarian insufficiency (FXPOI), and fragile X-associated tremor/ataxia syndrome (FXTAS). Several lines of evidence suggest that carriers of the *FMR1* premutation allele present with a higher risk of medical, psychiatric, and cognitive symptoms than the general population. On the other hand, there is controversy regarding smaller expansions named as "gray zone" (45-54 CGG) or intermediate alleles (IA). Although the frequency of these alleles varies greatly among different populations, several studies have reported an excess of these alleles in patients either with Parkinsonism, cognitive/behavioural phenotypes (CBP) or primary ovary insufficiency (POI). It is well accepted that depending on the AGG interruptions number IA confers an increased risk of expansion. In an attempt to provide new insights on the clinical impact of IA in the *FMR1*-associated phenotypes, we have studied the frequency of IA in three pathologies, spinocerebellar ataxias (SCAs), POI and CBP, in Spain. Among 5115 CBP patients 93 IA were detected (1.82%), 5 among 304 patients with spinocerebellar ataxia (1.64%) and 9 among the 346 patients presenting POI (2.60%). Regardless of whether we take into account patient or allele frequencies (ID, 1.77%; SCAs, 1.12%; POI, 1.30%), the results obtained did not differ from those observed in the control population (1.71%). Although our results do not support an excess of IA in the *FMR1*-associated pathologies, it is advisable to refer individuals carrying an IA to a genetic counsellor in order to study the AGG interruptions and evaluate the possible risk of expansion of these alleles during transmission.

2430F

A genetic survey of Charcot-Marie-Tooth disease in Turkey. *E. Battaloglu[1], A. Candayan[1], H.I. Akcay[2], H. Durmus[2], M. Sivaci[1], F. Deymeer[2], P. Oflaz-er-Serdaroglu[2], Y. Parman[2].* 1) Department of Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey; 2) Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey.

Charcot-Marie-Tooth (CMT) (MIM 118220) disease is the most common hereditary polyneuropathy that is characterized with distal weakness and atrophy. More than 40 CMT genes have been identified that is a challenge for genetic diagnostics especially in populations with frequent consanguineous marriages as in Turkey.   In this study, we investigated 310 Turkish CMT patients using direct gene sequencing and thirty three of these patient samples were analyzed by whole exome sequencing (WES). Mutations were identified in 159 patients (51%). Eighty percent had *PMP22* duplication or *GJB1, GDAP,* or *MFN2* mutations. *PMP22* duplication was the most common (46%) cause and *SH3TC2, MTMR2, MTMR13, FGD4, KIFB1, DNM1, and HSPB1* mutations were rarely observed. Age of onset was 1- 57 with a mean of 13.85±12.2 among genetically diagnosed cases and the inheritance patterns were autosomal dominant (62%), X-linked (13%), or autosomal recessive (21%). Isolated cases constituted about 2% of this group. Most common CMT type was CMT1 (81/159) that was followed by CMT2 (28/159), CMT4 (28/159), and CMTX (22/159). Mean age of onset was higher in CMT1 than in other types, but age of onset was also heterogeneous in this group of patients that was in between 4-57 years. Majority of CMT1 patients were carrying *PMP22* duplication (93%) and five had *MPZ* mutations. These patients usually showed typical phenotype that was relatively benign compared with other subtypes and almost all patients remained ambulatory throughout their life. Most of the recessive cases (CMT4 and CMT2) had an earlier age of onset and severe clinical picture. The causative gene could not be identified in 151 patients (49%) most of which were isolated cases. For the 33 patients analyzed by WES, genetic diagnosis was possible in 13 (39%). Absence of mutations in known CMT genes in 49% of our cohort implicates further genetic heterogeneity for the disease. WES analysis is a suitable approach to be used for genetic diagnosis of this genetically heterogeneous disorder especially for isolated cases and in countries with high rate of consanguineous marriages. .

2431W

First report of mutations in human *CLDN5*: A neurological phenotype associated with disruption of blood-brain barrier tight junctions. *C. Cytrynbaum[1], S. Walker[2], A.G. Minassian[2], C. King[3], R. Loh[4], G. Jones[5], H. Lango Allen[5], S. Ellard[5], S.W. Scherer[2], R. Newbury-Ecob[3], R. Weksberg[1].* 1) Div Clinical & Metabolic Gen, Hosp Sick Children, Toronto, ON, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON Canada; 3) Department of Clinical Genetics, University Hospitals Bristol NHS Foundation Trust, Bristol, UK; 4) Royal United Hospital, Bath, UK; 5) Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK.

Claudins are a family of proteins that are crucial structural and functional components of tight junctions (TJ). TJs have important roles in regulating paracellular permeability and maintaining cell polarity in epithelial and endothelial cell sheets. The claudins interact with multiple proteins and are intimately involved in signal transduction to and from the tight junction. Several human diseases have been shown to be caused by mutations in claudin genes (i.e. *CLDN1*, *CLDN14*, *CLDN16*, *CLDN19*). Claudin-5 (*CLDN5*) is the predominant claudin expressed in endothelial TJs which are particularly important in vascular endothelium of the blood-brain barrier (BBB). Here we report two unrelated patients with similar clinical features including microcephaly, brain calcifications and seizures who were both found to have *de novo* heterozygous single base substitutions in *CLDN5*. The first patient had intractable epilepsy and died at 6 months of age. The second patient has global developmental delay (not sitting independently and not verbal at 21 months of age). Both patients have missense mutations in exon 1b (p.Gln148Lys and p.Val126Met respectively) resulting in changes to highly conserved amino acids. In silico prediction programs (SIFT/Mutation Tester) predict both variants to be deleterious. Several claudin-deficient mouse models have been generated and the diversity of phenotypes observed demonstrates their important roles in the maintenance of tissue integrity in various organs. CLDN-5-deficient mice demonstrate size-selective loosening of the BBB. It is possible that impairment of the BBB, even in a size selective manner, could be harmful to CNS activity. In fact, Claudin-5 null mice died within 10 hours of birth, although the underlying cause was unclear. Based on the role of *CLDN5* in the BBB, we propose that mutations in *CLDN5* lead to a severe neurologic phenotype as seen in our patients. Functional studies including transcriptional and protein assays in fibroblast cells from the second patient are ongoing.

2432T

Cerebral vascular malformations in patients with *ACTA2* mutations. *C. Abe[1,2], T. Yokoi[2], M. Takagi[3], Y. Enomoto[2], Y. Tsurusaki[2], T. Naruto[4], T. Goto[3], H. Ueda[5], Y. Fujii[6], N. Aida[6], K. Kurosawa[2].* 1) Pediatric Neurology, Kasugai, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center; 3) Department of Neurology, Kanagawa Children's Medical Center; 4) Institute of Health Biosciences, The University of Tokushima Graduate School; 5) Department of Cardiology, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Department of Radiology, Kanagawa Children's Medical Center, Yokohama, Japan.

α2-actin protein, encoded by *ACTA2*, is expressed in vascular smooth muscle cells of various organs. Heterozygous *ACTA2* mutations are reported to cause aortic aneurysm diseases and strokes. Strokes caused by *ACTA2* mutations indicate severe cerebral arteriopathy that is different from other diseases. Detailed radiological investigation including the neurovascular survey should be considered in the patients supposed to be caused by the *ACTA2* mutations. Here we report on 2 patients with *ACTA2* mutation representing cerebral vascular malformations. Patient 1 had a history of patent ductus arteriosus (PDA) ligation at 2 months old and aneurysm resection of left brachial artery at 4 months old. At age of 13 years, he presented dilatation of Valsalva sinus and started taking angiotensin-converting-enzyme inhibitor (ACEI). He also presented transient left paresis. His cerebral MRI revealed reduced diffusion in diffusion-weighted images (DWIs) involving the territories of the right middle cerebral artery (MCA) that is consistent with infarction. Magnetic resonance angiography (MRA) also showed narrowing of the right vertebral artery, and fusiform dilatation of CAs. After medication for brain infarction, his symptoms improved and he showed only slight hypoesthesia of left body at the time of discharge. Patient 2 was 11 years old. At age of 4 days, she showed poor sucking and diagnosed to have PDA, dilatation of ascending aorta, and congenital midriasis. She had an episode of nonsymptomatic dilatation of urinary duct at 6 months old. At age of 6 years, MRA showed irregular forms of MCAs and 5 years later MRI showed chronic ischemic lesions in parietal lobe. MRA also showed diffuse fusiform dilatation of both CAs and narrowing in the branches covering the multiple areas. We performed a targeted high-throughput sequencing of genes. We identified heterozygous mutations c.772C>T: p.R258C in patient 1 and c.536G>A: p.R179H in patient 2 in ACTA2, respectively. Sanger sequencing confirmed the mutations. Specific features of cerebral images (MRIs and MRAs) in both patients are confident with previous reports. ACTA2 mutations cause vascular malformations in multiple organs including central nervous system (CNS). Cerebral infarctions found in the patients are part of CNS involvements resulted from vascular anomalies. Molecular analysis and neurovascular investigation are important for the patients with the disorder.

2433F

A bioinformatic pipeline for studying non-coding variants in disease: Application to hereditary sensory and autonomic neuropathy patient genomes. *M.H. Couse[1], F.R. Zahir[1], G. Horvath[1], C. Shyr[1], M. Tarailo-Graovac[1], C. Nislow[2], C. van Karnebeek[1], J.M Friedman[1].* 1) BC Children's Hospital, Vancouver, Canada; 2) University of British Columbia, Vancouver, Canada.

Hereditary and sensory autonomic neuropathies (HSAN) are a group of clinically and genetically heterogeneous disorders of the peripheral nervous system. We are working with eight children affected by early-onset HSAN, whose symptoms include loss of pain sensation, developmental delay, and gastrointestinal complications. We have performed whole genome sequencing on these patients and analyzed coding variants, but have only reached a diagnosis in two cases. We hypothesize that the pathogenic variants in these patients lie in regulatory regions of genes associated with HSAN. We investigated rare variation in regulatory elements of the genome associated with 50 hereditary neuropathy genes: promoters, enhancers, untranslated regions, and regions sensitive to purifying selection. We associated distal regulatory elements to genes of interest by their presence within topological associated domains encompassing hereditary neuropathy genes. We implemented a custom bioinformatics pipeline that filters rare regulatory SNVs, indels, and SVs from patient whole genome vcf and bam files. We found a single heterozygous SNV in one patient at a conserved locus in the 5'UTR of *PMP22* that is predicted by CADD and FATHMM-MKL scores to be deleterious. Although the phenotype of this patient is not typical of the classical CMT1A seen with *PMP22* duplications or of the variant CMT phenotypes reported in association with missense mutations of the *PMP22* coding segment, PMP22 is dosage-sensitive and can cause a broad range of phenotypes. Further validation assessment of our non-coding variant is in progress. The pipeline developed for this study is easily modified to analyze other rare genetic diseases, and may be valuable for discriminating non-coding, pathogenic variants in future studies.

2434W

WDR45 **mutations in three male patients with West syndrome.** *M. Nakashima[1], K. Takano[2,3], Y. Tsuyusaki[2], S. Yoshitomi[4], M. Shimono[5], Y. Aoki[6], M. Kato[7,8], N. Aida[9], T. Mizuguchi[1], S. Miyatake[1], N. Miyake[1], H. Osaka[2,10], H. Saitsu[1,11], N. Matsumoto[1].* 1) Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 4) Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka, Japan; 5) Department of Pediatrics, University of Occupational and Environmental Health, Kitakyushu, Japan; 6) Department of Pediatrics, Asahi General Hospital, Asahi, Japan; 7) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 8) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 9) Department of Radiology, Kanagawa Children's Medical Center, Yokohama, Japan; 10) Department of Pediatrics, Jichi Medical University, Tochigi, Japan; 11) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

West syndrome is an early-onset epileptic encephalopathy characterized by clustered spasms with hypsarrhythmia seen on electroencephalogram. West syndrome is genetically heterogeneous, and its genetic causes have not been fully elucidated. *WDR45* resides on Xp11.23, and encodes a member of the WD repeat protein interacting with phosphoinositides (WIPI) family, which is crucial in the macroautophagy pathway. *De novo* mutations in *WDR45* cause beta-propeller protein-associated neurodegeneration characterized by iron accumulation in the basal ganglia. In this study, we performed whole exome sequencing of individuals with West syndrome and identified three *WDR45* mutations in three independent males. Two novel mutations occurred *de novo* and the remaining mutation detected in a male patient and his affected sister was inherited from the mother, harboring the somatic mutation. The three male patients showed early-onset intractable seizures, profound intellectual disability and developmental delay. Their brain magnetic resonance imaging (MRI) scans showed cerebral atrophy. We found no evidence of somatic mosaicism in the three male patients. Our findings indicate that hemizygous *WDR45* mutations in males lead to severe epileptic encephalopathy.

2435T

The epidemiological and molecular analysis of Paroxysmal Kinesigenic Dyskinesia (PKD) in Japan. *N. Kurotaki[1], Y. Morimoto[1], K. Kurotaki[1], Y. Kusumoto[1], K. Shirotani[3], N. Iwata[3], H. Ozawa[1], K. Yoshiura[2].* 1) Neuropsychiatry, Nagasaki University School of Medicine, Nagasaki, Japan; 2) Department of HUMAN GENETICS, Atomic Bomb Disease Institute, Nagasaki University , Nagasaki, Japan; 3) Genome-based Drug Discovery, Nagasaki Univ. Sch. of Pharmaceut. Sci., Nagasaki, Japan.

Paroxysmal kinesigenic dyskinesia (PKD) is characterized by recurrent, brief attacks of involuntary movement triggered by sudden voluntary movement. In 2011, PRRT2 on chromosome 16 is responsible for PKD reported by the Chinese group. A year later, our group found that the mutation is almost concentrated into c.649dupC (p.Arg217Profs*8) and involved in benign infantile seizure (Ono et al., 2012). Additionally, several reports show that PRRT2 is related to not only infantile seizure, but also migraine, and other epileptic disorders. The papers of both Wood H. in the Nat Rev Neurology (2012) and one of Scheffer IE.'s in the Nat Rev Neurology (2013) emphasized the PRRT2 gene in clinical neurology. In this situation, we proposed epidemiological study of PKD in Japan and have performed the molecular analysis to estimate the relationship between PKD and its complications. First, we conducted the survey by using a questionnaire about the frequency and the complications of PKD, and the medical treatments of the members of both Japanese Society of Neurology and Japanese Society of Child Neurology. Japanese Society of Psychiatry and Neurology kindly recommended the study. A total of all members are 6000 and the response rate of a questionnaire was about 50%. In this survey, 426 cases of PKD have reported by members of these societies. Interestingly, 147 cases had various complications including epileptic seizure, autistic disorders, bipolar disorders and other neurological disorders. Until now, we analyzed PRRT2 mutations in 68 cases resulting c.649dupC of PRRT2 is main gene alterations as reported previously. However, some cases have no mutation of PRRT2. We will perform further investigation utilizing the next generation sequence technology in the cases without any PRRT2 mutation. Otherwise, we are estimating Clinical features of pure PKD and PKD with other complications combined with genotype of each case. Our study will show the genotype - phenotype correlation of PKD. Also, we could shed light on the molecular genetic aspect of neuropsychiatric diseases including autistic disorders, mood disorders based on the results of molecular analysis of PKD. This study received the ethical approval from the Committee for Ethical Issues in Human Genome and Gene Analysis at Nagasaki University, Japan. Also, this study is supported by Health Labour Sciences Research Grant.

2436F

Heterozygous mutations in *TUBB2A* associated with brain abnormalities, severe delay and joint contractures. *R. Ejaz[1], A. Lionel[2], S. Blaser[4], C. Hawkins[5], S. Walker[2], S. Scherer[2], R. Babul-Hirji[1], C. Marshall[2,3], D. Stavropoulos[3], D. Chitayat[1,6].* 1) Division of Clinical & Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Genome Diagnostics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Pathology, Mount Sinai Hospital, Toronto, Ontario, Canada; 6) The Prenatal Diagnosis and Medical Genetics Program; Department of Obstetrics and Gynecology, University of Toronto, Toronto, Ontario, Canada.

The formation of the mammalian cortex requires the generation, migration, and differentiation of neurons. The discovery that mutations in tubulin genes such as *TUBA1A*, *TUBB2B* and *TUBB3* cause different neurodevelopmental conditions, including lissencephaly, polymicrogyria, and an ocular motility disorder shows the important role of these genes in brain formation and function. We described a third individual with a germline *TUBB2A* mutation, with the hitherto undescribed phenotype of joint contractures, facial dysmorphism, vocal cord paralysis, severe developmental delay, seizures, optic nerve hypoplasia and MRI findings of polymicrogyria. The patient was born at term to a 21-year-old primigravida mother via emergency caesarean section for fetal distress. Prenatal ultrasounds were unremarkable. He was hypotonic with overriding sutures, prominent nasal bridge, downslanted palpebral fissures, midface hypoplasia, low-set and posteriorly-rotated ears, and a short neck. Musculoskeletal examination revealed camptodactyly, second and fifth fingers overlapping the third and fourth, single palmar creases, right rocker bottom foot, and joint contractures. Brain MRI at 2 days of life showed bilateral posterior sylvian polymicrogyria with asymmetric ventricles and frontal gyration. Ophthalmologic assessment revealed bilateral optic nerve hypoplasia. The patient had global developmental delay, contractures and developed seizures at 19 months of age. He died at 28 months of age of acute bronchopneumonia. Given the association of polymicrogyria and optic nerve hypoplasia with *TUBA8* mutations, the suspicion of a tubulinopathy was raised. However, as the child had other features not previously reported in tubulinopathies, whole exome sequencing (WES) was pursued and revealed a de novo *TUBB2A* (NM_001069) mutation [c.G785A:p.R262H], in the same highly conserved GTP-binding domain as the two previously reported cases. Our patient further supports the association between heterozygous mutations in the *TUBB2A* gene and abnormal cortical development. While the two previously reported individuals had simplified gyral patterning, our patient had polymicrogyria-like cortical dysplasia, with microcephaly, facial dysmorphism, and musculoskeletal anomalies. Seizures and developmental delay were shared. Our case suggests a broader phenotype to *TUBB2A* mutations than currently reported. More reports are needed for further delineation of the manifestations associated with this gene mutation.

2437W

A missense mutation in the *PPIG* gene encoding the peptidyl-prolyl isomerase G in patients with autosomal dominant benign adult familial myoclonic epilepsy. *R. Fujioka[1], S. Miura[2], K. Yamada[3], T. Morikawa[3], H. Motooka[4], K. Uchiyama[5], Y. Aso[6], N. Kimura[6], K. Sano[2], K. Yamamoto[7], T. Taniwaki[2], H. Shibata[3].* 1) Department of Food and Nutrition, Beppu University Junior College, Beppu, Oita, Japan; 2) Division of Respirology, Neurology, and Rheumatology, Department of Medicine, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 3) Division of Genomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 4) Department of Neurpsychiatry, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 5) Department of Radiology, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 6) Department of Neurology, Faculty of Medicine, Oita University, Yufu, Oita, Japan; 7) Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Fukuoka, Japan.

Benign adult familial myoclonic epilepsy (BAFME) is an inherited epileptic syndrome characterized by adult-onset cortical tremors and seizures. Here, we investigate the responsible variant for an autosomal dominant BAFME ascertained from Oita prefecture, Japan using exome sequencing combined with linkage analysis. The cardinal clinical features of the present family were as follows: 1) slow progression; 2) myoclonic tremor; 3) very low frequency of the occurrence of generalized seizure; 4) upper motor neuron sign; 5) bradycardia; 6) deep sensory involvement in the lower limbs. By the multipoint linkage analysis of seven family members including five patients, we identified ten regions located in eight chromosomes (2, 3, 4, 8, 11, 14, 15, and 19) in which the logarithm of odds (LOD) scores were over 1.0. Subsequently, we performed exome sequencing of the seven family members with the average depth of 88.7x. From the total of 6,062,467 variants observed, we identified 23 patient-specific novel nonsynonymous variants. These variants were then filtered with the criteria of the minor allele frequency < 0.4% in the 1000G Project and the absence in 30 in-house exomes of unrelated Japanese individuals. From the remaining 13 nonsynonymous variants, we selected three variants located within the ten linkage regions. By Sanger validation of the 3 candidates, we confirmed only two variants, His429Asp in *PPIG* and Val240Leu in *LRTM1* are cosegregated with the disease in our pedigree. We further confirmed their absence in 500 healthy unrelated Japanese individuals. Since the variant in *PPIG* was predicted to be damaging (SIFT score = 0.04, PolyPhen-2 = possibly damaging), while the variant in *LRTM1* was predicted to be benign (SIFT score = 0.36, PolyPhen-2 = benign), we concluded the novel single nucleotide substitution, c.1505C>G (p.His429Asp), located in the *PPIG* gene, is highly likely to be the causative variant in our BAFME pedigree. *PPIG* encodes peptidyl-prolyl isomerase G also known as cyclophilin G that is a member of the cyclophilin family of peptidyl-prolyl cis/trans isomerases. PPIG is expressed in multiple tissues including brain and is known to be associated with transcription elongation, as well as splicing. This is the first report of a human disease related to PPIG.

2438T

Adenylate kinase 9 (*AK9*) mutation is a novel disease gene for congenital myasthenic syndrome with limb girdle myasthenia. *C.W. Lam.* Department of Pathology, The University of Hong Kong, Hong Kong, China.

Neuromuscular junction (NMJ) is clinically important because abnormal neuromuscular transmission can cause myasthenic syndrome. Single-gene defects causing myasthenic syndrome are called as congenital myasthenic syndrome (CMS). A special type of CMS causes only limb-girdle myasthenia. Here, we aim to determine the molecular basis of CMS in a consanguineous family with limb girdle type of CMS. There were two clinically affected siblings (P1, P2) and two clinically unaffected sisters (S1 and S2). Homozygosity mapping was performed using high-density single-nucleotide polymorphism microarray. Mutations were identified by WES and WGS and genotypes of the family members were confirmed by Sanger sequencing. A region of homozygosity (ROH) of 20 MB was mapped on chromosome 6q15–21. This ROH region is the only ROH region that is present in all the clinically affected siblings (P1 and P2), absent in all the clinically unaffected siblings (S1 and S2). Bioinformatics analysis of WES and WGS data showed a novel single-nucleotide variant (SNV) of the adenylate kinase (*AK9*) gene located in this region. P1 and P2 were homozygous for the SNV, S1 is heterozygous for the SNV, and the SNV is absent in S2. The parents were heterozygous for the SNV. This mutation was not present in EXAC, ESP6500, 1000 Genomes, and ClinVar databases.This SNV creates a cryptic transcriptional signal in intron 5 of the *AK9* gene by creating a start codon, i.e., start-gain mutation. An *AK9* knockout zebrafish model created by TALEN-mediated gene targeting showed fatigue in swimming consistent with the human CMS phenotype. The gene *AK9* is a novel disease gene for limb girdle myasthenia and CMS. AK9 is located in the cytosol and is involved in production of NDP and NTP. NTP is required for synthesis of dolichol phosphate and nucleotide-sugars for N-glycosylation of proteins. This novel link of *AK9*, N-glycosylation will pave a new way for studying the biology of NMJ and synapses. Since genes of the N-glycosylation pathway are associated with CMS with limb-girdle myasthenia, i.e, *GFPT1*, *DPAG1*, *ALG14*, *ALG2, GMPPB,* increasing the nucleotide pools in NMJ by dietary supplements may be a new therapeutic option of treating patients with CMS with limb girdle myasthenia.

2439F

Biallelic mutations of *VAC14* in pediatric onset neurological disease.
G.M. Lenk[1], K. Szymanska[2,3], G. Debska-Vielhaber[4], M. Rydzanicz[5], A. Walczak[5], M. Bekiesinska-Figatowska[6], S. Vielhaber[4], K. Hallmann[7], P. Stawinski[5,8], S. Buehring[9], D.A. Hsu[10], W.S. Kunz[7], R. Ploski[5], M.H. Meisler[1]. 1) Human Genetics, Univ Michigan, Ann Arbor, MI; 2) Dept of Child Psychiatry, Warsaw Medical Univ, Poland; 3) Dept of Exp and Clin Neuropathology, Medical Research Center, Polish Academy of Sciences, Warsaw, Poland; 4) Klinik für Neurologie, Universitätsklinikum Magdeburg, Magdeburg, Germany; 5) Dept of Medical Genetics, Warsaw Medical University, Warsaw, Poland; 6) Dept of Diagnostic Imaging, Institute of Mother and Child, Warsaw, Poland; 7) Klinik für Epileptologie and Life&Brain Center, Universitätsklinikum Bonn, Bonn Germany; 8) Institute of Phys and Path of Hearing, Nadarzyn, Poland; 9) Madison, Wisconsin; 10) Dept of Neurology, Univ of Wisconsin, Madison, Wisconsin.

   The low abundance endo-lysosomal lipid $PI(3,5)P_2$ is generated from $PI(3)P$ by a protein complex with 3 major subunits: the lipid kinase PIKFYVE, the phosphatase FIG4, and the scaffold protein VAC14. Mutations of *FIG4* result in several inherited neurological disorders. We now describe inherited variants of *VAC14* in two unrelated children with neurological dysfunction. At the ages of 18 months and 3 years the children experienced sudden onset of progressive neurological decline characterized by dystonia and regression of developmental milestones. Both became nonambulatory and nonverbal. Striatal abnormalities were present on MRI. A diagnosis of Leigh syndrome was rejected due to normal lactate profiles. Exome sequencing identified biallelic variants of *VAC14* that were inherited from unaffected heterozygous parents in both families. Proband 1 inherited a splice site mutation that results in skipping of exon 13, p.Ile459Profs*4 (not reported in public databases), and the missense variant p.Trp424Leu (reported in one heterozygote in ExAC). Proband 2 inherited two missense mutations in the dimerization domain of VAC14 that result in the amino acid substitutions p.Ala582Ser and p.Ser583Leu (not reported in public databases). Cultured skin fibroblasts from the probands exhibited the accumulation of vacuoles that is characteristic of $PI(3,5)P_2$ deficiency. Vacuolization was rescued by transfection of wildtype *VAC14* cDNA. The similar age of onset and similar neurological progression in these two unrelated children suggest that a specific recessive disorder results from compound heterozygosity for deleterious variants of *VAC14*. To characterize the functional consequences of these missense variants, we are examining the biochemical interactions of the mutant proteins. We are interested in evaluating novel variants of VAC14 identified in individuals with neurological dysfunction. (Supported by R01 NS34509 and R03 NS090142).

2440W

Mutations in *Citron-Kinase* cause recessive micro-lissencephaly with multinucleated neurons.
A. Moccia[1], B. Harding[2], O. Soukarieh[3], H. Tubeuf[3], S. Drunat[4,5], L. Chitty[6], A. Verloes[4,5], P. Gressens[5,7,8], V. El Ghouzzi[5], S. Joriot[9], S. Passemard[4,5,7], F. Di Cunto[10], A. Martins[3], S. Bielas[1]. 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania and Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, U.S.A. 19104; 3) Inserm-U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Interactive Biosoftware, Rouen, France 76183; 4) Département de Génétique, Hôpital Robert Debré, Paris, France 75019; 5) Inserm U1141, Protect, Hôpital Robert Debré, Paris, France 75019; 6) Genetics and Genomic Medicine, UCL Institute of Child Health and Great Ormond Street NHS Foundation Trust, London, U.K. WC1N 1EH; 7) Université Paris Diderot, Hôpital Robert Debré, Paris, France 75019; 8) Center for Developing Brain, King's college, St Thomas' Campus, London, United Kingdom WC2R 2LS; 9) Service de Neuropédiatrie, Centre Hospitalier Régional Universitaire de Lille, Lille, France 59037; 10) Department of Molecular Biotechnology and Health Sciences, University of Turin, Italy 10126.

   Primary microcephaly is a neurodevelopmental disorder caused by a reduction in brain size attributed to defects in proliferation of neural progenitor cells during development. Mutations in genes encoding proteins that localize to the mitotic spindle and centrosomes have been implicated in the pathogenicity of primary microcephaly. In contrast, the contractile ring and midbody required for cytokinesis, the final stage of mitosis have not previously been implicated in the molecular mechanisms of this phenotype. Citron Kinase (CIT) is a multi-domain protein that localizes to the cleavage furrow and midbody of mitotic cells, where it is required for the completion of cytokinesis. Rodent models of *Cit* deficiency highlighted the role of this gene in neurogenesis and microcephaly over a decade ago. Here we identify recessively inherited pathogenic truncating and splice site variants in *CIT* as the genetic basis of severe microcephaly and neonatal death. We present postmortem data showing that CIT is critical to build a normally sized human brain. Consistent with cytokinesis defects attributed to CIT, multinucleated neurons are observed throughout the cerebral cortex and cerebellum of an affected proband, expanding our understanding of mechanisms attributed to primary microcephaly.

2441T

NR1H3 **and** *P2RX4/P2RX7* **mutations cause multiple sclerosis in families.** *C. Vilariño-Güell[1], W. Song[2], A.L. Traboulsee[3], J.S. Wiley[4], B.J. Gu[4], A.D Sadovnick[1,3].* 1) Medical Genetics; 2) Psychiatry; 3) and Division of Neurology, University of British Columbia, Vancouver, British Columbia, Canada; 4) Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia.

   Introduction: Multiple sclerosis (MS) is an inflammatory disease characterized by myelin loss and neuronal dysfunction. Despite the aggregation observed in some families, pathogenic mutations have remained elusive. **Objectives:** The goal of this study is to identify pathogenic mutations responsible for Mendelian forms of MS and variants of major effect on disease risk. **Methods:** Exome sequencing analysis was applied to patients from multi-incident families; and variants of interest validated in a case control series consisting of 2211 MS patients and 880 healthy controls. **Results:** Exome analysis identified pathogenic mutations for MS in *NR1H3* and a *P2RX7-P2RX4* haplotype. *NR1H3* (LXRA) is nuclear receptors that controls the transcriptional regulation of genes involved in inflammation and innate immunity. A LXRA p.R415Q substitution was observed in seven MS patients from two multi-incident families. All patients presented a severe and progressive form of disease, with an average age at onset of 34 years. The p.R415Q position is highly conserved in orthologs and paralogs, and disrupts LXRA heterodimerization and transcriptional activation of target genes. Protein expression analysis revealed that mutant LXRA alters gene expression profiles, suggesting a disruption in transcriptional regulation as one of the mechanisms underlying MS pathogenesis. Additionally, association analysis of common variants in *NR1H3* identified rs2279238 conferring a 1.35-fold increased risk of developing progressive MS. Purinergic receptors are proinflammatory mediators containing genetic variants associated with MS susceptibility risk. In this study we identified a three mutation haplotype encompassing purinergic receptors *P2RX7* (p.T205M and p.N361S) and *P2RX4* (p.G135S) in a multi-incident family with high incidence of MS. This haplotype was observed in five blood-related patients with an average age at onset of 31 years, and a clinical course consistent with relapsing-remitting or secondary progressive MS. Functional analysis revealed that this disease haplotype causes over 95% loss in P2RX7 pore function, and almost completely abolishes its phagocytic activity. This loss of function results from near absence of P2RX7 surface expression due to impaired trafficking or accelerated turnover. **Conclusion:** Exome sequencing analysis of multi-incident MS families can be successfully used for the identification of pathogenic mutations, and to unravel the underlying molecular mechanisms of disease.

2442F

Independent variant analysis of *TEAD1* **and** *OCEL1* **in 38 Aicardi Syndrome patients.** *B.K.Y. Wong[1,6], V.R. Sutton[2], R.A. Lewis[2,3,4,5], I.B. Van den Veyver[1,2,6].* 1) Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Medicine, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Department of Ophthalmology, Baylor College of Medicine, Houston, TX; 6) Jan and Dan Duncan Neurological Research Institute Texas Children's Hospital.

   Aicardi syndrome is a severe neurodevelopmental disorder characterized by infantile spasms, typical chorioretinal lacunae, agenesis of the corpus callosum, and other neuronal migration defects. It has been reported recently that *de novo* variants in *TEAD1* and *OCEL1* each may cause Aicardi syndrome in a single individual of a small cohort of females with this clinical diagnosis. These data were interpreted to suggest that the clinical diagnosis of Aicardi syndrome may be genetically heterogeneous. To investigate this further, we sequenced *TEAD1* and *OCEL1* coding regions using DNA from 38 clinically well-characterized girls with Aicardi syndrome. We did not detect the previously reported or any other deleterious variants in any of the analyzed samples. This suggests that the published variants represent either an extremely rare cause of Aicardi syndrome or an incidental finding.

2443W

Biallelic mutations in *UBA5* **reveal the UFM1 cascade as a new pathway for early-onset encephalopathy.** *E. Colin[1,2], J. Daniel[3], A. Ziegler[1,2], J. Wakim[2], A. Scrio[4], T.B. Haack[11], S. Khiati[2], A.S. Denomme[1,2], P. Amati-Bonneau[1,2], M. Charif[2], V. Procaccio[1,2], P. Reynier[1,2], K.A. Aleck[5], L.D. Botto[6], C.L. Herper[3], C.S. Kaiser[3], R. Nabbout[7], S. N'Guyen[8], J.A. Mora-Lorca[9], B. Assmann[10], S. Christ[10], T. Meitinger[11,12], T.M. Strom[11,12], H. Prokisch[11,12], A. Miranda-Vizuete[9], G.F. Hoffmann[11], G. Lenaers[2], P. Bomont[4], E. Liebau[3], D. Bonneau[1,2], The FREX Consortium.* 1) Department of Biochemistry and Genetics, University Hospital of Angers, Angers, France; 2) Mitolab, PREMMi, Mitovasc, UMR CNRS 6214-INSERM 1083; 3) Department of Molecular Physiology, Westfälische Wilhelms-University Münster, Münster, Germany; 4) Avenir-Atip team, IN-SERM U1051, Institute of Neurosciences of Montpellier, University of Montpellier, Montpellier, France; 5) Department of Genetics and Metabolism, Phoenix Children's Medical Group, Phoenix, United States; 6) Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City, United States; 7) Department of Pediatric Neurology, National Reference Center for Rare Epilepsies, University Hospital Necker-Enfants-Malades, Paris, France; 8) Department of Pediatric Neurology, University Hospital, Angers, France; 9) Institute of Biomedicine of Seville, University Hospital Virgen del Rocío/CSIC/University of Seville, Seville, Spain; 10) Department of General Pediatrics, Division of Pediatric Metabolic Medicine and Neuropediatrics, University Hospital Heidelberg, Heidelberg, Germany; 11) Institute of Human Genetics, Technische Universität München, München, Germany; 12) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.

   Using whole-exome sequencing, we identified rare autosomal recessive variants in *UBA5* in 5 children from 4 unrelated families affected with a similar pattern of severe intellectual deficiency, microcephaly, movement disorders and/or early-onset intractable epilepsy. *UBA5* encodes the E1-activating enzyme of the ubiquitin-fold modifier 1 (UFM1) protein, a recently identified ubiquitin-like protein. Most members of the UFM1 cascade and target proteins are localized in a large protein complex at the luminal site of the endoplasmic reticulum (ER) and are involved in the regulation of the unfolded protein response (UPR) and ER-stress mediated apoptosis. The UFM1 cascade has also been involved in the development of various cancers. However, the specific biological function of ufmylation and the clinical implications of its dysfunction remain largely uncharacterized. Biochemical studies of mutant UBA5 proteins and studies in fibroblasts from affected individuals revealed that *UBA5* mutations impair the process of ufmylation, resulting in an abnormal endoplasmic reticulum structure. In Caenorhabditis elegans, knockout of *uba-5* and of human orthologous genes in the UFM1 cascade alter cholinergic, but not glutamatergic, neurotransmission. In addition, *uba5* silencing in zebrafish decreased motility while inducing abnormal movements suggestive of seizures. These clinical, biochemical, and experimental findings support *UBA5* mutations as a novel pathophysiological cause for early-onset encephalopathies due to abnormal protein ufmylation.

2444T

The changing face of type 2 Gaucher disease: A natural history study. *T.R. Roshan Lal[1,2], K. Weiss[2], G. Monestime[2], N. Tayebi[2], G. Lopez[2], E. Sidransky[2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

   Background Gaucher disease (GD), the most common lysosomal storage disorder, results from the inherited deficiency of the enzyme glucocerebrosidase. Neuronopathic type 2 GD is the rarest, as well as the most severe and progressive form. Traditionally, the clinical presentation is considered to be stereotypic; however it can range from hydrops fetalis to the collodion baby phenotype to infants presenting after 6 months of life. Neurological involvement occurs very early on in the disease process, resulting in severe degeneration and death in infancy or early childhood. However in the past decades, infants have been living longer due to more aggressive interventions. The purpose of this study is to delineate the changing clinical course of type 2 GD. Establishing this data may help with the management of these affected individuals as well as their families Also, the collected information will provide an essential baseline for future treatment protocols. **Methods** A cohort of both living and deceased patients with clinically and molecularly confirmed type 2 GD is being identified from patient databases, patient support groups and referring physicians. A structured telephone interview consisting of in-depth medical questions will be conducted with parents of affected type 2 GD individuals. Questions asked during the telephone interview include information regarding disease presentation, disease progression, surgical history, medications, feeding issues, family history, treatment or interventions, disease complications, hospice care, cause of death and the impact of disease on other family members. Medical records will be reviewed when available. **Results** Currently available treatment options, while not curative, appear to be changing the clinical course of type 2 GD. Affected individuals are living longer, but interventions do not appear to markedly impact the neurological manifestations. This study has just begun and is anticipated to highlight newer complications or disease manifestations seen in these patients. **Conclusion** Diagnosing and treating infants with type 2 GD has always been a challenge to clinicians. It is paramount that we understand the baseline and any changes in the clinical course of this disease, as it often impacts not just the patient, but also the entire family. The data generated will be essential in the design of any future interventional protocols, helping to identify goals of treatment and milestones to determine efficacy.

2445F

Examining the risk of Parkinsonism in six sibling pairs with Gaucher disease. *E. Sidransky[1], G. Monestime[1], C. Groden[1], E. Wiggs[1], N. Tayebi[1], A. Zimran[2], G. Lopez[1].* 1) NHGRI, Bethesda, MD; 2) Gaucher Clinic, Share Zedek Medical Center, Jerusalem, Israel.

   Background While the association between mutations in the gene encoding glucocerebrosidase (*GBA1*) and parkinsonism is well established, most *GBA1* mutation carriers do not develop parkinsonian features. In Parkinson disease (PD), by the time patients present with motor symptoms, more than 50% of dopaminergic neurons in their substantia nigra have been lost. Identifying patients at highest risk for developing PD is critical in order to impact patients' prognoses and quality of life. **Methods** We studied six sibling pairs with Gaucher disease longitudinally at the National Institutes of Health Clinical Center. Five of the pairs had Ashkenazi Jewish background. In each discordant sibling pair (DSP), both siblings had the same genotype in *GBA1*, but only one sibling was clinically diagnosed with PD. In all 12, neurological evaluations, family pedigrees, neuroimaging, neuropsychological testing and olfactory testing were performed and motor and non-motor symptoms of PD were assessed. Medical information including age at onset, genotype, clinical presentation, family history, and current medications were also collected. DNA and RNA samples were banked. Subjects were evaluated 1-4 times over 1-10 years. **Results** Five genotypes were detected; N370S/N370S (in two pairs), N370S/L444P, N370S/V394L, c.84insG/N370S and N370S/IVS2+1. The general severity of Gaucher disease was determined based on the degree of skeletal involvement, hematological abnormalities, and organomegaly. In four of the six pairs, the sibling with PD was younger and had milder Gaucher symptoms. No subtle parkinsonian features were observed in any of the non-PD sibs, even among those who underwent multiple sequential evaluations. **Conclusions** Continued longitudinal evaluations and recruitment of additional DSPs will help to power this study and elucidate factors contributing to and/or preventing the development of parkinsonism in these individuals. Genetic studies will explore potential genetic modifiers or environmental factors that may diminish or augment risk. .

2446W

Arthrogryposis associated with a homozygous frameshift variant in *TOR1A*: Fetal-onset dystonia may cause severe congenital joint contractures. *R.D. Clark, J.A. Gold, J. Camacho.* Division of Medical Genetics/Peds, Loma Linda University Medical Center, Loma Linda, CA.

   Regions of homozygosity (ROH), identified by SNP microarray, increase the chance of autosomal recessive (AR) disease due to a homozygous mutation at a locus within the ROH. A term female, born to healthy parents, in the breech presentation had severe arthrogryposis, lack of spontaneous movement, tremors and bulbar dysfunction. She was spastic and hyperreflexic, did not nipple and was fed by G-tube. She had a right inguinal hernia, femoral head dislocation and bilateral acetabular dysplasia. An EMG documented anterior horn cell involvement. MRI of the brain showed colpocephaly with asymmetric enlargement of the occipital and temporal horns and thinning of the corpus callosum. Parents were from the same small town in Mexico but denied consanguinity. A SNP microarray revealed 46.3 Mb of ROH in several chromosomes, including 9q33.2q34.11. The patient died at 4 months of age. Results from clinical exome sequencing, available posthumously, revealed only one homozygous variant, in *TOR1A*: c.790_793del, p.Asp267Ilefs*13, which was reported to be "likely pathogenic" because a frameshift introduced a premature stop codon. *TOR1A* encodes torsin A, which is expressed throughout the CNS in humans. Heterozygous mutation in *TOR1A* causes torsion dystonia type 1 (DYT1, MIM 128100, 605204), a variable autosomal dominant (AD) trait of childhood onset. The common mutation in *TOR1A*, an in-frame GAG deletion in exon 5, has reduced penetrance for dystonia of 30-40%. Arthrogryposis has not previously been associated with *TOR1A* sequence changes. We propose that homozygosity for this variant *TOR1A* allele caused arthrogryposis in our patient due to fetal-onset of severe dystonia that rendered her almost immobile. Other AD genes that cause dystonia may be candidate genes for AR arthrogryposis, especially when a hypomorphic allele or other modifying factor reduces the penetrance. Finally, when ROH are interrogated for genes that could explain a rare phenotype, both AR and AD traits should be considered.

2447T

Family exome sequencing in two founder populations for personalized genetic diagnostics and novel disease gene identification. *C. Gonzaga-Jauregui[1], E. Puffenberger[2], K. Brigatti[2], C. Van Hout[1], A. King[1], K. Williams[2], K. Praveen[1], C. Perry[1], E. Maxwell[1], M. Kuebler[2], J. Staples[1], J.G. Reid[1], J.D. Overton[1], A. Baras[1], S. Mellis[3], A. Shuldiner[1], F. Dewey[1], K.A. Strauss[2].* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY; 2) Clinic for Special Children, Strasburg, PA; 3) Regeneron Pharmaceuticals Inc, Tarrytown, NY.

Founder populations derive from a small number of individuals and are often culturally or geographically isolated from gene inflow from cosmopolitan populations. Old Order Amish and Mennonite populations maintain a high-degree of cultural and genetic isolation, making them ideal for discovering links between gene variation and disease. We have performed exome sequencing in families of Amish and Mennonite ancestry seen for genetic evaluation at the Clinic for Special Children. We sequenced the exomes of the proband, parents, and all available affected and unaffected siblings (i.e. 'family exome'). We performed pedigree-based variant analyses and segregation in order to identify candidate disease genes and pathogenic alleles in these probands. To date, we have sequenced and analyzed more than 100 probands and their relatives in order to identify the molecular causes of their genetic disease. As expected for founder populations, we identified variants in recessive disease genes, including a case of two recessive disorders segregating together in this population (*SLC26A2*, diastrophic dysplasia; and *SH3CT2*, Charcot-Marie-Tooth type 4C) to produce an unusually severe, early-onset neuro-skeletal syndrome. Interestingly, 55% of the cases where we were able to assign a genetic diagnosis were due to novel *de novo* mutations in recently reported disease associated genes such as *ARID1B*, *SETBP1*, *PURA*, *SYNGAP1*, *TCF4*, *CHD7*, *CTNNB1*, *TUBB3*, among others. In addition, we identified potential novel disease gene associations in 30 probands with previously unreported rare disorders. We demonstrate the utility of genomic sequencing in individuals with genetic disorders and their family members to reach a fast and accurate molecular diagnosis and guide treatment. As in any population, a broad spectrum of pathogenic molecular mechanisms is observed, ranging from autosomal recessive homozygous variants specific to these relatively genetically homogeneous populations to sporadic *de novo* variants. Genetic testing proves especially useful for the evaluation of developmental disabilities and polygenic disease mechanisms. We identified novel candidate disease gene associations that may inform molecular diagnosis for other cases with similar clinical presentations.

2448F

Exome utility in phenotype/genotype association for 6,769 exomes spanning >500 phenotypes. *S.N. Jhangiani[1], Z.H. Coban Akdemir[2], J. Posey[2], W. Charng[2], T. Gambin[2], N. Sobreira[3], K. Doheny[3], C. Boehm[3], D. Avramopoulos[3], A. Hamosh[3], T. Harel[2], E. Karaca[2], Y. Bayram[2], M. Eldomery[2], M.N. Bainbridge[1], H. Dinh[1], Y. Han[1], H. Doddapaneni[1], J. Hu[1], V.R. Sutton[2], D.M. Muzny[1], E.A. Boerwinkle[1], D. Valle[3], J.R. Lupski[2], R.A. Gibbs[1], Baylor-Hopkins Center for Mendelian Genomics.* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD.

The Centers for Mendelian Genomics (CMGs) have sequenced ~19,000 patient samples in collaboration with more than 500 investigators from 36 countries. At its inception in 2011 the Baylor-Hopkins Center for Mendelian Genomics (BHCMG), as one of then three NIH funded CMGs, began its efforts towards: i) novel gene/mutation discovery, ii) elucidating the molecular bases of disease, iii) understanding the genetic susceptibility to disease traits, and iv) determining the genetic/genomic architecture of presumably Mendelian disease. The BHCMG has learned many lessons from its contribution of 6,769 exomes with study subjects spanning 550 clinical phenotypes and presented our findings in more than 100 publications. The Human Genome Sequencing Center at the Baylor College of Medicine has generated 48Tb of whole exome sequencing (WES) data using the HGSC-VCRome capture reagent and now includes 'Spike-in PKv2' to capture more difficult regions and additional gene targets. The HGSC-VCRome capture reagent, along with a multiplex strategy and use of full-length blocking oligos employed for hybridization, has yielded 7.7Gb of data per exome providing a coverage of 96% at 20X or greater. The Spike-in PKv2 reagent is comprised of 3,643 additional unique gene targets derived from GeneTests, OMIM, selected cancer genes and Baylor Miraca Genetics Laboratory positive cases. The addition of this reagent converts >700-800 genes from partially covered to fully covered at ≥ 20X coverage. This high throughput sequencing pipeline, producing 3,000 capture libraries per month, is automated from library construction through annotated VCF generation in part using the 'Mercury' analysis pipeline. Success rates in this program have varied by phenotype and cohort collections ranging from 37% to ~90% of cases in which a likely molecular diagnosis contributing to the disease phenotype is achieved. The BHCMG has established valuable sample acquisition approaches and resources, enhanced sequencing methodology, curated a well-characterized phenotype-rich genetic database enabling genotype/phenotype relationships (i.e. PhenoDB) and encouraged collaborative efforts (i.e. GeneMatcher) to implicate 589 disease genes including 222 novel, 231 known and 136 phenotypic expansion genes. Each discovery has demonstrated the diagnostic capabilities in using whole exome sequencing and has taught lessons in disease mechanisms that continue to drive investigation into those cases that remain unsolved.

2449W

Biallelic mutations in *IARS*, encoding cytosolic isoleucyl-tRNA synthetase, cause growth retardation with prenatal onset, intellectual disability, muscular hypotonia, and infantile hepatopathy. *R. Kopajtich[1,2], K. Murayama[3,4], A.R. Janecke[5,6], T.B. Haack[1,2], M. Breuer[7], A.S. Knisely[8,9], I. Harting[10], T. Ohashi[11], Y. Okazaki[12,13], D. Watanabe[14], Y. Tokuzawa[13], U. Kotzaeridou[7], S. Kölker[7], S. Sauer[7], M. Carl[15], S. Straub[5], A. Entenmann[5], E. Gizewski[16], R.G. Feichtinger[17], J.A. Mayr[17], K. Lackner[9], T.M. Strom[1,2], T. Meitinger[1,2], T. Müller[5], A. Ohtake[18], G.F. Hoffmann[7], H. Prokisch[1,2], C. Staufner[7].* 1) Institute of Human Genetics, Helmholtz Zentrum Muenchen, Munich / Neuherberg, Germany; 2) Institute of Human Genetics, Technical University Muenchen, Munich, Germany; 3) Department of Metabolism, Chiba Children's Hospital Chiba, Japan; 4) Chiba Cancer Center Research Institute, Chiba, Japan; 5) Department of Pediatrics I, Medical University of Innsbruck, Innsbruck, Austria; 6) Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria; 7) Department of General Pediatrics, Division of Neuropediatrics and Metabolic Medicine, University Hospital Heidelberg, Heidelberg, Germany; 8) Institute of Liver Studies, King's College Hospital, London, United Kingdom; 9) Institute of Pathology, Medical University of Graz, Graz, Austria; 10) Department of Neuroradiology, University Hospital Heidelberg, Heidelberg, Germany; 11) Department of Pediatrics, The Jikei University School of Medicine, Tokyo, Japan; 12) Division of Translational Research, Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan; 13) Division of Functional Genomics & Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan; 14) Department of Large Animal Clinics, School of Veterinary Medicine, Kitasato University, Towada, Aomori, Japan; 15) Department of Cell and Molecular Biology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; 16) Department of Neuroradiology, Medical University of Innsbruck, Innsbruck, Austria; 17) Department of Paediatrics, Paracelsus Medical University, SALK Salzburg, Salzburg, Austria; 18) Department of Pediatrics, Faculty of Medicine, Saitama Medical University, Moroyama, Saitama, Japan.

tRNA synthetase deficiencies are a growing group of genetic diseases associated with tissue-specific, mostly neurological phenotypes. In cattle, cytosolic isoleucyl-tRNA synthetase (*IARS*) missense mutations cause hereditary weak calf syndrome. Exome sequencing in three unrelated individuals with severe prenatal-onset growth retardation, intellectual disability, and muscular hypotonia revealed biallelic mutations in *IARS*. Two of the individuals had infantile hepatopathy with fibrosis and steatosis, leading to liver failure in one during the course of infections. Zinc deficiency was present in all affected individuals and supplementation with zinc showed a beneficial effect on growth in one. The functional relevance of the identified variants was confirmed by using a yeast model. All individuals studied are compound heterozygous for one *IARS* loss-of-function allele and one allele with significantly reduced but some residual activity. In situ hybridization and morpholino knockdown of the identified gene homolog in zebrafish were performed to investigate the expression pattern and the gene's role for embryonic development. These experiments suggest an important role for *IARS* in embryogenesis. Downregulation of *IARS* causes high lethality with surviving embryos exhibiting a severe and consistent brain phenotype and a shortening of the body axis reminiscent of the human phenotype. The pathomechanisms of IARS deficiency is not yet fully understood. Given the ubiquitous presence of ARSs in all kingdoms of life and their indispensable function, it is remarkable that mutations in ARSs cause appreciably distinct clinical pictures with tissue-specific phenotypes in humans. In recent years, whole exome sequencing studies have revealed various causes of infantile liver failure with partly specific clinical phenotypes. This study delineates biallelic mutations in *IARS* as underlying a multisystemic disease affecting mainly growth, brain, and liver. For reasons yet to be determined, this disease is associated with zinc deficiency and can be palliated by zinc supplementation.

2450T

Mutations in DCC cause isolated agenesis of the corpus callosum and mirror movements associated with sex-specific penetrance and failure of corticospinal axonal decussation. *A.P.L. Marsh[1,2], D. Heron[3,4,5], T.J. Edwards[6,7], A. Quartier[8], C. Galea[9], C. Nava[3,10], A. Rastetter[10], M. Moutard[11,12,13], S.A. Mandelstam[2,14,15], G. Mcgillivray[16], A. McIlroy[17], A. Méneret[10,19], S.E. Stephenson[1,2], G. Rouleau[20,22], E. Roze[10,19], A. Piton[8,23], M. Bahlo[18,21], T. Billette de Villemeur[5,11,12,24], E.H. Sherr[25], R.J. Leventer[2,26,27], L.J. Richards[6,28], C. Depienne[3,8,10,23], P.J. Lockhart[1,2].* 1) Paediatrics, Murdoch Childrens Research Institute, Parkville, Melbourne, Victoria, Australia; 2) Department of Paediatrics, University of Melbourne, Parkville, Victoria 3052, Australia; 3) AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Génétique, F-75013, Paris, France; 4) Groupe de Recherche Clinique (GRC) "déficience intellectuelle et autisme" UPMC, 75013 Paris, France; 5) Centre de Référence "déficiences intellectuelles de causes rares", 75013 Paris, France; 6) The University of Queensland, Queensland Brain Institute, St Lucia Campus, Brisbane, 4067, Australia; 7) The University of Queensland, School of Medicine, St Lucia Campus, Brisbane, 4067, Australia; 8) Département de Médecine translationnelle et Neurogénétique, IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg, 67400 Illkirch, France; 9) Medicinal Chemistry and Drug Delivery, Disposition and Dynamics (D4), Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia; 10) INSERM, U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France; 11) AP-HP, Hôpital Trousseau, service de neuropédiatrie, 75012 Paris, France; 12) UPMC, GRC ConCer-LD, Sorbonne université, Paris France; 13) Centre de référence "neurogénétique", Paris France; 14) The Florey Institute of Neuroscience and Mental Health, Melbourne, Victoria, Australia; 15) Department of Radiology, University of Melbourne, Royal Children's Hospital, Parkville, Victoria 3052, Australia; 16) Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Parkville, Victoria 3052, Australia; 17) Murdoch Childrens Research Institute, Parkville, Victoria 3052, Australia; 18) Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia; 19) AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Neurologie, F-75013, Paris, France; 20) Department of Neurology and Neurosurgery, McGill University Health Center, Montreal, Quebec, H3A 2B4 Canada; 21) Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville 3052 VIC, Australia; 22) Montreal Neurological Institute and Hospital, McGill University, Montréal, Quebec H3A 2B4, Canada; 23) Laboratoires de génétique, Institut de génétique médicale d'Alsace, Hôpitaux Universitaires de Strasbourg, 67 000 Strasbourg, France; 24) INSERM U1141, 75019 Paris, France; 25) Department of Neurology, UCSF Benioff Children's Hospital, San Francisco, California, USA; 26) Neuroscience Research Group, Murdoch Childrens Research Institute, Parkville, Victoria 3052, Australia; 27) Department of Neurology, University of Melbourne, Royal Children's Hospital, Parkville, Victoria 3052, Australia; 28) The University of Queensland, School of Biomedical Sciences, St Lucia Campus, Brisbane, 4067, Australia.

Background: Deleted in Colorectal Cancer (DCC) is a conserved Netrin-1 binding axon-guidance receptor critical for corpus callosum development. Mice with homozygous *Dcc* mutations display agenesis of the corpus callosum (ACC), as well as corticospinal (CS) tract defects. In humans, *DCC* haploinsufficiency has been associated with mirror movements (MM, OMIM#157600) but not ACC (OMIM#217990). **Methods:** Genetic studies, including whole exome sequencing, were used to investigate the cause of ACC in 4 multigenerational families and 70 sporadic cases. MRI and probabilistic constrained spherical deconvolution (CSD) MRI tractography were utilised to assess brain structure and CS tract wiring. *DCC* expression in human neural stem cells (hNSC) was measured using RNAseq and RT-qPCR following testosterone treatment. **Findings:** Genetic analyses identified *DCC* mutations in 9 unrelated families with isolated ACC with/without MM. All individuals with DCC-related ACC had an intellectual quotient within normal/borderline range. CSD MRI tractography identified a failure of CS axonal decussation in all affected individuals (n=7) from 2 families. All 7 individuals displayed MM, with concomitant ACC in 5. Sex bias in phenotype expression was observed, with *DCC* haploinsufficiency in males significantly correlated with MM while females only presented with ACC. This sex bias was not observed in individuals with missense mutations. Males and females with missense mutations located within the DCC/Netrin-1 binding interface presented with ACC and MM; individuals with missense mutations located outside this region presented with ACC only. Testosterone-treated hNSC showed significant upregulation of *DCC*. **Interpretation:** We show that heterozygous missense and loss-of-function *DCC* mutations

cause isolated ACC, with/without MM. Contrary to the multitude of genes known to cause syndromic ACC, *DCC* mutations are associated with isolated ACC and a relatively mild phenotype with a favourable cognitive outcome. We provide new evidence that MM in *DCC*-mutant individuals results from failure of CS axonal decussation. Our data indicates a distinct molecular mechanism leading to disease in missense mutant compared to haploinsufficient individuals. We provide novel evidence that sex bias in phenotype expression may be the result of increased *DCC* expression caused by higher testosterone levels in males. This is the first report linking mutation of a gene to isolated ACC in multiple, unrelated individuals.

2451F

Clinical and genetic features of 94 Japanese Charcot–Marie–Tooth disease patients with *MFN2* **mutations.** *A. Masahiro, A. Yoshimura, A. Hashiguchi, Y. Hiramatsu, J. Yuan, Y. Higushi, Y. Okamoto, H. Takashima.* neurology, Kagosima university, Kagoshima, Kagoshima, Japan.

   Background and Objective Charcot-Marie-Tooth disease (CMT) is one of the most common inherited peripheral neuropathies(IPN) for which more than 70 causative genes have been reported. The clinical features of CMT are significantly heterogeneous even among patients with mutations in the same gene. In this study, we aimed to describe the clinical spectrum of CMT patients harboring *MFN2* mutations in Japan. **Methods** We analyzed 1513 unrelated cases with clinically suspected IPN who were referred by neurological and neuropediatric departments throughout Japan. All these cases had been proved negative for *PMP22* duplication/deletion. We conducted mutation screening using DNA microarray, targeted resequencing, and/or whole-exome sequencing successively. Candidate variants detected using various platforms were validated by Sanger sequencing, and if possible, segregation analysis was performed. **Results** In 94 suspected CMT2A patients, we identified 49 heterozygous and two homozygous different mutations in *MFN2*. Herein, 20 novel mutations were detected in 27 cases, whereas 31 mutations in 67 cases had been previously described. Segregation analysis was conducted in seven pedigrees with novel mutations. The primary symptom might emerge at 10–50 years of age, with a mean onset age of 12.03 ± 14.3 years in 75 cases with a specific record. The inheritance pattern was determined in 85 cases comprising 43 autosomal dominant, three autosomal recessive, and 39 sporadic cases. The mean onset age in the sporadic cases was lower than that in the autosomal dominant cases (7.58 ± 10.24 vs 14.95 ± 16.06). We reviewed all electrophysiological analyses of 81 cases; five cases with demyelinating type (motor conduction velocity < 38 m/s) and 76 cases with axonal type (motor conduction velocity > 38 m/s) were verified. Interestingly, we also detected mutations in two genes, *MFN2* and *PMP22*, from one patient. In addition, one cases characterized by late-onset axonal CMT, which is rare in *MFN2*-related CMT patients, were found. Furthermore, a patient harboring homozygous R280H mutation presented with a more severe phenotype than that of the heterozygous cases. **Conclusion** *MFN2* was identified as the most common causative gene of CMT2, and the clinical features of patients significantly varied. Notable intragenic mutation diversity could be observed in our Japanese cohort. Regardless of the onset age and inheritance pattern, it is necessary to consider the possibility of *MFN2* mutation in axonal CMT.

2452W

A recurrent mutation in *MYH14* causes autosomal dominant Charcot-Marie-Tooth disease with suspected mitochondrial pathology. *C. Smith[1], T. Shutt[1,2], G. Pfeffer[1,3], R.E. Lamont[1,4], F. Bernier[1,4], J. Parboosingh[1,4], A.M. Innes[1,4], Care4Rare Canada Consortium.* 1) Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada; 2) Department of Biochemistry & Molecular Biology, University of Calgary, Calgary, Alberta, Canada; 3) Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada; 4) Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada.

Mutations in myosin heavy chain 14 (*MYH14*) are an established cause of autosomal dominant non-syndromic hearing loss (DFNA4A, OMIM: 600652). However, there are two literature reports of a NM_001077186.1 (MYH14): c.2822G>T (p.Arg941Leu) mutation being associated with Charcot-Marie-Tooth Disease (CMT) in addition to hearing loss (PNMHH, OMIM: 608568). Using exome sequencing we have identified a three generation family with CMT and hearing loss that also has the p.Arg941Leu mutation. *MYH14* encodes a non-muscle myosin that has been ascribed roles in cytokinesis, karyokinesis and neuritogenesis. In order to better understand the pathogenesis of the disease, we attempted to characterize the role of MYH14 protein in cell culture. Expression of a GFP tagged protein showed localization to mitochondria. Moreover, MYH14 was observed at sites of mitochondrial fission, where it co-localized with the well studied fission factor DRP1. Overexpression of MYH14 caused cells to take on a subtly more fragmented mitochondrial morphology, while cells overexpressing p.Arg941Leu mutant MYH14 were no different from cells containing an empty vector. Based on these results, we suspect that MYH14 has a role in mitochondrial fission and that dysfunction in this role contributes to the pathogenesis of CMT and hearing loss. This is an attractive hypothesis because disruption of other genes involved in mitochondrial dynamics also cause CMT (*MFN2, GDAP1*). Our experiments in cell culture further suggest a loss-of-function mechanism. Finally, the identification of a third family confirms the association between the p.Arg941Leu mutation and CMT with hearing loss.

2453T

Genotype-phenotype study of patients with VCP valosin-containing protein multisystem proteinopathy. *E. Al-Obeidi[1], A. Surampalli[1], S. Al-Tahan[1], N. Goyal[2], C. Smith[3], M. Omizo[4], D. Waggoner[5], A. Hermann[6], T. Mozaffar[2], V. Kimonis[1].* 1) Div. of Genetics and Genomic Medic, Dept. Pediatrics, University of California- Irvine Irvine, CA; 2) Div. of Neuromuscular Neurology, Dept. of Neurology, University of California- Irvine, Orange, CA; 3) Dept. of Neurology, University of Kentucky Medical School, Lexington, KY; 4) Oregon Osteoporosis Center, Bend, OR; 5) Div. Human Gen, Dept. Pediatrics, University of Chicago, Chicago IL; 6) Bereich Neurodegenerative Erkrankungen, Klinik und Poliklinik für Neurologie, Dresden, Germany.

Missense mutations in valosin-containing protein (VCP) is associated with a unique limb-girdle/ inclusion-body myopathy (IBM) associated with Paget's disease of bone (PDB), frontotemporal dementia (FTD), and amyotrophic lateral sclerosis (ALS). VCP is an ATPase found in all cells and is a key player in protein degradation and autophagy. VCP disease is a progressive, fatal disorder that has an autosomal dominant adult onset progressive disease. To date, 44 disease-causing missense mutations in the *VCP* gene have been reported worldwide. We report the clinical, natural history and molecular findings in the largest cohort in the world. The goal of this study was to examine genotype-phenotype data from 231 individuals (118 males, 113 females) from 36 families carrying 15 different VCP mutations (three of which are novel). We analyzed whether the different mutations were related to age of onset and severity of IBM, PDB, FTD and ALS, and measured the prevalence of these conditions and other comorbidities. Myopathy was present in 89% of the patients, beginning at an average age of 43 years. PDB was diagnosed in 43% of patients with an average age of onset of 41 years. FTD presented in 29% of the individuals at an average age of 56 years. Approximately 9% of patients with VCP mutations had an ALS phenotype, 4% had been diagnosed with Parkinson's disease, and 2% had been diagnosed with Alzheimer's disease. Because of the wide inter and intra-familial variation establishing a genotype-phenotype correlation was difficult. We however did find that the R159C had a later age of onset of myopathy and the absence of PDB. By understanding the typical clinical presentation and age of onset of symptoms, clinicians will be better able to identify and diagnose VCP-related diseases and thus proactively manage associated features more effectively.

2454F

Identification of a post-translational modification with ribitol-phosphate and its defect in muscular dystrophy. *T. Toda[1], M. Kanagawa[1], K. Kobayashi[1], M. Tajiri[2], H. Manya[3], A. Kuga[1], Y. Yamaguchi[4], Y. Wada[2], T. Endo[3].* 1) Division of Neurology / Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Japan; 2) Department of Molecular Medicine, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 3) Molecular Glycobiology, Research Team for Mechanism of Aging, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan; 4) Structural Glycobiology Team, Systems Glycobiology Research Group, RIKEN-Max Planck Joint Research Center for Systems Chemical Biology, RIKEN Global Research Cluster Wako, Japan.

Glycosylation is an essential post-translational modification that underlies many biological processes and diseases. α-dystroglycan (α-DG) is a receptor for matrix and synaptic proteins that causes muscular dystrophy and lissencephaly upon its abnormal glycosylation (α-dystroglycanopathies). Here we identify the glycan unit ribitol 5-phosphate (Rbo5P), a phosphoric ester of pentose alcohol, in α-DG. Rbo5P forms a tandem repeat and functions as a scaffold for the formation of the ligand-binding moiety. We show that enzyme activities of three major α-dystroglycanopathy-causing proteins are involved in the synthesis of tandem Rbo5P. Isoprenoid synthase domain-containing (ISPD) is cytidine diphosphate ribitol (CDP-Rbo) synthase. Fukutin and fukutin-related protein are sequentially acting Rbo5P transferases that use CDP-Rbo. Consequently, Rbo5P glycosylation is defective in α-dystroglycanopathy models. Supplementation of CDP-Rbo to ISPD-deficient cells restored α-DG glycosylation. These findings establish the molecular basis of mammalian Rbo5P glycosylation and provide insight into pathogenesis and therapeutic strategies in α-DG-associated diseases.

2455W

A novel mutation p.Glu632Lys in *SH3TC2* gene causing Charcot-Marie-Tooth disease type 4C in a Chinese family. *M. Gu[1], Z. Yu[1], J. Zhang[1], Y. Xu[1], W. Shen[2].* 1) Dept. of Medical Genetics, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 2) Dept. of Laboratory Medicine, Renji Hospital, Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China.

Background The Charcot-Marie-Tooth disease (CMT) is one of the most common human inherited peripheral neuropathies. Major pattern of inheritance is autosomal dominant, with less often occurrence autosomal recessive and X-linked dominant/recessive inheritance. CMT is generally divided into three forms: demyelinating forms (CMT1), axonal forms (CMT2) and intermediate forms. The autosomal recessive form (AR-CMT1 or CMT4) is accompanied by progressive distal muscle weakness and atrophy of the limbs, pes cavus and claw-like hands, both often observed in CMT. In addition, CMT4 is also characterized by early onset, rapid progression, and varying degrees of sensory loss and spinal deformities. Eleven subtypes of CMT4 have been identified: from CMT4A to CMT4J. Some of these subtypes were clear in pathogenic mechanisms, some had founder mutation, but some still had limited clinical description and mutation analysis. **Objective** To identify potential mutation in a Chinese family affected with CMT4C and analyse the SH3TC2 protein structure. **Methods** Clinical data of a three-generation family from the Zhejiang Province of China was collected, and genomic DNA was extracted from peripheral blood samples of the family members. Seventy-two candidate genes of the proband were captured and sequenced by targeted next-generation sequencing, and the results were confirmed by Sanger sequencing. The protein structure was predicted with PyMOL-1 software. **Results** The proband presented with certain typical CMT4 phenotypes, including symmetrical muscle wasting and a predominating weakness of the distal parts of the lower limbs, pes cavus and claw-like hands. Furthermore, a novel homozygous missense mutation c.1894G>A (p.Glu632Lys) was identified in the exon 11 of *SH3TC2* gene in the index case. The heterozygous mutation (p.Glu632Lys) was also detected in the index case's father, mother and daughter, but not in the healthy family members and 300 normal controls. Retrieval of the NCBI, HGMD and 1000 genome databases has verified the c.1894G>A to be as a novel mutation. Computer simulation has suggested that the mutation has altered the 3D structure of the SH3TC2 protein. **Conclusions** The index case was identified as CMT4C, for which the underlying gene was *SH3TC2* gene. Our finding has expanded the spectrum of *SH3TC2* mutation in associated with CMT4C. This work is partially supported by National Natural Science Foundation of China (Proj. No. 30470951, 31071107).

2456T

Recurrent mutations in *HEXB* gene in using targeted sequencing. *S. Mikaeeli[1], R. Shervin-Badv[2,3], N. Mahdieh[4], S. Mikaeeli[5], B. Rabbani[2,4].* 1) Islamic Azad University, North Tehran Branch, Tehran, Iran, tehran, Alborz, Iran; 2) Children's Hospital Center, Tehran University of Medical Center; 3) Growth and Development Research, Tehran University of Medical Center; 4) Genetic Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran; 5) Islamic Azad University of Medical Branches, Tehran, Iran.

Sandhoff disease [MIM 268800] or GM2-gangliosidosis type 2 is a rare lysosomal storage disorder with autosomal recessive inheritance. There are different phenotypes for Ssandhoff disease based on biochemical varieties and age of onset including; infantile (acute form; <0.1% activity), late infantile and juvenile (subacute form; 0.5% activity) and adult (chronic; 2-4% activity). There are three types of lysosomal β-hexoaminidase enzyme which deficiency of hexoaminidase A (αβ) and B (ββ) activity is seen in Sandhoff disease. The *HEX* gene encoding β-subunits of these enzymes located on 5q13 chromosome and consists of 14 exons. Mutations in this gene lead to GM2 accumulation in neuronal lysosomes and causes fatal neurodegeneration and apoptosis of neurons. Here we present a patient with a mutation in *HEXB* gene. Genotype was determined by targeted-next generation sequencing (NGS) [P117] for 4 candidate genes (*HEXB, HEXA, GM2A, GLB1*). The determined mutation was confirmed by PCR-RFLP assay. Mutational analysis showed a homozygous mutation at position c.1602C>A of *HEXB* gene in exon 13 that has pathogenesis effects leading to early termination of amino acid sequence (Cys534Ter). RFLP-PCR of the consanguineous parents showed a heterozygous variant at position c.1602C>A. This mutation leads to elimination of last 23 amino acids from the C terminal loop. Also, a homozygous mutation was found at position p.Lys121Arg (c.362A>G) for another patient. The panel based analysis could be important for molecular diagnosis and confirmation of heterogonous phenotype in medical genetic.

2457F

Novel homozygous *SPG11* mutation in the molecular diagnosis of hereditary spastic paraplegia. *B. Burnyte, L. Ambrozaityte, V. Kucinskas, A. Utkus.* Dept. of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania.

Hereditary spastic paraplegia (HSP) is a heterogeneous group of neurodegenerative disorders. *SPG11* mutations are the most common cause of autosomal recessive HSP. *SPG11*, encoding spatacsin, is expressed ubiquitously in the nervous system, but most prominently in the cerebellum, cerebral cortex, hippocampus and pineal gland. The specific role of peptide remains unknown, though it seems to be essential to the survival of neurons. We report the case of a 26-year old male with early onset spasticity, axonal neuropathy, ataxia, dysarthria, narrowed field of vision, persistent acne and cognitive impairment. Next generation sequencing of the 98 genes related to neurodegenerative disorders was negative. Whole exome sequencing was performed to determine the genetic cause of our patient's phenotype. A novel homozygous mutation c.2431C>T; p.Q811* has been identified in the *SPG11* (rs756134516) gene consistent with the diagnosis of HSP type 11 (SPG11). In silico analysis predicts this variant to be disease causing, as the majority of missense/nonsense mutation group of *SPG11* created termination codon. Targeted whole exome sequencing data analysis based on a number of genes of selected neurodegenerative disorders allowed to exclude the presence of pathogenic mutations in other possible genes in the proband. Clinical resemblance between our patient and reported SPG11 cases suggests that the detected homozygous mutation is the phenotype causal mutation in our patient. Further studies are necessary to clarify the pathogenic mechanism of the detected mutation at the protein level as well as the role of SPG11 RNA transcript under both physiological and pathological conditions.

2458W

Enrichment of unamplified DNA and long-read SMRT Sequencing to unlock repeat expansion disorders. *J.M. Ekholm, Y. Tsai, D. Greenberg, T.A Clark.* Pacific Biosciences, Menlo Park, CA.

Nucleotide repeat expansions are a major cause of neurological and neuromuscular disease in humans, however, the nature of these genomic regions makes characterizing them extremely challenging. Accurate DNA sequencing of repeat expansions using short-read sequencing technologies is difficult, as short-read technologies often cannot read through regions of low sequence complexity. Additionally, these short reads do not span the entire region of interest and therefore sequence assembly is required. Lastly, most target enrichment methods are reliant upon amplification which adds the additional caveat of PCR bias. We have developed a novel, amplification-free enrichment technique that employs the CRISPR/Cas9 system for specific targeting of individual human genes. This method, in conjunction with PacBio's long reads and uniform coverage, enables sequencing of complex genomic regions that cannot be investigated with other technologies. Using human genomic DNA samples and this strategy, we have successfully targeted the loci of Huntington's Disease (HTT; CAG repeat), Fragile X (FMR1; CGG repeat), ALS (C9orf72; GGGGCC repeat), and Spinocerebellar ataxia type 10 (SCA10; variable ATTCT repeat) for examination. With this data, we demonstrate the ability to isolate hundreds of individual on-target molecules in a single SMRT Cell and accurately sequence through long repeat stretches, regardless of the extreme GC-content. The method is compatible with multiplexing of multiple targets and multiple samples in a single reaction. This technique also captures native DNA molecules for sequencing, allowing for the possibility of direct detection and characterization of epigenetic signatures.

2459T

Large-scale exome sequencing in patients with hereditary spastic paraplegia (HSP) reveals molecular epidemiology of AR-HSP and related diseases including AR-HSP with mutations in *CAPN1*. *H. Ishiura[1], M. Tanaka[1], K. Koh[2], H. Shimazaki[3], Y. Ichinose[2], Y. Takahashi[4], J. Mitsui[1], J. Goto[5], T. Matsukawa[1], Y. Ichikawa[6], J. Yoshimura[7], K. Doi[7], S. Morishita[7], R. Saika[8], T. Kondo[9,10,11,12], S. Matsumoto[9], Y. Ohnaka[13,14], K. Hagiwara[15], J. Kira[16], M. Kuwahara[17], Y. Nakamura[18], H. Sasaki[19], G. Sobue[20], S. Miyawaki[21], N. Saito[21], Y. Takiyama[2], S. Tsuji[1], JASPAC (Japan Spastic Paraplegia Research Consortium).* 1) Department of Neurology, The University of Tokyo, Tokyo, Japan; 2) Department of Neurology, Yamanashi University; 3) Department of Neurology, Jichi Medical University, Tochigi, Japan; 4) Department of Neurology, National Center of Neurology and Psychiatry; 5) Department of Neurology, International University of Health and Welfare Mita Hospital; 6) Department of Neurology, Kyorin University; 7) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo; 8) Department of Neurology, Shimane University Faculty of Medicine; 9) Department of Neurology, Tazuke Kofukai Foundation, Medical Research Institute, Kitano Hospital; 10) Department of Clinical Network and Collaborative Medicine, Kyoto University Hospital; 11) Department of Neurology, Kyoto University Graduate School of Medicine; 12) Department of Neurology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine; 13) Department of Neurology,;School of Medicine,;Showa University; 14) Department of Neurology, Showa University Fujigaoka Hospital; 15) Department of Clinical Neurophysiology, Neurological Institute, Faculty of Medicine, Graduate School of Medical Sciences, Kyushu University; 16) Department of Neurology, Neurological Institute, Faculty of Medicine, Graduate School of Medical Sciences, Kyushu University; 17) Department of Neurology, Kindai University, Faculty of Medicine; 18) Department of Neurology, Sakai Hospital, Kindai University, Faculty of Medicine; 19) Department of Neurology, Hokkaido University; 20) Department of Neurology, Nagoya University; 21) Department of Neurosurgery, The University of Tokyo.

[Background] Hereditary spastic paraplegia (HSP) is a neurodegenerative disorder characterized by spasticity and pyramidal weakness of the lower limbs. The causative genes for autosomal recessive HSP (ARHSP) are more heterogeneous than those of autosomal dominant HSP and about 60% of ARHSP patients remain to be elucidated. In the majority of the families, affected individuals are singletons, making it difficult to efficiently narrow the candidate region with linkage analysis. We tried to identify novel causative gene for ARHSP based on exome sequencing of index patients from individual families. [Methods] From 253 patients with HSP likely having autosomal recessive inheritance, 136 patients of undiagnosed HSP were enrolled. Non-synonymous, frameshift, and splicing variants whose minor allele frequencies were <0.2% were selected from exome data. We then searched for genes carrying biallelic rare variants that were shared by multiple families and were not shared by control subjects. [Results] We identified biallelic mutations in *CAPN1* in four patients. We found two homozygous nonsense mutations (c.461G>A, p.W154*; c.1153C>T, p.R385*) and one homozygous missense mutations (c.658G>A, p.G220R) in two families. Homozygous nonsense mutations strongly support loss-of-function mutations and the missense mutation is located in the conserved catalytic domain. The haplotypes surrounding the missense mutation were identical in the two families. The clinical presentations of these four patients are characterized by ataxia or dysarthria in addition to spastic paraparesis. Taken together with a missense mutation was found in dogs presenting with ataxia and spasticity, *CAPN1* is considered to be causative gene for complicated ARHSP. In addition, we identified nine other possible candidate genes including *AP5B1*, in which two index patients share biallelic mutations. In the present study, the causative mutations were found in 121 of the 253 patients (47.8%), including five SPG28 families showing retinitis pigmentosa as a novel phenotype in addition to spastic paraparesis. [Discussion] Utilizing exome data of singletons, we attempted to identify causative genes for ARHSP, focusing on genes with biallelic variants shared by multiple families. The fact that we identified *CAPN1* as a novel causative gene for ARHSP indicated that our study paradigm is efficient for identifying causative genes even though only singletons from individual families are available for the analysis.

2460F

Novel *GRN* mutation c.687T>A, p.(Tyr229*): Clinical and neuropathological features. *L. Kuuluvainen[1,2], M. Pöyhönen[1,2], P. Pasanen[3], M. Siitonen[3,2], J. Rummukainen[4], P.J. Tienari[5,6], A. Paetau[7], L. Myllykangas[7].* 1) Department of Clinical Genetics, Helsinki University Central Hospital , Helsinki, Finland; 2) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 3) Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland; 4) Department of Pathology, Kuopio University Hospital, Kuopio, Finland; 5) Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 6) Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; 7) Department of Pathology, University of Helsinki, Helsinki, Finland.

Frontotemporal lobar degeneration (FTLD) is the third most common cause of dementia and accounts for about 5-10% of all dementia patients. Mutations in the granulin (*GRN*) gene represent about 5-10% of frontotemporal lobar degeneration (FTLD). We present the clinical and neuropathological features of a family with a novel *GRN* c.687T>A, p.(Tyr229*) mutation. The mutation was found in the proband of the family using whole exome sequencing. The mutation caused a severe FTLD neuropathological phenotype with TDP43 inclusions in the proband. The nephew of the proband was found to have similar but milder FTLD pathology and several other family members had had early-onset dementia. Gene expression studies using commercial TaqMan gene expression assay showed decreased *GRN* gene expression in mutation carriers' blood samples. The novel *GRN* mutation results in haploinsufficiency of *GRN* and a severe neuropathologic FTLD phenotype. .

2461W

Dominant mutations in the *ALDH18A1* gene cause hereditary spastic paraplegia SPG9. *E. Panza[1,2], J.M. Escamilla-Honrubia[3], C. Marco-Marín[3,4], N. Gougeard[3,4], G. De Michele[5], V. Brescia Morra[5], R. Liguori[6,7], L. Salviati[8,9], M.A. Donati[10], R. Cusano[11,12], T. Pippucci[2], R. Ravazzolo[13,14], A.H. Nemeth[15,16], S. Smithson[17], S. Davies[18], J.A. Hurst[19], D. Bordo[20], V. Rubio[3,4], M. Seri[2].* 1) Human Genetics, University of Utah, SLC, UT; 2) Department of Medical and Surgical Science, University of Bologna, Bologna 40138, Italy; 3) Instituto de Biomedicina de Valencia of the CSIC, Valencia 46010, Spain; 4) Group 739, Centro para Investigación Biomédica en Red sobre Enfermedades Raras CIBERER-ISCIII, Valencia, Spain; 5) Department of Neurosciences and Reproductive and Odontostomatologic Sciences, University Federico II, Naples 80138, Italy; 6) IRCCS Istituto delle Scienze Neurologiche di Bologna 40139, Italy,; 7) Department of Biomedical and NeuroMotor Sciences, University of Bologna 40139, Italy,; 8) Clinical Genetics Unit, Department of Woman and Child Health, University of Padova 35128, Italy; 9) IRP Città della Speranza, Padova 35129, Italy; 10) A. Meyer Institute, Firenze 50139, Italy; 11) Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Cagliari 09042, Italy; 12) CRS4, Parco Tecnologico della Sardegna, Cagliari 09010, Italy; 13) Medical Genetics Unit, G. Gaslini Institute, Genova 16147, Italy; 14) University of Genova, DI-NOGMI Department, Genova 16147, Italy; 15) Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, UK; 16) Department of Clinical Genetics, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford, UK; 17) Department of Clinical Genetics, St Michael's Hospital Bristol BS8 4LL, UK; 18) Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK; 19) Department of Clinical Genetics, Great Ormond Street Hospital, London, UK; 20) IRCCS AOU San Martino – IST. Istituto Nazionale per la Ricerca sul Cancro, Genova 16132, Italy.

SPG9 (MIM 601162) is a form of autosomal dominant hereditary spastic paraplegia, mapped in two families of Italian and English origins. This complex form of spastic paraplegia is characterized by the presence of early cataracts, persistent vomiting and upper motor neuron signs. In the SPG9 critical region the *ALDH18A1* gene was present and previously associated with recessive disorders (MIM 219150), with some common traits among these forms. *ALDH18A1* encodes for Δ^1-pyrroline-5-carboxylate-synthetase (P5CS) a bifunctional enzyme (glutamate-5-kinase and reductase domains G5K and GPR), catalyzing the first step of ornithine and proline biosynthesis. In each family we identified one *ALDH18A1* mutation in heterozygosis, cosegregating with the disease. The mutations identified, c.727G>C (p.Val243Leu) or c.755G>A (p.Arg252Gln), are in the exon 7 of the *ALDH18A1* gene, and are positioned nine amino acids apart in the sequence of the glutamate-5-kinase (G5K) component of P5CS. These mutations did not prevent synthesis nor cause cellular mislocalization of the P5CS protein, as judged by fluorescence microscopy of the transiently expressed GFP-tagged P5CS in CHO cells, or by P5CS immunofluorescence of patient fibroblasts. Low fasting plasma ornithine, citrulline and hydroxyproline in SPG9 patients suggested a loss-of-function of P5CS. Indeed, when these mutations have been introduced into recombinant human P5CS, they essentially inactivated the enzyme activity and promoted dissociation of the P5CS hexamer. To reconcile the loss-of-function effect with the dominant inheritance, we propose a dominant negative effect in which even the presence of a single mutant subunit introduces architectural defects in the P5CS oligomer that inactivate the entire oligomer of bifunctional subunits. In summary, dominant mutations in *ALDH18A1* cause SPG9. The existence of another very rare condition associated to *ALDH18A1*, presenting with a similar phenotype but having recessive inheritance indicates that *ALDH18A1* mutations can be associated with recessive or dominant inheritance in an allele-specific way. An accurate comparison of the phenotypes and mutations of the *ALDH18A1*-related disorders is necessary to better understand the pathogenetic mechanism of these group of conditions. *Grants: Telethon Fondation Onlus, GGP06209 & GGP10121 (AMS,Gubbio) & It.Min.Health; (VR) PrometeoII/2014/029 (Valencian Government) and BFU2014-58229-P (Spanish Government).*

2462T

First report of a contraction of the *C9ORF72*-**hexanucleotide repeat expansion.** *P. Weydt[1], T. Mussotter[2], V. Müller[1], A. Knehr[1], J. Weishaupt[1], A. Ludolph[1], C. Kubisch[3], A. Volk[3].* 1) Neurology, Univ Ulm, Ulm, Germany; 2) Human Genetics, Univ Ulm, Germany; 3) Human Genetics, Univ Medical Center Hamburg-Eppendorf, Germany.

The GGGGCC-repeat expansion in *C9orf72* is the most frequent mutation in patients with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) and one of the most common genetic causes of a neurodegenerative disease in general. The exact cut-off between normal, intermediate and pathologic alleles is currently unclear. Also, our understanding of possible genotype-phenotype correlations is still evolving. Studies applying Southern-blot (SB) analyses reported smaller repeat expansion in patients with an earlier age of onset. However another study reported a healthy elderly man carrying an intermediate repeat who transmitted larger expanded alleles to several affected children. The latter would be in line with the phenomenon of anticipation as observed in the majority of currently known repeat diseases. Here, we present a family in which a 70 year old male was diagnosed with ALS according to the revised El Escorial criteria. As his father had died of ALS at the age of 57 a familial form was suspected. Indeed, genetic testing confirmed the diagnosis and revealed a non-expanded allele with 5 repeats and expanded allele with 1800-2400 repeats as determined by PCR-based techniques and SB analysis. Subsequently, one of his children opted for predictive genetic testing to learn about his personal risk for *C9orf72*-associated diseases. The neurological exam in this 38 year old man was unremarkable. Genetic testing showed one allele with 2 repeats and one carrying 100-120 repeats. As an inheritance of the expanded allele by the mother was excluded and paternity was confirmed by genetic testing with polymorphic markers the most plausible explanation for the presence of the 100-120 repeat allele is a contraction of a larger expanded allele or somatic mosaicism in the father. Both scenarios highlight our currently limited knowledge on the *C9orf72* repeat and the resulting challenges for genetic counselling. If the father has a somatic mosaicism at least a larger expansion could be detected in blood cells and somatic instability with larger expansions within the brain have been described in other patients before. Also, a repeat contraction seems difficult to interpret in this context as the exact cut-off for pathologic alleles remains to be determined and smaller repeat expansions had been associated with an earlier age of onset.

2463F

Genetic analysis of Parkinsonism among Norwegian patients. *E. Gustavsson[1,2,3], J. Trinh[1], I. Guella[1], M. McKenzie[1], J. Aasly[2,3], M. Farrer[1].* 1) Djavad Mowafaghian Centre for Brain Health, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Neurology, St. Olav's Hospital, Trondheim, Norway.

Parkinson's disease (PD) is an age-associated disease and the second most common neurodegenerative disease. It is a heterogeneous clinical syndrome characterized by a range of motor symptoms. Definitive and early diagnosis is imperative for optimal treatment, clinical trials and biomarker research. Approximately 14% of patients report a positive family history of PD. Many causal genes have been linked to parkinsonism including mutations in *SNCA*, *LRRK2, VPS35, DNAJC13, ATP6AP2* and *CHCHD2* in late-onset dominant disease, and in *PRKN and PINK1* in recessive early-onset parkinsonism. Mutations in *DJ-1, ATP13A2, PLA2G6, FBXO7, DNAJC6* and *SYNJ1* also result in recessive early/juvenile-onset parkinsonism albeit with atypical presentations. We have studied 105 unrelated patients with early-onset parkinsonism and the probands of 11 pedigrees with multi-incident late-onset dominant PD. All subjects are of Norwegian origin. Genome-wide SNP genotyping has been performed using the Illumina Multi-Ethnic Genome Array (MEGA) and whole-exome sequencing has been completed using the Ion AmpliSeq™ Ion Proton system. Initially we describe variability in genes previously shown to cause PD/parkinsonism for all patients with early-onset disease, familial affected probands and 50 control subjects. Disease in none of the 11 pedigrees is explained by previously known genes. However, 16 patients with early-onset disease were seen carrying pathogenic variants in genes previously implicated in parkinsonism. Subsequently, CNV analysis from MEGA has identified a significant burden of duplications among patients, including a subset highlighting genes for dystonia, a prevalent phenotype among patients with parkinsonism. Mapping of novel variability within shared runs of homozygosity, in patients with early-onset parkinsonism and segregation analysis within multi-incident pedigrees has nominated novel genetic variability as causal. These findings and their further evaluation within additional case-control series will be reported.

2464W

Analysis of mitochondria-related gene from clinically suspected Charcot-Marie-Tooth patients. *Y. Hiramatsu[1], Y. Okamoto[1], A. Yoshimura[1], M. Ando[1], J. Yuan[1], Y. Higuchi[1], A. Hashiguchi[1], S. Tsuji[2], H. Takashima[1].* 1) Department of Neurology and Geriatrics, Kagoshima University Graduate School of Medicine and Dental Sciences, Kagoshima city, Japan; 2) Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.

Background Charcot-Marie-Tooth (CMT) disease is a clinically and genetically heterogeneous disorder that causes inherited distal symmetric peripheral neuropathy. In spite of more than 80 causative genes linked to CMT till now, the unexpected low detection rate suggests more undiscovered genes that could result in CMT phenotype. Mitochondria-related gene like *MFN2* and *GDAP1* are well known CMT disease-causing genes, and patients with mitochondrial disorders could develop peripheral neuropathy. Methods We enrolled 768 clinically suspected CMT patients. We first screened the CMT disease-causing genes using DNA microarray (Affymetrix; 27 genes) or targeted resequencing (Illumina MiSeq; 40 genes). All of the mutation-negative patients in these pre-screening underwent a whole-exome sequencing (WES). On the basis of their inheritance pattern, 247 CMT patients grouped as autosomal recessive (AR) or sporadic were selected for a further analysis with 167 mitochondria-related genes. Results In the pre-screening state, the most common mitochondria-related gene was *MFN2*, in which pathogenic mutations were detected from 47 patients. *GDAP1* mutation, as the second common reason, was identified in 6 patients. All of these mutations were previously described or validated by segregation analysis. Additionally, using WES, we detected seven novel likely pathogenic variants from 167 mitochondria-related genes. These patients presented with diverse neurological symptoms, comprising encephalopathy, mental retardation, optic atrophy, and hearing loss. Conclusions Since mitochondrial dysfunction frequently leads to polyneuropathy, it is essential to check the mitochondria-related genes in patients with a CMT phenotype, particularly for the patients accompanied with additional neurological findings like encephalopathy or optic atrophy.

2465T

Loss of glycine transporter 1 (GLYT1) causes congenital abnormal muscle tone, arthrogryposis, respiratory failure and mildly elevated CSF glycine. *A. Kurolap[1,2], T. Hershkovitz[1], A. Mory[1], T. Paperna[1], G. Tal[3], Y. Nijem[4], E. Sella[4], N. Ekhilevitch[1,2], H. Mandel[2,3], H.N. Baris[1,2].* 1) The Genetics Institute, Rambam Health Care Campus, Haifa, Israel; 2) Rappaport School of Medicine, Technion – Israel Institute of Technology, Haifa, Israel; 3) Metabolic Unit, Rambam Health Care Campus, Haifa, Israel; 4) Pediatric and Neonatal Unit, Nazareth Hospital EMMS, Nazareth, Israel.

Glycine is a major neurotransmitter in inhibitory synapses, activating the glycine receptor, as well as a co-agonist at the glutamatergic NMDA receptors in excitatory synapses. Two high-affinity substrate-selective transporters, glycine transporter-1 and glycine transporter-2 (GLYT1 and GLYT2), play critical roles in regulating intra and extracellular glycine concentrations within the CNS. Glycine dysregulation has been linked to several conditions, including hyperekplexia, and nonketotic hyperglycinemia (NKH). We report four patients originating from two unrelated consanguineous Muslim-Arab kindred from Israel. All patients presented at birth with facial dysmorphism, encephalopathy, severe arthrogryposis with extreme hyperextension of the knee joints, hip dislocations, respiratory failure necessitating mechanical ventilation, and hypotonia progressing to hypertonicity with clonus. Three babies died between ages two days and eight months. One patient, currently two years-old, has been ventilation free for the last several months; she gained social smile and reaches for objects. Mildly elevated cerebrospinal fluid (CSF) glycine levels with normal serum glycine were found in two patients. Whole exome sequencing of the first studied patient of family 1 revealed a rare homozygous frameshift variant in exon 6 of *SLC6A9* which encodes GLYT1. The variant co-segregated as expected within the extended family. Variants related to NKH and hyperekplexia genes were excluded. The phenotypic similarity of the three siblings of family 2 prompted Sanger sequencing of *SLC6A9*, which revealed a homozygous nonsense variant in exon 12, co-segregating appropriately. The essential role of GLYT1 in CNS function has been underscored by the lethal phenotype of homozygous knockout mice, which had severe motor and respiratory deficits and died during the first postnatal day. Zebrafish with *SLC6A9* homozygous missense mutation show diminished escape responses and absent swimming. Noteworthy, several GlyT1-knockout mice escaped premature death, as was seen in the surviving patient; therefore, it is thought that GlyT1 is critical early in life but its role may be compensated thereafter. While GLYT1 has been previously proposed in the etiology of NKH, this is the first description of *SLC6A9* mutations causing a unique human neurological syndrome. The mildly elevated CSF glycine in the presence of normal serum glycine may constitute a diagnostic marker for glycine transporter 1 defects.

2466F

Mutations in *RALGAPB* cause seizures, intellectual disability and cortical malformations. *A.A. Larson[1], A.M. Quintana[2], H.C. Yu[1], E.A. Geiger[1], J.A. Hernandez[2], N.V. Stence[3], R.I. Kuzniecky[4], N.J.L. Meeks[1], C.R. Coughlin II[1], T.H. Shaikh[1].* 1) University of Colorado School of Medicine, Department of Pediatrics, Section of Genetics; 2) University of Texas, El Paso, Department of Biological Sciences, Border Biomedical Research Center; 3) University of Colorado School of Medicine, Department of Radiology; 4) New York University School of Medicine, Department of Neurology, NYU Epilepsy Center.

Malformations of cortical development (MCD) are disorders of neurological morphogenesis that may be diagnosed in the setting of seizures and developmental delay. Ocular and endocrine abnormalities may coexist with MCD. We performed trio-based whole exome sequencing (WES) on subjects with MCD as well as optic nerve hypoplasia (ONH) and pituitary hormone deficiency (also known as septo-optic dysplasia). One subject with infantile spasms (IS), severe developmental delay, ONH, perisylvian polymicrogyria and panhypopituitarism was found to have compound heterozygous rare missense variants in *RALGAPB*: c.1980A>G (p.Lys640Glu) and c.2324G>T (p.Arg775Leu). *RALGAPB* encodes a Ral GTPase activating protein. It forms a heterodimeric enzyme with RALGAPA1, for which haploinsufficiency leads to brain malformations. The RALGAP complex downregulates signaling of RalA and B, which are involved in several intracellular signaling pathways. Notably, one of these pathways is the mTOR signaling pathway, which is disrupted in many known causes of MCD. Additionally, *RALGAPB* has been identified as a candidate gene in two large published cohorts of individuals with neurodevelopmental diseases that did not include detailed phenotypic data. We present detailed phenotypic data on one of these individuals reported by the Epi4K consortium with a *de novo* nonsense variant, c.3860C>G (p.Ser1287*), in *RALGAPB*, who had IS, focal cortical dysplasia and developmental delay. The second *de novo* variant, c.865_866delAT (p.Met289Valfs*3), was reported as part of a large cohort of individuals with autism. We were unable to obtain further phenotypic data for this individual. We performed morpholino knockdown of *ralgapb* in zebrafish, showing disruption of brain development, including malformation of the optic tectum and enlargement of the fourth ventricle. Injection of human wild type *RALGAPB* mRNA resulted in phenotype that overlapped those observed in the morphants, while higher concentrations of *RALGAPB* mRNA resulted in nonviable embryos, suggesting that *RALGAPB* overexpression may also be detrimental. Thus, our data suggest that either haploinsufficiency, due to dominant *de novo* variants or compound heterozygosity of missense variants, leading to partial loss of function in both alleles of *RALGAPB*, may lead to abnormalities in nervous system development and resultant neurological disease.

2467W

Genetic and clinical modifiers of age at onset in SPAST inherited spastic paraplegia. *L. Parodi[1], G. Banneau[2], F. Danjou[1], J. Guegan[1], E. Petit[1,2], S. Aitsaid[2], S. Fenu[1,2], S. Tezenas du Montcel[3], G. Stevanin[1,2,4], A. Brice[1,2], C. Depienne[5], A. Durr[1,2].* 1) ICM Institut du Cerveau et de la Moelle, INSERM U1127, CNRS UMR7225, Sorbonne Universités – UPMC Université Paris VI UMR_S1127, Paris, France; 2) APHP Department of Genetics, Groupe Hospitalier Pitié-Salpêtrière, Paris France; 3) Sorbonne Universités, UPMC Univ Paris 06 UMR_S1136, and INSERM UMR_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France; 4) Ecole Pratique des Hautes Etudes (EPHE), Paris Sciences et Lettes (PSL) Research University, Neurogenetics team, 75013 Paris, France; 5) Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM U964, CNRS UMR7104, Strasbourg, France.

Hereditary spastic paraplegias (HSPs) are a group of neurological disorders caused by a progressive upper motoneuron degeneration, resulting in spastic gait as core feature. HSPs are genetically heterogeneous, with at least 59 genes involved. *SPAST*/SPG4 encoding Spastin, a microtubule severing protein, is the most frequently mutated HSP gene, involved in 20-40% of HSPs. The resulting phenotype is characterized by an extreme variable age at onset, ranging from birth up to 70 years, including also unaffected mutation carriers. The aim of our study is to analyze *SPAST* mutation carriers and to search for modifiers such as genetic variants and/or environmental factors. We performed Sanger sequencing and WES in French HSP patients, collected through the SPATAX network and national reference centers for rare diseases in France. We applied a web-based questionnaire assessing environmental factors impact on the disease. Ninety-two questionnaires from *SPAST* carriers have been collected yet. One third (n=621) out of 2123 HSP patients screened for *SPAST* harbored a mutation, with 80% of point mutations mainly affecting Spastin ATPase activity domain. Mean age at onset was 29 ±19 years (from birth to 74 years, n=328) and mean disease duration 17±14 years (from 0 to 66 years, n=328). After assessing the disease severity using the SPATAX disability scale (0, absence of any functional handicap, to 7, bedridden; n=158), we observed that patients with a disease onset after 30 years tended to have a faster disease progression, i.e. faster disease progression seemed to be linked to older age. Survival analysis showed earlier onset associated with missense mutations, predicted pathogenic by at least 3 algorithms. Focusing on the detection of variants possibly modifying age at onset and severity, we identified 12 patients (0,6 %) carrying p.S44L variant, a previously described *SPAST* intragenic modifier, in combination with another major *SPAST* mutation. It was associated to early onset in a *SPAST* mutated family and to late onset in a SPG7 patient. To research other *SPAST* genetic modifiers, we performed WES on 34 SPG4-HSP patients. We detected in two early onset patients the presence of *HSPD1* variant p.G563A, already proposed as age at onset modifier. Our analysis showed evidence of age at onset variations linked to the nature of *SPAST* mutations in a large French cohort and confirmed the role of variant p.S44L, and possibly of variant p.G563A, as age at onset modifiers.

2468T

Whole-exome sequencing in a four-generation family with neuromyotonia and muscle weakness. *E.M. Ramos[1], A. Huang[1], D. Dokuru[1], Q. Wang[1], B. Merriman[2], G. Bettinsoli[1], D. Geschwind[1], A. Woo[3], G. Coppola[1].* 1) Department of Psychiatry and Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine at the University of California Los Angeles; 2) Department of Human Genetics, David Geffen School of Medicine at the University of California Los Angeles; 3) Santa Monica Neurological Consultants, David Geffen School of Medicine at the University of California Los Angeles.

We identified a large, four-generation family with an autosomal dominant history of neuromuscular disease, with affected individuals showing proximal hip flexor weakness, neuromyotonia and periodic paralysis attacks. To identify the genetic determinants in this family, we initially performed identity-by-descent mapping in five affected and three unaffected family members using genome-wide SNP genotyping with 250K arrays. This approach identified five candidate loci (segregating haplotypes on chromosomal regions 1p31.1-p22.2, 1q32.3-q41, 2p11.2-q14.1, 11q22.1-q22.3 and 11q25) spanning a total of 43.4Mb. We then performed whole-exome sequencing in all subjects. Briefly, paired-end sequence reads (sequenced on an Illumina HiSeq2500 using a SeqCap EZ Exome capture library) were mapped to the GRCh37/hg19 reference genome and variants were interactively joint-called with GATK Haplotype Caller according to GATK Best Practices recommendations. A total of ~116,000 variants were identified in the five affected and three unaffected samples sequenced. We filtered out variants not segregating with the disease and all non-coding and synonymous changes. Then, under the assumption that the mutation causing this rare autosomal dominant disease is extremely rare, we also excluded variants that are present in the databases with a minor allele frequency greater than 1%. After applying these filtering criteria, we narrowed our search to seven putative disease-causing variants. Of these, six are located in genes within the candidate loci, one of which is a strong biological candidate gene. We are currently following up on these results and recruiting additional family members.

2469F

Partial duplication of *Desert hedgehog* **(***DHH***) gene in patients with minifascicular neuropathy accompanied by 46XY partial gonadal dysgenesis.** *N. Sato[1,2], R. Maekawa[2], H. Ishiura[1], J. Mitsui[1], H. Naruse[2], K. Sugie[3], J. Shimizu[1], J. Goto[4], S. Tsui[1], Y. Shiio[2].* 1) Neurology, The University of Tokyo, Tokyo, Japan; 2) Department of Neurology, Tokyo Teishin Hospital, Tokyo, Japan; 3) Department of Neurology, Nara Medical University, Nara, Japan; 4) Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, japan.

Minifascicular neuropathy (MN) is an extremely rare developmental malformation in which peripheral nerves are composed of many small fascicles. Only one case, who was associated with 46XY gonadal dysgenesis (GD), carried a homozygous missense mutation at the initiating codon in exon 1 of the *desert hedgehog* (*DHH*) gene. Molecular genetics of the other cases, however, have not been delineated. *DHH* is a member of the hedgehog family which also includes sonic hedgehog and indian hedgehog, and responsible for signals for the formation of the connective tissue sheathing around peripheral nerves during early development. Here, we analyzed molecular basis of two consanguineous Japanese families with MN. Two of the patients showed 46XY GD, whereas a patient with 46XX karyotype apparently shows normal sexual development. We performed multipoint parametric linkage analysis based on results of genotyping by SNP-array 6.0 (Affymetrix, Santa Clara, CA) in samples from the three affected individuals. After multipoint LOD score calculations using the pipeline software SNP-HiTLink under the assumption of autosomal recessive inheritance, five regions in four chromosomes with the highest multipoint LOD scores of 2.10. The candidate regions overlapped with the locus of *DHH* on chromosome 12. We next tried to amplify the entire *DHH* region employing a long-range PCR method. The size of the PCR products of each individuals was larger by several hundred basepairs than that obtained from a control, suggesting a rearrangement of the genome. Direct nucleotide sequence analysis of the PCR products revealed the tandem duplication of which the 5'-end was in intron 1 and 3'-end in exon 2 of *DHH* gene. Haplotype analysis showed a common founder chromosome. In conclusion, we identified a novel mutation in a human *DHH* gene. The patients with 46XY karyotype exhibited MN accompanied by 46XY GD, whereas a patient with 46XX karyotype apparently shows normal sexual development. These data further support that *DHH* is an important molecule in both male gonadal differentiation and perineural formation in peripheral nerves.

2470W

Whole-exome sequencing identifies a missense mutation in *hnRNPA1* **in a family with flail arm amyotrophic lateral sclerosis.** *S. Shu[1,3], Q. Liu[2,3], R.R. Wang[1,3], F. Liu[1,3], B. Cui[2,3], X.N. Guo[1,3], C.X. Lu[1,3], X.G. Li[2,3], M.S. Liu[2,3], B. Peng[2,3], L.Y. Cui[2,3], X. Zhang[1,3].* 1) McKusick-Zhang Center for Genetic Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China; 2) Department of Neurology and Laboratory of Clinical Genetics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China; 3) Neuroscience Center, Chinese Academy of Medical Sciences, Beijing 100005, China.

Flail arm syndrome (FAS), one of the atypical subtypes of amyotrophic lateral sclerosis (ALS), is typically described as slowly progressive, predominantly proximal, symmetric weakness and wasting of the upper limbs, with no significant lower limb or bulbar muscle disability for several years. Here we report the identification of the disease-causing gene of a family diagnosed as FAS. Using targeted next-generation sequencing, we excluded 24 known ALS genes. We detected a novel missense mutation in *hnRNPA1*, c.862/1018C>T (p.P288S/P340S) through whole-exome sequencing, and it was co-segregated with disease in the FAS family. HnRNPA1 belonging to the hnRNP family, is a widely expressed nuclear protein and shuttles continuously between the nucleus and cytoplasm. The mutation substituted the exact proline residue in the encoded PY-nuclear localization signal (NLS) of hnRNPA1 protein, and the residue is highly conserved in evolution. Mutant *hnRNPA1* showed altered intracellular localization, resulting in formation of cytoplasmic inclusions that co-localized with stress granules in transfected Hela cells. Further mutation screening of *hnRNPA1* in additional 251 Chinese patients with ALS (including 7 sporadic FAS) detected two rare variants with unknown significance. These variants lie in the prion-like domain of hnRNPA1 long isoform, which was detected exclusively in the central nervous system. Our results suggest that *hnRNPA1* is the causative gene in the family with flail arm ALS. This further expanded disease phenotype of *hnRNPA1* mutations.

2471T

Recessive porphyria due to homozygous *HMBS* **mutation: A cause of leukodystrophy with cataracts.** *C. Stutterd[1,2,3], V. Lukic[4], P. Diakumis[4], R. Fullinfaw[5], J. Archer[6], R. Leventer[1,2,3], M. Bahlo[4], P. Lockhart[2,3], M. Delatycki[2,3].* 1) Royal Children's Hospital , Melbourne, Victoria, Australia; 2) Murdoch Childrens Research Institute, Melbourne, Victoria, Australia; 3) University of Melbourne, Melbourne, Victoria, Australia; 4) Walter and Eliza Hall Institute, Melbourne, Victoria, Australia; 5) Royal Melbourne Hospital, Melbourne, Victoria, Australia; 6) Austin Hospital, Melbourne, Victoria, Australia.

We have identified a new genetic and potentially-treatable cause for leukodystrophy with cataracts and expanded the known phenotype associated with the rare occurrence of recessive acute intermittent porphyria (AIP). Two adult siblings from a consanguineous family were referred for whole exome sequencing (WES) as part of a gene discovery project investigating unclassified leukodystrophies. They share a novel phenotype characterised by leukodystrophy, childhood-onset cataracts, progressive ataxia, lower limb spasticity and peripheral neuropathy. Their cerebral MRI shows periventricular demyelination with thalamic cavitations. WES identified a homozygous variant in *HMBS* (c.251C>A, p.A84D) in both affected siblings. Their parents and two unaffected siblings were found to be heterozygous for this variant. The pathogenicity of this variant has been demonstrated by reduced erythrocyte HMBS activity in the proband to 20% normal. Heterozygous variants in this gene cause acute intermittent porphyria (AIP), a dominantly inherited neurological disorder with reduced penetrance. Only six cases of recessive AIP have been reported, all presenting in early childhood with severe central and peripheral neurological impairment, four with childhood onset cataracts and one found to have leukodystrophy. All had <4% of normal HMBS activity and died in childhood. The level of residual HMBS activity appears to correspond with disease severity. There are no reported cases of adults with recessive AIP and so this finding represents a novel phenotype and a new genetic cause for leukodystrophy with cataracts. The less affected of the two sibs is being worked up for liver transplantation with the aim of halting progression of this neurodegenerative disorder.

2472F

Genome-wide association study identifies potential genetic modifiers in Charcot-Marie-Tooth Disease type 1A. *F. Tao[1], G. Beecham[1], A. Rebelo[1], S. Blanton[1], L. Abreu[1], F. Baas[2], B.O. Choi[3], D. Pareyson[4], M. Reilly[5], M. Shy[6], S. Zuchner[1], Inherited Neuropathy Consortium.* 1) University of Miami Med School, Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Genome Analysis, Academic Medical Centre, Amsterdam, The Netherlands; 3) Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 4) Department of Clinical Neurosciences, C. Besta Neurological Institute, Milan, Italy; 5) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, London, UK; 6) Department of Neurology, University of Iowa, Iowa City, Iowa.

   Charcot-Marie-Tooth disease type 1A (CMT1A) is caused by a uniform 1.5-Mb duplication on chromosome 17p11.2, which includes peripheral myelin protein 22 (*PMP22)* gene. It is the most common subtype of inherited peripheral neuropathies, and is estimated to affect 1 in 10,000 individuals worldwide. CMT1A patients usually present the classic CMT phenotype, but also with high clinical variability in their phenotypic presentation and disease severity. The cause of the phenotypic variability is largely unclear. This study uses genome-wide association study (GWAS) to identify potential genetic variants that modify the severity of motor function involvement in CMT1A. Using Illumina OmniExpress platform, we analyzed 647,267 genomic markers of 303 CMT1A patients from European ancestry: 165 individuals with severe foot dorsiflexion (MRC score of 0 to 3), and 138 individuals with mild foot dorsiflexion (MRC score of 5). Association analyses were performed using R package GWAF, adjusting for patients' age at exam, sex and family relations in the framework of logistic regression via generalized estimating equations. After Bonferroni corrections, three SNPs showed genome-wide significance in association with severe foot dorsiflexion: rs10910527 (chr1:232509189, P=1.851E-08, odds ratio=12.3, 95% confidence interval=(5.1,29.5)), rs7536385 (chr1:232601208, P=1.851E-08, odds ratio=12.3, 95% confidence interval=(5.1,29.5)), rs4649265 (chr1:232591925, P=2.010E-08, odds ratio=10.1, 95% confidence interval=(4.5,22.7)). The three SNPs are in high linkage disequilibrium, and all locate within intronic regions of signal induced proliferation associated 1 like 2 (*SIPA1L2*), which codes for a putative GTPase with unknown functions. Neuro-2a cells transfected with SIPA1L2-Myc plasmid showed colocalization of SIPA1L2 with actin in growth cones of axonal projections, suggesting potential involvement of *SIPA1L2* in neuronal cell functions. Our results suggest genetic variants in *SIPA1L2* as potential modifiers in the severity of motor function impairment in CMT1A. We are currently analyzing whole genome sequencing data of phenotypically extreme patients to pinpoint specific genetic variation underlying this statistical signal. We will also conduct functional studies to further characterize the role of *SIPA1L2* in the disease.

2473W

Identification of a novel *LAMC3* variant in a girl with seizures, global developmental delay and cortical malformation. *J. Zambonin[1], S. Venkateswaran[2], Y. Xi[3], T. Hartley[4], K. Boycott[1,4], J. Parboosingh[3], D. Dyment[1,4], Care4Rare Canada Consortium.* 1) Department of Medical Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Department of Pediatrics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 3) Department of Medical Genetics and Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada; 4) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.

   Occipital cortical malformations (OCCM; MIM 614115) are rare autosomal recessive malformations characterized by pachygyria and polymicrogyria of the occipital lobes. Affected individuals have childhood onset seizure disorders including absence, myoclonic-atonic and generalized tonic-clonic semiology. Homozygous or compound heterozygous mutations in the gamma-3 chain isoform of laminin (gamma-3 or *LAMC3*) gene are causative for this condition. Laminins are a family of extracellular matrix glycoproteins that are necessary for cell adhesion, differentiation, migration and neurite outgrowth. LAMC3 is expressed in the developing fetal brain and is vital for neuronal migration and cortical organization. Until now, only four cases of OCCM with variants in *LAMC3* have been reported in three unrelated individuals and six siblings from a large consanguineous family. All mutations were predicted to result in premature truncation of the *LAMC3* gene product with one individual having a compound heterozygous missense mutation and truncating mutation. Here, we report a novel and likely pathogenic variant in *LAMC3* identified by next-generation sequencing. Our patient was a 13 year old female born to consanguineous parents. She presented with myoclonic, atonic and atypical absence seizures and global developmental delay in keeping with others with mutations in *LAMC3*. Of note, neuroimaging showed bilateral temporoparietal pachygyria and polymicrogyria with only mild involvement of the occipital lobe rather than isolated occipital involvement. Microarray testing did not show any pathogenic copy number variants however there was extensive loss of heterozygosity. Next-generation sequencing was performed using a capture kit that targets over 4800 clinically-relevant disease genes. She was found to have the homozygous and likely pathogenic variant NM_006059.3:c.3190C>T (p.Gln1064*) in *LAMC3*. This mutation has not previously been reported and was predicted to cause a premature truncation of LAMC3. We describe this novel variant in *LAMC3* and expand the phenotype of OCCM to include pachygyria and polymicrogyria in a distribution pattern that involves the posterior parietal and temporal lobes. Therefore the role of LAMC3 appears to extend beyond cortical organization of the occipital regions.

2474T

***CDH18* and *C4orf21* (*ZGRF1*) variants segregate separately in two multigenerational families with overlapping phenotypic presentations of childhood apraxia of speech.** *B. Peter[1,2,3], E.M. Wijsman[4,5], A.Q. Nato[4], M.M. Matsushita[4], K.L. Chapman[6], I.B. Stanaway[7], J. Wolff[4], K. Oda[4], V.G. Gabo[4], W.H. Raskind[4,8], University of Washington Center for Mendelian Genomics.* 1) Speech & Hearing Sci, Arizona State University, Tempe, AZ; 2) Communication Sciences and Disorders, Saint Louis University, Saint Louis, MO; 3) Speech and Hearing Sciences, University of Washington, Seattle, WA; 4) Medical Genetics, Dpt. of Medicine, University of Washington, Seattle, WA, United States; 5) Biostatistics, University of Washington, Seattle, WA, United States; 6) Communication Sciences and Disorders, University of Utah, Salt Lake City, Utah, United States; 7) Environmental and Occupational Health, Environmental Toxicology, University of Washington, Seattle, WA, United States; 8) Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States.

Childhood apraxia of speech (CAS) is a severe speech sound disorder characterized by difficulty translating articulatory intentions into precisely timed and coordinated movement sequences. There is evidence that CAS has a genetic etiology, for instance rare inherited *FOXP2* point mutations or a *de novo BCL11A* deletion, but in most cases, the causal variations remain unknown. Here, we present two multigenerational families with familial CAS. Phenotypic manifestation in the two families overlap partially whereas genetic profiles differ completely. Linkage analysis, IBD analysis, and exome sequencing showed that in Family A, affected members share a *CHD18* variant; additional variants of interest were found in *MYO10, NIPBL, GLP2R, NCOR1, FLCN, SMCR8, NEK8,* and *ANRKD12.* In Family B, the affected members share a rare and deleterious *C4orf21* (*ZGRF1*) variant.   Here, we extend these results to phenotypic characteristics associated with inheritance of each genomic region. In both families, affected children did not produce intelligible speech until age 5 or 6 years and required between three and seven years of speech therapy to acquire normalized speech. Affected children and adults with a CAS history in both families had difficulty with nonword imitation and rapid nonword reading, tasks that require intact phonemic awareness, sequential processing, and motor programming. In Family A, affected members showed additional sequential processing deficits in the oral motor domain in that speeds during rapid syllable repetition in the alternating (e.g., /pata/) but not repetitive (e.g., /papa/) condition were slower than expected. In Family B, additional sequential processing deficits were seen during nonword imitation in that substantially more phoneme sequencing than phoneme substitution errors were made. *CDH18, C4orf21,* and five of the other genes of interest are expressed in the cerebellum, a brain region previously implicated in imaging studies of individuals with *FOXP2* mutations and known to influence temporal coordination of not only motor but also linguistic processes. To what extent the phenotypic similarities and differences in the sequential processing deficit are due to gene-specific influences on the brain remains to be investigated. Results are consistent with the view that, similar to other neurodevelopmental disorders CAS is both phenotypically and genetically heterogeneous and may result from single or multiple variations.

2475F

Evaluating motor functions and biomarkers for Fukuyama type congenital muscular dystrophy. *M. Taniguchi[1], R. Harada[2], N. Tatsuya[3], H. Awano[1], L. Tomoko[4], Y. Takeshima[4], K. Kobayashi[5], I. Morioka[1], K. Iijima[1], T. Toda[5].* 1) Pediatrics, Kobe Univ Grad Sch Med, Kobe, Hyogo, Japan; 2) Rehabilitation Medicine, Kobe Univ Grad Sch Med, Kobe, Hyogo, Japan; 3) Radiology, Kobe Univ Grad Sch Med, Kobe, Hyogo, Japan; 4) Pediatrics, Hyogo Medical College, Nishinomiya, Hyogo, Japan; 5) Neurology/Mol Bra Sci, Kobe Univ Grad Sch Med, Kobe, Hyogo, Japan.

Fukuyama type congenital muscular dystrophy (FCMD) is an autosomal recessive, severe childhood muscular dystrophy with brain and eye involvement. FCMD is mainly caused by an ancestral insertion of 3-kb SVA (SINE-VNTR-*Alu*) type retrotransposal element into the 3' untranslated region of the causative gene, *fukutin*. Recently, we testified that pathogenic exon trapping by SVA transposon cause splicing abnormality in FCMD. We also succeeded in optimizing antisense oligonucleotides therapy for FCMD model mice and FCMD myoblasts. Since FCMD is an incurable disease, there has been no biomarkers for FCMD. Moreover, there is few study on natural history of FCMD. There is a need to develop sensitive, non-invasive outcome measures of FCMD patients that can be readily available to human clinical trials in the near future. To find specific serum biomarkers and to comprehend natural history of FCMD patients, we collected serum and clinical data from patients. We tested on serum biomarkers by measuring muscle specific microRNAs. Clinical data contains motor function score, muscle elastography (shear wave elastography (SWE), and whole body muscle computed tomography with ultralow level irradiation. As a result, microRNA 206 (miR206), which is highly expressed in regenerating muscle fibers was significantly overexpressed in FCMD patients compared to normal controls. Correlation coefficient of miR206 with serum CK level, serum creatinine level, and motor function scores were also high. MiR 206 was especially high in FCMD patients with high muscle contents, suggesting remaining sparing capacity of muscle regeneration. SWE showed significantly high elasticity in biceps brachii and brachial muscle but not high in lower extremities in FCMD patients, compared to normal control, because of high content of fat infiltration due to disuse of lower extremities. In conclusion, serum miR206 and SWE is useful for monitoring muscle wasting progression and motor function level of FCMD.   .

2476W

Spinal muscular atrophy carrier frequency in Russian Federation. *V.V. Zabnenkova[1], E.L. Dadali[1], M.G. Spiridonova[2], R.A. Zinchenko[1], A.V. Polyakov[1].* 1) Research Centre for Medical Genetics, Moscow, Russian Federation; 2) The Research Institute for Medical Genetics, Tomsk, Russian Federation.

Introduction: The proximal spinal muscular atrophy (SMA) is a severe autosomal recessive neuromuscular disease characterized by degeneration of alpha motor neurons in the spinal cord, resulting in progressive proximal muscle weakness and paralysis. 95% SMA patients have homozygous deletion of *SMN1* gene. The remaining 5% cases are caused by compound heterozygous mutations: *SMN1* deletion on the one allele and a subtle mutation on the other allele. SMA carrier frequency is 1 per 40–50, incidence is 1 per 6000-10000 live births in the general population. SMA carrier frequency has not been estimated in Russia previously. **Objective:** This study sought to evaluate the incidence of deletion carrier state in the gene *SMN1* in Russian Federation. **Methods:** Genomic DNA was isolated from the peripheral blood cells from 2253 individuals of the general population in Russia using Wizard® Genomic DNA Purification Kit (Promega, United States). The number of *SMN1* gene copies was detected using a non-commercial kit, developed in laboratory of RCMG on base of Multiplex Ligation-dependent Probe Amplification (MLPA). The capillary electrophoresis was carried out in an ABI3130 DNA analyzer (Applied Biosystems). **Results:** 63 heterozygous carriers of the exons 7 and/ or 8 *SMN1* gene deletion were identified. The frequency of SMA carriers estimated in this study is 1 in 36 individuals (2.8 %). The calculated incidence of SMA is equal to 1 per 5184. **Conclusions:** Detected high frequency carrier of spinal muscular atrophy in Russia shows the importance of carrier screening. It is highly necessary for IVF program donors and for identifying of asymptomatic carrier couples with no family history but having a risk for offspring with SMA.

2477T

Aicardi-Goutières syndrome and the *ADAR1* dominant negative mutation: A case report. *A.I. Scott[1], A. Myers[2], G.M. Enns[2], T.M. Cowan[1].* 1) Department of Pathology, Stanford University School of Medicine, Stanford, CA; 2) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.

Aicardi-Goutières syndrome (AGS) [MIM: 225750] is an autoimmune encephalopathy with hallmark features of basal-ganglia calcification, leukodystrophy, and CSF lymphocytosis. Psychomotor retardation, dystonia, and spasticity are present in many cases. While there is clinical and genetic heterogeneity, all six genetic forms abnormally accumulate DNA or RNA and induce an inflammatory response. Despite increased identification of patients with AGS, the relationship between genotype and phenotype remains unclear. Approximately 7% of AGS patients have mutations in an RNA adenosine deaminase (ADAR1), a double stranded RNA editing enzyme. Sixteen *ADAR1* mutations have been described in AGS patients; of these, a single dominant change, G1007R, abolishes normal enzyme function and is predicted to be clinically severe. Within this specific group there is significant heterogeneity: the seven reported probands share features of gait disturbances, dystonia, and neurological involvement, but differ in age of onset, rate of progression, and extent of skin findings. At least two heterozygous parents of affected probands had no noted clinical abnormalities. We describe a new case of the G1007R mutant ADAR1 with an early onset and rapid progression of AGS symptoms. Our patient presented with gait disturbances at 15 months of age following normal development. A brain MRI at 20 months was normal. He is currently four years old and has progressive muscle involvement including truncal hypotonia, choking, and lower extremity spasticity. He has freckling on his extremities: hands, feet, and knees. Whole exome sequencing identified *ADAR1* c.3019G>A (p.G1007R) leading to the diagnosis of AGS. Our case underscores the phenotypic variability seen among patients with ADAR1 G1007R, which can lead to extensive testing; five of the probands had non-specific diagnoses including bilateral striatal necrosis, nonsyndromic spastic paraplegia, and dyschromatosis symmetrica hereditaria before entering research studies to identify a causative mutation. The underlying mechanism of anti-viral immune activation links these patients to the diagnosis of AGS, but the specific disturbances of nucleic acid metabolism may provide clues to the phenotypic variability. Identification of more patients with ADAR1 G1007R will more fully define clinical features and may help identify critical targets of ADAR1 activity.

2478F

Whole Genome Next Generation Sequencing [WGNGS] identifies inherited variants contributing to asperger syndrome in a South African family with apparent autosomal dominant Mendelian inheritance. *M. Kambouris[1,2], Y. Al-Sarraj[3], D. Ahram[4], H. Shaath[3], R. Thompson[3], F. Al-Shaban[3], H. El-Shanti[5,6]. 1) Pathology-Genetics, Sidra Medical & Research Center, Doha, Qatar; 2) Genetics, Yale University School of Medicine, New Haven, CT, USA; 3) Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 4) Translational Medicine, Sidra Medical & Research Center, Doha, Qatar; 5) Pediatrics, University of Jordan School of Medicine, Amman, Jordan; 6) Pediatrics, University of Iowa, Iowa City, IA, USA.*

A South African family with a father and three sons all affected with Asperger syndrome that presents with high intellectual functionality, significant difficulties in social interaction and nonverbal communication, repetitive behaviors and physical gestures was studied by WGNGS and comparative Genome analyses. The father is a high-ranking banker with superb banking skills. He does not interact with colleagues and co-workers except through a single buffer individual. At home, he interacts only with his spouse and is quite distant from all immediate and extended family members. All three affected sons are exceptionally intelligent, learned written & spoken English as a second language by the age of 5, without tutoring just from television and books. The oldest son, published a book in English at the age of 7, but did not graduate high school since he can not undergo the required testing. The second son completed the academic high school curriculum by age 12, but also could not undergo the required testing until the age of 16. The third son completed the high school academic curriculum very early, although he did not attend school until the age of 8, since he refused to wear clothes and remained naked for the first 8 years of his life. All four affected members have certain routines, stereotypic behaviors and gestures which they control to a limited extend. Due to the strong heritability in this family, data was mined for monoallelic variations in the father inherited by all three affected sons (autosomal dominant inheritance) in protein coding genes with in-silico predicted damaging effects that could possibly include a major offending gene. Among others, these include genes and biological pathways associated with Autism Spectrum Disease (ASD): a nonsense mutation *HYAL4* [c.628C>T/p.210Arg>*] in a hyaluronidase that intracellularly degrades the glycosaminoglycan hyaluronan involved in brain function and development and two missense mutations that affect highly conserved amino acids: *PARK2* [c.110C>T/p.37Pro>Leu] a component of a multiprotein E3 ubiquitin ligase complex involved in protein degradation and *FAT1* [c.2563C>A/p.855Gly>Arg], a cadherin required for cell-cell association and actin organization. Intronic, regulatory and and non-coding variants should also have been considered. Determination of a single Autosomal Dominant offending mutation is difficult. This family presents a Mendelian model of the polygenic-multifactorial Asperger variant of ASD.

2479W

Precision medicine in juvenile psychosis at Boston Children's Hospital. *C.A. Brownstein[1], A.H. Beggs[1], P. Agrawal[1], J. Gonzalez-Heydrich[2]. 1) Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Developmental Neuropsychiatry Program, Boston Children's Hospital, Boston MA.*

Diagnostic Sequencing is a cornerstone of Precision Medicine, especially at a tertiary hospital. Two initiatives at BCH are illustrative: 1) the Developmental Neuropsychiatry Program (DNP), and 2) the Manton Center for Orphan Disease Research. The study of rare Mendelian forms of Juvenile Psychosis is an effective way to discover new candidate genes for the condition. Boston Children's Hospital has created the infrastructure needed for large-scale psychiatric research and treatment discovery with the Developmental Neuropsychiatry Program (DNP). The DNP seeks to develop therapeutics to prevent the development of schizophrenia in at-risk children by identifying causative mutations in the youngest patients presenting with psychosis, and providing targeted care for others who present with the same conditions. Once a causative mutation or copy number variation is found and functionally confirmed, we have the ability to identify whether there are other patients with the same mutation or variation at BCH, and assess whether they are in need of psychiatric services. Using this method, we have identified patients with unaddressed psychotic or prodromal symptoms and have some success stories of getting patients and families necessary treatment. We also have some case examples of identifying patients too late- once we identified that 16p13.11 copy number loss could present as juvenile psychosis, we performed a search and identified a family with two psychotic breaks that perhaps could have been prevented. We are able to do this work due to the Manton Center for Orphan Disease Protocol. The Manton Center is one of the first centers in the world solely devoted to the study of rare diseases, focusing on patient-centered research and funding existing research efforts on rare diseases. The Center maintains a general IRB protocol suitable for enrolling subjects with any rare or undiagnosed condition, and provides consenting, enrollment and genotyping services and training, enabling discovery. If something actionable is found, it is CLIA verified and able to be returned to the medical record. Through the use of this common genomic toolkit allowing for seamless transfer between the clinical and research realms, BCH is developing shared informatics resources and intellectual expertise. The success of this program has provided precision medicine to this vulnerable population. .

2480T

Genotype-phenotype correlation in *ADNP*-related syndromic autism. *F. Kooy[1], A. Van Dijck[1], I. Gozes[2], A. Vulto-van Silfhout[3], E. Eichler[4], B. de Vries[3], G. Vandeweyer[1].* 1) University of Antwerp, Antwerp, Belgium; 2) Tel Aviv University, Tel Aviv, Israel; 3) Radbout University, Nijmegen, the Netherlands; 4) University of Washington.

Among the many hundreds of genes linked to autism, only a few have been clearly characterized. Recently, we discovered that ADNP is one of a still very limited number of genes that lead to autism in a substantial proportion of cases. ADNP encodes a neuroprotective protein that is involved in chromatin remodeling. Interestingly, the mutations in the original description of the disorder were all located in the fith and last exon of the gene and were demonstrated to escape non-sense mediated decay and we hypothesized a dominant negative mutational mechanism. Since the description of the first 10 patients, we have collected an additional 50 patients. Detailed and structural clinical information was collected from all as well as the nature of the mutations. In this cohort, mutations in other but the last exon are present and by comparing the genotype with the genotype of all patients, we have been able to estabish a detailled genotype-fenotype correlation. Such information is crucial both in understanding the mutational mechanism as well as for the interpretation of ADNP variants discovered in autistic patients. Strikingly, the identified mutations appear to cluster at specific positions in the gene. Almost one third of the patients have an almost identical 4 bp deletion mutation, either g.49508752_49508755delTTTA or g.49508757_49508760delTTAA in exon 5. Clustering of de novo, rare variants is suggestive of a mutation predisposition mechanism, potentially as a result of a particular local genomic architecture. Mfold analysis (web server for nucleic acid folding and hybridization prediction) showed that the recurrent mutation is located in the stem of the same short hairpin, suggestive of an underlying mechanism involving a DNA repair defect following pausing of a replication fork at these hairpins.

2481F

Characterizing a novel variant in histone demethylase gene *KDM5C* associated with intellectual disability and autism spectrum disorder. *C. Vallianatos, C. Keegan, S. Iwase.* Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI.

Dysregulation of histone methylation has emerged as a major contributor of neurodevelopmental disorders such as autism spectrum disorder (ASD) and intellectual disability (ID). The X chromosome gene *KDM5C* encodes a histone H3 lysine 4 (H3K4) demethylase, and *KDM5C* mutations are frequently associated with syndromic X-linked ID, Claes-Jensen type (MIM 300534). Genome-wide, H3K4 methylation is enriched at promoters and enhancers of actively transcribed loci, and tri-methylated H3K4 directly interacts with transcription machinery to promote gene expression. KDM5C is therefore regarded as a transcriptional repressor and key chromatin regulator of neuronal development and function. How individual *KDM5C* mutations lead to cognitive impairment remains unclear. Here, we report a new *KDM5C* variant and assessment of its molecular and cellular pathogenicity. We performed exome sequencing of a young male with ID, ASD, and developmental abnormalities, and revealed a maternally inherited new missense mutation in *KDM5C*, p.Arg1115His (c.3344G>A). Known *KDM5C* mutations tend to cluster near the JmjC catalytic domain and lead to decreased enzymatic activity, suggesting a loss-of-function pathogenic mechanism. In contrast, p.Arg1115His is distant from the enzymatic core, and lies upstream of a PHD domain (PHD2) which has unknown function. PHD domains can be histone binding modules, however our histone peptide binding studies suggest PHD2 may bind non-histone substrates. We hypothesize p.Arg1115His may affect interaction of KDM5C with known binding partners. We focus on RACK7, a histone binding protein known to interact with KDM5C to suppress over-activation of enhancers. Immunoprecipitation of KDM5C shows binding to RACK7 is dramatically reduced by p.ArgR1115His. KDM5C p.ArgR1115His exhibits normal nuclear localization in mouse cortical neurons, yet leads to decreased protein stability. Our results suggest p.Arg1115His interferes with the KDM5C-RACK7 interaction, an important enhancer-suppressing complex, and reduces KDM5C stability. Current efforts are focused on assessing H3K4 demethylase activity of KDM5C p.ArgR1115His. Future studies will test the ability of this variant to be recruited to chromatin and restore aberrant gene expression in *Kdm5c*-null mouse cortical neurons. Our work informs us about the pathogenicity of this new variant and the role of KDM5C PHD2, and lead to a better understanding of H3K4 methylation dynamics dysregulated in neurodevelopmental disorders.

2482W

SATB2-associated syndrome: Report of a new case associated with a new mutation in the SATB2 gene. *Z.L. Xu, G. Cowing, M. Szybowska, C. Li.* Department of Genetics, McMaster Children's Hospital, Hamilton, ON, Canada.

Special AT-rich sequence-binding protein 2 (*SATB2*) is a highly-conserved gene that encodes a transcription regulator and is involved in chromatin remodeling. *SATB2* haploinsufficiency results in *SATB2*-associated syndrome (SAS;OMIM#608148); a clinically recognizable condition characterized by intellectual disability, limited or absent speech and language, facial dysmorphism that includes hypertelorism, downslanting palpebral fissures, dental anomalies, and cleft palate. Other features of SAS include behavioral problems, sleep disturbances, seizures, and osteopenia. To our knowledge, only ten patients have been described in the literature. We report a 13-year-old male in whom trio-based whole-exome sequencing revealed a unique *de novo* mutation in exon 9 of *SATB2*, c.1286G>A (p.Arg429Gln). The genetic change has not been previously reported but we believe it to be pathogenic based on: 1) clinical phenotype. The proband demonstrates features previously reported in SAS—severe global developmental delay, dysmorphic features with downslanting eyes, prominent low-set ears, a prominent nose, an open-mouthed hypotonic facies as well as high-arched palate, bifid uvula. The patient also had macrodontia and retrognathia. Bilateral 5th finger clinodactyly and 2-3 toe syndactyly were also part of his physical features. Behavioral problems are notable for hyperactivity, aggressiveness, skin-picking, sleep disturbance and autism spectrum disorder. 2) molecular and database information. The mutation is a semi-conserved amino acid substitution that occurs at an evolutionarily conserved position of the gene. It is absent in the normal population from the NHBI exome sequencing project with an allelic frequency of GMAF 0.00140 (A). In summary, whole-exome sequencing identified a new genetic change in the *SATB2* gene that is likely pathogenic in a patient with novel as well as overlapping clinical features of SAS. The case therefore may serve to expand the phenotype of SAS.

2483T

Large-scale untargeted metabolomic profiling in Pitt-Hopkins syndrome uncovers phenotypic-associated amino acid and lipid abnormalities. *J.T. Alaimo, S.H. Elsea.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Transcription factor 4 (*TCF4*) encodes a helix-loop-helix transcription factor that, when haploinsufficient, results in the neurodevelopmental disorder Pitt-Hopkins syndrome (PTHS). PTHS is characterized by severe intellectual disability, absent speech, seizures, breathing abnormalities, behavioral problems, and delayed motor milestones in which more than half fail to walk independently. In addition, natural variation at this locus is associated with the complex disease schizophrenia. The phenotypic spectrum associated with *TCF4* variation suggests a unique role for the protein in pathways mitigating neurodevelopment and cognitive function; therefore, we sought to undercover functional pathways regulated by TCF4 through the application of a small molecule metabolomics screen. We examined plasma samples of 13 individuals with a molecularly confirmed diagnosis of PTHS that were predominately male and ranged in age from 1.5 to 20 years. Our cohort had a history of global developmental delay, absent or sparse speech, seizures and sleep anomalies. Metabolomic profiles from this cohort were significantly enriched for alterations in lipid and amino acid associated pathways. All individuals displayed abnormal lipid profiles, but consistent alterations were observed in fatty acid and phospholipid metabolism. We identified elevated oleamide levels, an endogenously metabolized fatty acid amide associated with sleep problems, a phenotypic feature of our cohort. In addition, abnormalities across palmitic acid metabolism and utilization were identified, including elevated palmitic amide a known marker for schizophrenia. Amino acid abnormities were observed across 12 of 13 samples with consistent alterations found in pathways associated with oxidative stress and mitochondria function, such as methionine sulfoxide and N-acetylaspartate. Interestingly, 3 individuals within our cohort displayed alterations in carnitine metabolism suggesting carnitine biosynthesis defects and were recommended for further clinical testing. Taken together, the application of metabolic profiling in individuals with PTHS suggests that TCF4 plays a functional role in small molecule pathways that may contribute to the phenotypic features of the disorder. Future studies discerning the cellular defects associated with these metabolic alterations will likely present new opportunities for therapeutic interventions in PTHS.

2484F

Exome sequencing identifies a novel mutation in *C10orf2* in a consanguineous family with intellectual disability. *A. Alkhateeb[1,2], S. Aburahma[1], W. Habbab[2], I.R. Thompson[2].* 1) Jordan University of Science and Technology, Irbid, Jordan; 2) QBRI-Hamad Bin Khalifa University, Doha, Qatar.

Intellectual disability is a heterogeneous disease with an increasing number of causal genes and mutations identified. Here we examined a first-cousin-in-marriage Arab family with two male and female sibs affected with intellectual disability. Patients were hypotonic with weakness, delayed speech and walking, bilateral foot drop, absent deep tendon reflexes and no history of epilepsy. MRI of the brain showed delayed myelination for age. The two patients and their parents were subjected to SNP microarray genotyping to identify regions of autozygosity in the patients. Four autozygosity regions were identified on chromosomes 10, 11 and 16. Whole exome sequencing was done for patients and their parents. Data were filtered and analyzed by Ingenuity Variant Analysis tool. Only four variants survived stringent filtering and only three of those resided in the autozygous regions. Two of the three variants were listed in dSNP and predicted benign in Polyphen-2. The third variant c.205A>T (p.I69F) in *C10orf2* gene was not listed in dSNP, localized in a highly conserved region of the protein, and predicted to be possibly damaging by Polyphen-2. Mutations in *C10orf2* gene has been implicated in intellectual disability. This novel mutation is located 10 amino acids upstream of the primase domain that is known to play a critical role in positioning the helicase to its target, and thus, seems likely to affect the helicase function of the protein leading to disease. Our analysis illustrates the power of whole exome and SNP genotyping to identify disease casual mutations.

2485W

***De novo* deletion of chromosome 19p13.3 involving *STK11* in a girl with intellectual disabilities and dysmorphic features without findings of Peutz-Jeghers syndrome: Further delineation of a contiguous gene deletion syndrome.** *C. Long, R. Olson, G. Scharer.* Division of Genetics, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN.

Peutz-Jeghers syndrome (PJS) is a rare hereditary cancer predisposition syndrome characterized by gastrointestinal polyposis and mucocutaneous pigmentation. Onset of characteristic hamartomatous polyps is often within the first 10 years of life, and there is an increased lifetime risk for gastrointestinal and extraintestinal malignancies. The majority of individuals with PJS have a mutation identified in the *STK11* gene. Microdeletions of 19p13.3 involving the *STK11* gene have been previously reported with the majority of individuals (4/6) presenting with either mucocutaneous pigmentation or GI polyps. We present a patient with a *de novo* 740 kb deletion of 19p13.3 involving the *STK11* gene (chr 19: 183996-1223710 / hg18). At 17 years of age, this individual did not have a history of mucocutaneous pigmentation or any identified polyps. Her clinical features included left Duane syndrome, bilateral ear keloid scarring, bifid uvula, scoliosis, developmental delays, mild intellectual disability, and dysmorphic features. Many of her features overlap with the features seen in previously reported individuals with 19p13.3 deletions. The deletion identified in our patient partially overlaps with the previously published reports of 19p13.3 microdeletions involving the *STK11* gene and allows for further characterization of the associated phenotype. Over 40 OMIM genes are encompassed by this deletion, including *STK11*, *ELANE*, *KISS1R*, *ARID3A*, and *GRIN3B*. *ARID3A* is a member of the ARID family of DNA binding proteins; other members of this family, specifically *ARID1A* and *ARID1B*, are associated with Coffin-Siris syndrome (MIM #603024, #614556), a condition characterized by intellectual disabilities and dysmorphic features. Deletion of this gene may contribute to some of the clinical findings seen our patient. In light of the haploinsufficiency of the *STK11* gene identified for this patient, we have recommended age-related PJS screening per the NCCN guidelines. Additional investigation regarding this contiguous gene deletion syndrome is warranted to further delineate the variable findings and the associated lifetime cancer risks.

2486T

Mapping autosomal recessive intellectual disability: Combined microarray and exome sequencing identifies 69 different genes in 192 consanguineous families. *J.B. Vincent[1], R. Harripaul[1], N. Vasli[1], A. Mikhailov[1], M.A. Rafiq[2], K. Mittal[1], C. Windpassinger[3], T.I. Sheikh[1], A. Noor[4], H. Mahmood[1], S. Downey[1], M. Johnson[1], K. Vleuten[1], M. Moradi[5], M. Ayaz[6], F. Naeem[7], A. Heidari[8], I. Ahmed[9], S. Ghadami[10], Z. Agha[2], S. Zeinali[10], R. Qamar[2], H. Mozhdehipanah[5], P. John[9], A. Mir[11], M. Ansar[12], L. French[1], M. Ayub[7].* 1) Neurogenetics Sect, R30, Ctr Add/Mental Hlth,Clarke Div, Toronto, ON, Canada; 2) COMSATS Institute of Information Technology, Islamabad, Pakistan;; 3) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 4) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada;; 5) Qazvin University of Medical Science, Qazvin, Iran;; 6) Lahore Institute of Research & Development, Lahore, Pakistan;; 7) Department of Psychiatry, Queen's University, Kingston, ON, Canada;; 8) Division of Hematology/Oncology, Hospital for Sick Children, Toronto, ON, Canada;; 9) Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan;; 10) Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran;; 11) Human Molecular Genetics Lab, Department of Bioinformatics and Biotechnology, FBAS, International Islamic University, Islamabad, Pakistan;; 12) Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan.

About 1% of children worldwide are affected by intellectual disability (ID). The vast majority with ID currently receive no molecular diagnosis. Knowing the causative gene may have huge benefits for diagnosis, care, overall health and lifespan, and family counseling and planning. Understanding the biological pathways that, when disrupted, may lead to ID, will provide a major step towards identifying therapeutic opportunities. Studies indicate high levels of genetic heterogeneity, with estimates of over 2500 autosomal ID genes, with the majority being recessive. Here, we present findings for a gene mapping/mutation identification project using SNP microarray analysis combined with whole exome sequencing, for 192 Pakistani and Iranian consanguineous, mainly non-syndromic ARID families with multiple affected members. We identified definite or likely causative mutations or CNVs in ~48% of families, including 26 genes not previously reported and 11 genes first reported from this cohort. The 26 new genes include 12 with homozygous loss-of-function mutations (*IMPA1, TMEM135, TRAPPC6B, USP44, MBOAT7, TBC1D23, SLAIN1, CCDC82, MAPK8, MPDZ, PIDD1,* and *UBA7*). The high yield of autosomal recessive mutations for ID signals that this approach not only has excellent clinical potential for consanguineous populations, but will also inform clinical diagnostics, including clinical whole exome and genome sequencing, for outbred populations.

2487F

Coffin-Lowry syndrome: Identification of a novel variant. A case report. *A.I Sanchez[1], J.C Prieto[1,2], P. Sanchez[2].* 1) Instituto de genética humana, Universidad Javeriana, Bogota, Bogota, Colombia; 2) Hospital La victoria SDS, Bogota, Colombia.

Introduction: Coffin-Lowry syndrome (CLS) is a rare X-linked dominant disorder caused by mutations in the RPS6KA3 gene, which encodes a growth factor-regulated serine threonine kinase. CLS affects 1 in 100,000 individuals, being 70-80% of them sporadic cases. It is characterized by severe to profound intellectual disability, global developmental delay, dysmorphic facial features, short stature and skeletal malformations. **Case presentation**: This is a 6 y/o male patient who is the 7th child of a healthy non-consanguineous couple. Family history was remarkable for two sisters and one brother having short stature and intellectual disability. He was born by uncomplicated spontaneous vaginal delivery at home after a full non-controlled term gestation. Birth weight at 38 weeks was 2380 grs; he was treated with crystalline penicillin G because of highly suspicious congenital syphilis and placed under phototherapy because of hiperbilirrubinaemia. Patient was hypotonic since birth and had history of global development delay with hampered verbal communication. Physical examination revealed microcephaly, large forehead, ocular hypertelorism, downslanting palpebral fissures, long philtrum, anteverted nares, thick lips, mild facial coarsening, broad tapering fingers and short stature. Hearing tests, thyroid function tests and MRI of the brain showed no abnormalities, as well as karyotype and microarray-based comparative genomic hybridization. Whole genome sequence analysis on DNA extracted from patient was performed. A novel pathogenic hemizigous mutation in exon 13 in the RPS6KA3 gene was detected. **Discussion**: RPS6KA3 gene is located at Xp22.2 and encodes for the RSK2 protein, which is expressed in developing central nervous system and is known to be involved in neuronal survival, neurite growth, functional maturation, and synaptic plasticity. The variant c.1004T>C (p.L335P) was identified; it was neither found in ExAC nor 1000G. Therefore we propose a novel mutation (single base exchange) which has had result in loss of function of the protein. This case is also noteworthy because patient's sisters had history of intellectual disability and short stature **Conclusions**: To our knowledge, this is the first time this variant is reported in a boy with a classic phenotype of CLS. To date, more than 140 RSK2 mutations have been identified in CLS patients. Further studies are necessary to elucidate the whole correlation between genotype and possible familial cases.

2488W

GRIN2B **encephalopathy – Novel phenotypic, genetic and functional aspects.** *J.R. Lemke[1], K. Platzer[2], H. Yuan[2], W. Chen[2], C. Hu[2], H. Kusomoto[3], H. Schütz[3], B. Laube[1], H.O. Heyne[4], K.L. Helbig[5], N. DiDonato[6], W.B. Dobyns[2], S.F. Traynelis[1], GRIN2B study group.* 1) University of Leipzig, Leipzig, Germany; 2) Emory University School of Medicine, Atlanta, GA, USA; 3) Technical University Darmstadt, Darmstadt, Germany; 4) Ambry Genetics, Aliso Viejo, California, USA; 5) Technical University Dresden, Dresden, Germany; 6) University of Washington, Seattle, Washington, USA.

   Objective: *GRIN2B* encoding the NMDA receptor subunit GluN2B has been associated with a variety of neurodevelopmental disorders. We aimed for a comprehensive delineation of the genotypic and phenotypic spectra, functional characterization of variants as well as exploration of potential prospects of personalized medicine. **Methods:** Molecular and clinical data of 84 individuals with *de novo GRIN2B* variants were collected from several diagnostic and research cohorts as well as from the literature. Functional consequences and response to the use-dependent partial NMDA receptor blocker memantine were investigated in *Xenopus laevis* oocytes and eventually translated into clinics. **Results:** We report 40 novel patients with *de novo GRIN2B* variants and review all 44 previously reported individuals. In 81 cases, variants were classified as pathogenic/likely pathogenic and associated with developmental delay, intellectual disability or autism spectrum disorder as well as muscular hypotonia, movement disorder, cortical visual impairment, cerebral volume loss and epilepsy. Four patients presented with an unusual malformation of cortical development (MCD). Missense variants cluster in transmembrane segments and ligand binding sites. Functional consequences were diverse, revealing various loss- and gain-of-function mechanisms, the latter partially retaining sensitivity to memantine. However, treatment responses in the respective patients still remain to prove significant benefits. **Conclusions:** In addition to previously known features of intellectual disability, epilepsy and autism spectrum disorder, we show that *GRIN2B* encephalopathy is also frequently associated with movement disorder, cortical visual impairment and MCD revealing a novel disease mechanism of channelopathies. The functional characterizations of variants enabled personalized therapeutic approaches which still remain challenging when translated from *in vitro* experiments into patient treatment.

2489T

Whole exome sequence identified disease-causing CNVs of *VPS13B* **responsible for Cohen syndrome in siblings with non-syndromic intellectual disability.** *Y. Enomoto[1], Y. Tsurusaki[1], T. Yokoi[2], M. Minatogawa[2], C. Abe-Hatano[2], K. Ida[2], T. Naruto[1], J. Mitsui[3], S. Tsuji[3], S. Morishita[4], K. Kurosawa[1].* 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Neurology, University of Tokyo Graduate School of Medicine, Tokyo, Japan; 4) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan.

   Cohen syndrome is rare autosomal recessive disease and characterized by persisting hypotonia, truncal obesity, visual impairment, short stature, intellectual disability(ID) and dysmorphic face. Causative gene is *VPS13B*. More than 150 cases with Cohen syndrome have been reported. Most of the cases are from Finnish population. The clinical phenotype is recognizable and homogenous for Finnish cases, but variable for other ethnic origin. Recently, disease-causing variants of *VPS13B* have been identified in the patients with non-syndromic ID by whole exome sequence (WES). We report on Japanese siblings with non-syndromic ID with compound heterozygous CNVs identified by normalized depth calculation using WES data. Older sister was born at 39 weeks of gestation. Birth weight was 2816g (-0.4SD), length 51cm (1.0SD), and OFC 34cm (0.5SD). She was able to speak one word at 1 year and 3 months and walk at 2 years and 6 months. At age of 6 years, she was able to walk very slowly and speak using 2-3 word sentences. She had truncal obesity, hypotonia, myopia, and mild to moderate ID, but did not have microcephaly and short stature. Younger sister was born at 39 weeks of gestation. Birth weight was 2500g (-1.3SD), length 46cm (-1.1SD), and OFC 32cm (-0.6SD). At 2 years and 8 months, her weight was 11.4kg (-0.8SD), length 83.7cm (-1.8SD) and OFC 44cm (-2.7SD). She was not able to walk and speak a comprehensive word. CGH array (8x60k) using standard filter setting which detects signal changes of 3 or more probes could not detect pathogenic CNV. However, normalized depth calculation using the WES data could detect pathogenic compound heterozygous CNVs in *VPS13B* common to the siblings. The two pathogenic CNVs included a deletion of exon8-15 derived from their mother and a deletion of exon 32-33 derived from their father. These results support the idea that the phonotype of Cohen syndrome is variable in individuals of non-Finnish populations and the normalized depth calculation using WES data is highly valuable for detecting the minor CNVs in diagnosis of non-syndromic ID.

2490F

Whole exome sequencing reveals a mutation in *ARMC9* as a cause of mental retardation, ptosis and polydactyly syndrome. *A. Kar[1,2], S. Phadke[3], A. Dalal[1,4].* 1) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 2) Graduate Studies, Manipal University, Manipal, India; 3) Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India; 4) Department of Medical Genetics, Kasturba Medical College, Manipal University, Manipal, India.

Intellectual disability has a prevalence of 1 -3% population and results from heterogeneous causes like environmental effect, chromosomal or monogenic causes. More than 230 genes have been reported to be involved in causation of syndromes of intellectual disability. We report on exome sequencing in three siblings in a consanguineous family with short stature, mental retardation, downslanting palpebral fissures, ptosis and polydactyly. Homozygosity mapping followed by whole exome sequencing was done in proband and two affected siblings after obtaining informed consent. Homozygosity mapping revealed five common regions of loss of heterozygosity region constituting about 15.69Mb and harbored 228 genes. Next generation sequencing reads were processed for mapping to reference genome, variant calling and annotation. Variants were filtered for polymorphisms using various databases. Comparison of variants among affected siblings revealed homozygous mutation in Armadillo repeat containing 9 (*ARMC9*) gene in all three. Mutation c.1073G>A in *ARMC9* is a synonymous splice site mutation, which is predicted to be 'disease causing' by various splice site mutation prediction software. Mutation was confirmed by Sanger sequencing in parents, probands, affected and unaffected siblings. *ARMC9* (KU-MEL-1) is known as melanoma/melanocyte antigen and is reported to be expressed in melanoma and melanocytes. It contains LISH1 domain, which is predicted to have involvement in microtubule dynamics and ARM domain, which exhibits binding with proteins due to presence of extensive solvent accessible surface. *ARMC9* is also involved in interaction with *SIAH1* (E3 ubiquitin ligase). Mutation c.1073G>A, which is a splice site mutation may alter downstream sequences and thus may abolish ubiquitination site at Lys441. We propose that mutation in *ARMC9* leads to mental retardation, ptosis and polydactyly syndrome in this family and *ARMC9* may be a novel gene involved in causation of syndromic intellectual disability. Functional studies are ongoing for validation of the results.

2491W

Whole exome sequencing reveals a missense mutation in the *ARX* gene in a family with X-linked non-syndromic mental retardation. *M.B. Petersen[1,2], I.S. Pedersen[3], V.Q. Le[3], A. Ernst[3], H. Krarup[3].* 1) Dept Clinical Genetics, Aalborg University Hosp, Aalborg, Denmark; 2) Dept Clinical Medicine, Aalborg University, Aalborg, Denmark; 3) Section of Molecular Diagnostics, Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark.

The *ARX* (Aristaless-related homeobox) gene belongs to a large family of homeodomain transcription factors with critical roles in development. Mutations in the *ARX* gene cause X-linked intellectual disability. The phenotypic spectrum comprises a series of X-linked developmental disorders ranging from lissencephaly to agenesis of the corpus callosum with abnormal genitalia to infantile spasms without brain malformations to syndromic and non-syndromic mental retardation. We describe a five-generation Danish family with three males at the age of 18, 44, and 61 years, respectively, showing mental retardation of unknown reason. Except from large ears, the clinical evaluation did not reveal any distinct dysmorphic feature. Array-CGH and FMR1 analysis were normal in two of the males. According to the pedigree, an X-linked inheritance pattern was strongly suspected. Whole exome sequencing (WES) with targeted analysis of the coding regions of the X chromosome was carried out in two of the affected males, a suspected obligate carrier female (mother of an affected), and a healthy male (brother of an affected). A filtering process searching for nonsynonymous variants and variants in the exon-intron boundaries revealed a novel variant, NM_139058.2:c.1A>G, in the first codon of the *ARX* gene. Disruption of an initiator codon may lead to reinitiation if an alternative initiator codon is present. Recently, premature termination mutations very early in the *ARX* gene have been reported to lead to reinitiation of translation to produce a truncated protein at markedly reduced level. These mutations are associated with a less severe phenotype than truncating mutations located further downstream in *ARX*. The finding of a mutation in the first codon of the *ARX* gene in a family segregating non-syndromic mental retardation supports the observed genotype-phenotype correlation reported for *ARX* mutations.

2492T

Heterozygous variants in *ACTL6A*, encoding a component of the SWI/SNF complex, are associated with intellectual disability. *R. Marom[1], M. Jain[1], L.C. Burrage[1], B.H. Graham[1], C.W. Brown[2], B.B. de Vries[3], S. Stevens[4], A.T. Gunter[5], J.D. Kaplan[5], R.H. Gavrilova[6], M. Shinawi[7], J.A. Rosenfeld[1], J.T. Lu[10], R.A. Gibbs[1,8], C.M. Eng[1], Y. Yang[1], J. Rousseau[9], P.M. Campeau[9], B.H. Lee[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX. USA; 2) Department of Pediatrics/ Genetics Division, University of Tennessee Health Science Center Memphis, TN. USA; 3) Department of Human Genetics and Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands; 4) Department of Human Genetics, Maastricht University Hospital, Maastricht, The Netherlands; 5) Department of Pediatrics, Division of Medical Genetics, University of Mississippi Medical Center, Jackson, MS. USA; 6) Department of Clinical Genomics, Mayo Clinic, Rochester, MN. USA; Department of Neurology, Mayo Clinic, Rochester, MN. USA; 7) Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, MO. USA; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX. USA; 9) Department of Pediatrics, CHU Ste-Justine and University of Montreal, Montreal Canada; 10) Helix, One Circle Star Way Bldg 1, San Carlos CA. USA.

Pathogenic variants in genes encoding components of the SWI/SNF complex and related interacting chromatin remodeling proteins have been associated with intellectual disability syndromes. By whole exome sequencing, we identified heterozygous, rare variants in *ACTL6A* (Actin-like 6A), a gene encoding a component of the SWI/SNF complex, in four patients with speech delay and varying degrees of intellectual disability. Other phenotypic features shared by some of our patients include a history of failure to thrive, congenital heart defect and finger and toe abnormalities. Three patients were found to have missense variants in highly conserved amino acid residues – a de novo p.R377W variant (exon 13), a maternally inherited p.R389W variant (exon 13), and p.E227Q (exon 8) with an unknown inheritance (due to the unavailability of parental testing). These three amino acids reside in a region with structural similarity to actin. Missense mutations in the homologous region in yeast actin were found to be dominant lethal (Wertman KF *et al*, Genetics 1992). Interestingly, a previously published study in cells transiently-transfected with *ACTL6A* showed that introducing mutations into these same three residues results in impaired binding of the ACTL6A protein to β-actin and BRG1 (Nishimoto N *et al*, J Cell Science 2012). These findings suggest that the variants identified in our patients may have a deleterious effect on the function of the protein, by disturbing the integrity of the SWI/SNF complex. A fourth patient was found to have a de novo splicing variant in intron 13, and RNA studies using a lymphoblastoid cell line derived from this patient showed abnormal splicing that creates an in-frame deletion of exon 13. Thus, three of the four variants involve exon 13 which is a highly evolutionarily conserved region of the protein. Whole exome sequencing did not detect other variants in genes known to cause intellectual disability that may explain the phenotype seen in these patients. *ACTL6A* is a member of the BAF (Brg/Brahama associated factors) complex that plays an essential role in neurodevelopment. To the best of our knowledge, *ACTL6A* has not been associated, thus far, with intellectual disability syndromes. Thus, heterozygous pathogenic variants in *ACTL6A* gene should be considered in patients with intellectual disability, learning disabilities or developmental language disorder.

2493F

Exome sequencing of 152 consanguineous families with intellectual disability reveals high diagnostic yield and identification of novel candidate genes. *R. Abou Jamra[1,2], M. Reuter[2], H. Tawamie[1,2], R. Buchert[2], T. Froukh[3], O. Hosny Gebril[4], T. Strom[5], S. Hoffjan[6], D. Wieczorek[7], H. Sticht[8], A. Reis[2].* 1) Institute of Human Genetics, University Medical Center Leipzig, Leipzig, Sachsonia, Germany; 2) Institute of Human Genetics, FAU Erlangen-Nürnberg, Erlangen, Germany; 3) Department of Biotechnology and Genetic Engineering, Philadelphia University, Amman, Jordan; 4) National Research Center, Doky, Cairo, Egypt; 5) Helmholtz Zentrum München, Neuherberg, Germany; 6) Institute of Human Genetics, University of Bochum, Bochum, Germany; 7) Institute of Human Genetics, University Duisburg-Essen, Essen, Germany; 8) Institute of Biochemistry, FAU Erlangen-Nürnberg, Erlangen, Germany.

Autosomal recessive inherited neurodevelopmental disorders are highly heterogeneous and many disease-associated genes are still unknown. In this study, we aim at identification of disease genes through confirmation of previously described and presentation of novel candidates, and to assess the diagnostic yield of exome sequencing in consanguineous families. We enrolled 152 consanguineous families with one (29%) or multiple (71%) affected children with intellectual disability. Most affected persons showed additional symptoms including muscular hypotonia, epilepsy and microcephaly. After autozygosity mapping in families and exome sequencing of the index we identified candidate variants in 105 genes. We could establish a genetic diagnosis in 53 families (35%) due to 54 (likely) pathogenic variants in genes already associated with neurodevelopmental disorders. Only four genes were hit recurrently (*AHI1, GRP56, PLA2G6, PRRT2*). In five of these families potentially treatable disorders were diagnosed, and in at least one patient we identified two homozygous variants in different genes leading to a combined dual diagnosis. In addition to recessive variants we also identified pathogenic X-linked variants in boys (*PIGA, SLC6A8*) and dominant de novo variants in known ID genes (*DYRK1A, KMT2B*) in singletons after performing trio analysis as a proof of principle. In the 52 families (34%), we identified convincing variants in candidate genes not previously associated with neurodevelopmental disorders. Of these, 16 were homozygous and truncating in *GRM7, STX1A, KIAA1967, EEF1D, METTL5, GALNT2, SLC44A1, LRRIQ3, MBOAT7, AMZ2, CLMN, SEC23IP, MBNL3, INIP, NARG2, KIAA1467*, and *TRAP1*. We also identified convincing, non-truncating variants in *RXRB, BDH1, C9ORF114, GRAMD1B, EZR, OGDHL, ADIPOR1, EDC3, CEP76, KCTD18, PPFIA1, HMG20A, SKIDA1 (C10orf140), EIF4A2, MAGI2, THG1L, LRCH3, TMTC3, HACL1, TSPAN18, FBXO11, SV2C, LENG8, CACNA2D1, ATP2C2, GCC2, TMEM132D, PTRHD1, KIAA0195, PPRC1, SMURF2, ENO2, NCAPD2, BTN2A2, GTF3C3, CHD1L*. We are aware that some of these could be false positive; however, we obtained supporting functional results and/or further independent families for ten of them. The diagnostic yield was higher in severe cases, in multiplex families, and in cases with detailed clinical description. We further confirm the utility of exome sequencing as first line diagnostic approach and propose 52 convincing candidate genes awaiting confirmation by independent families.

2494W

The TIDE protocol identifies treatable inborn errors of metabolism in 5% of children with unexplained intellectual developmental disorder. *G. Aisha[1], R. Salvarinova[2], G. Horvath[2], M. Connolly[2], M. Demos[2], P. Eydoux[3], H. Vallance[3], G. Sinclair[3], S. Stockler[1], C. van Karnebeek[1,4], TIDE-BC.* 1) BC Children's Hospital, Vancouver, British Columbia, Canada; 2) Dept of Pediatrics, University of British Columbia, Vancouver, Cada; 3) Dept of Pathology and Laboratory Medicine, University Of British Columbia; 4) Centre for Molecular Medicine and Therapeutics, CFRI.

   BACKGROUND: Intellectual developmental disorders (IDD) are characterized by significant impairment of cognitive functions, affecting 2.5% of the population worldwide with significant morbidity and associated healthcare costs. Inborn errors of metabolism (IEM) currently constitute the only group of genetic defects amenable to causal therapy. Early diagnosis prevents or minimizes brain damage. Our literature review identified 89 such treatable IEM; although evidence is limited, therapies are often effective, safe, accessible. **METHODS:** We translated this knowledge into the published TIDE *diagnostic protocol:* The 1st tier comprises metabolic screening tests in blood/urine with potential to identify 62% of treatable IDs. The *second tier* focuses on remaining disorders, requiring 'single test per disease' approach. A freely available App (www.treatable-id.org) supports the protocol. The protocol was superimposed to current diagnostic guidelines in the evaluation of unexplained IDD during 3 years in our 3 divisions in our tertiary care centre. **RESULTS:** In total, we established a causal diagnosis in 53% of 522 IDD patients (incl chromosomal copy number variants, syndromic diagnoses etc). Treatable IEMs were identified in 5% (n=27), including creatine deficiencies, amino-acidopathies, serine deficiencies, metal disorders, vitamin responsive disorders, neurotransmitter diseases, lysosomal storage disorders, organic acidurias etc. Analysis comparing these patients to those diagnosed in our hospital between 2000-2009 revealed that the TIDE protocol reduced 'time to diagnosis' by 6 months (range 1-50months) as well as costs of unnecessary testing (>$1500 per patient). 25% of patients with a treatable IEM presented with isolated stable ID; frequent co-morbidities include autism and seizures. 1st tier tests identified 55% of IEMs; of the 1st and 2nd tier testing, urine creatine metabolites and CSF neurotransmitter analyses identified highest number of IEMs. **DISCUSSION**: Our protocol for treatable forms of ID has proven effective in terms of increasing the diagnostic yield and reducing costs and diagnostic delay. Treatment effects vary from improvement of cognitive development, behavior, epilepsy, psychiatric disturbances or stabilization of disease. Overall better outcomes can be achieved via standard screening for treatable conditions in IDD patients. Coverage of the genes encoding treatable IEMs should be optimized if exome / genome sequencing is performed.

2495T

Genomic variants in *LINGO1* likely cause autosomal recessive intellectual disability and developmental delay. *M. Ansar[1], S. Riazuddin[2], P. Makrythanasis[1,3], M.T. Sarwar[4], Z. Iqbal[5,6], S.A. Paracha[4], J. Khan[4], M. Hussain[5,7], A. Razzaq[5,7], D.L. Polla[5], A. Suhail[4], A. Holmgren[8], D. Misceo[8], A.P.M. de Brouwer[5], M. Guipponi[1,3], F.A. Santoni[1,3], E. Frengen[8], J. Ahmed[4], H. van Bokhoven[5], S. Riazuddin[9,10], S.E. Antonarakis[1,3,11].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Department of Otorhinolaryngology-Head & Neck Surgery, School of Medicine, University of Maryland, Baltimore, MD, 21201, USA; 3) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 4) Institute of Basic Medical Sciences (IBMS), Khyber Medical University, Peshawar, Pakistan; 5) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; 6) Department of Neurology, Oslo University Hospital, Oslo, Norway; 7) National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan; 8) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 9) Allama Iqbal Medical College, University of Health Sciences, Lahore, Pakistan; 10) Shaheed Zulfiqar Ali Bhutto Medical University, Pakistan Institute of Medical Sciences, Islamabad, Pakistan; 11) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.

   LINGO1, a transmembrane receptor, is predominantly expressed in the central nervous system and is involved in the inhibition of neuronal myelination, through activation of the NgR1 signaling pathway. Various GWAS and expression studies have implicated increased expression of this gene with Multiple Sclerosis, Essential Tremor and Parkinson's disease. We have studied two unrelated Pakistani consanguineous families from different areas of Pakistan. All patients presented with a similar phenotype consisting of severe intellectual disability, aggressive behavior, speech and motor delay. We performed exome sequencing in order to identify the potential causal variant. We have identified a homozygous missense variant (NM_032808.6:c.869G>A:p.(Arg290His)) in family one and different homozygous variant in family two (NM_032808.6:c.863A>G p.(Tyr288Cys)) in *LINGO1*. Both variants segregated with the phenotype in the respective families and were predicted to be pathogenic by SIFT, PolyPhen and MutationTaster. NgR1 signaling pathway mediated by LINGO1-NgR1 complex, negatively regulates oligodendrocyte differentiation and neuronal survival. Previously mouse and zebrafish models have been described and documented the role of LINGO1 in abnormal neuronal development and central nervous system myelination. Taken together, our results indicate that *LINGO1* missense variants likely cause autosomal recessive intellectual disability. However, further functional studies are warranted to dissect the exact pathophysiological mechanism of the identified variants.

2496F

Bi-allelic *SPATA5* mutations cause severe intellectual disability, hypotonia and hearing impairment. *R. Buchert[1,2], A.I. Nesbitt[3], H. Tawamie[2], I.D. Krantz[4], L. Medne[5], I. Helbig[6,7], D.R. Matalon[4], A. Reis[2], A. Santani[3,8], H. Sticht[9], R. Abou Jamra[2,10].* 1) Institute of Medical Genetics and Applied Genomics, University Tübingen, Tübingen, Germany; 2) Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 3) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, USA; 4) Division of Human Genetics, The Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA; 5) Division of Child Neurology, Departments of Pediatrics and Neurology, Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA; 6) Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, USA; 7) Department of Neuropediatrics, University Medical Center Schleswig-Holstein, Kiel Campus, Kiel, Germany; 8) Department of Pathology & Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA; 9) Institute of Biochemistry, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 10) Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany.

   Intellectual disability is a genetically heterogeneous disorder which is characterized by significant limitations in both intellectual functioning and in adaptive behavior. In a study comprising 152 families with intellectual disability of probably autosomal recessive inheritance we examined an extended consanguineous Syrian family with seven individuals affected with severe intellectual disability, microcephaly, muscular hypotonia and hearing impairment. Autozygosity mapping followed by whole exome sequencing identified a homozygous, segregating variant in *SPATA5* (c.1822_1824del; p.Asp608del). In an unrelated family, we identified compound heterozygous variants in *SPATA5* (c.[2081G>A];[989_991delCAA]; p.[Gly694Glu];[.Thr330del]) in a further individual with global developmental delay, infantile spasms, profound dystonia, and sensorineural hearing loss. Molecular modeling showed that the aspartate at position 608 (p.Asp608) is situated at the subunit interface and forms stabilizing interactions with a lysine at position 517 (p.Lys517) of the adjacent subunit of homohexameric SPATA5. Thus, a deletion of p.Asp608 is expected to cause a loss of this salt bridge and reduce the stability of the hexameric structure as well as the stability of the subunit itself. Similar to p.Asp608, p.Gly694 is also located at the subunit of the interfaces in the hexameric enzyme. A replacement of p.Gly694 by glutamate causes electrostatic repulsion with p.Asp630 of the adjacent protein subunit. Therefore, this alteration is expected to decrease hexamer stability in a similar fashion as p.Asp608del. The alteration p.Thr330del has been previously reported and could not be modeled due to a lack of sequence similarity to known protein structures. Recently, bi-allelic variants in *SPATA5* have been reported in individuals with intellectual disability, microcephaly, epilepsy and hearing loss. The identification of further individuals with a comparable phenotype and bi-allelic mutations in *SPATA5* confirms *SPATA5* as a new candidate gene for intellectual disability and further delineates the associated phenotype.

2497W

***NR2F1* mutations: Human phenotypes and investigation of a mouse model.** *C. Chen[1,2], R. Pautler[1], S. Pedersen[1], A. Raman[1,2], L. Wang[1,2], M. van der Heijden[1,2], J. Yin[1,2], Z. Liu[1,2], C. Schaaf[1,2].* 1) Baylor College of Medicine, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX.

   Bosch-Boonstra-Schaaf optic atrophy syndrome (BBSOAS) is an autosomal dominant disorder characterized by optic atrophy and intellectual disability, caused by mutations in *NR2F1*. Other clinical phenotypes include hypotonia, seizures, autism spectrum disorder, oromotor dysfunction, thinning of the corpus callosum, and hearing defects. Structural abnormalities in the brain, such as hippocampal malrotation, occipital stenogyria, and chiasmal misrouting have also been reported. To better understand the disease mechanism, we studied *Nr2f1* heterozygous knockout mouse (*Nr2f1+/-*) as a disease model, given that all human patients identified to date are heterozygous mutation carriers. *Nr2f1+/-* mice manifest structural abnormalities in the brain based on DTI (diffusion tensor imaging). *Nr2f1+/-* mice have decreased hippocampal volume and increased fiber number in entorhinal cortex, however the corpus callosum appears structurally normal. Both hippocampus and entorhinal cortex are known to be critical for memory formation and maintenance. Behavioral studies show that *Nr2f1+/-* mice manifest prolonged retention of fear memory, indicating altered memory extinction. Additionally, neonatal hypotonia was noticed in p6 *Nr2f1+/-*. Although most BBSOAS patients have optic nerve atrophy, *Nr2f1+/-* mice manifest a visual performance that is not significant from wild type littermates (optokinetic reflex test). Given the reduced hippocampal volume and learning/memory abnormality in *Nr2f1+/-* mice, RNA-Seq was performed to compare the differential gene expression between wild type and *Nr2f1+/-* hippocampus. Pathway enrichment analysis showed an up-regulation of matrix metalloproteases (Mmp2, Mmp14, and Mmp15) in *Nr2f1+/-* hippocampus. The up-regulation of Mmp2 in *Nr2f1+/-* hippocampus was further validated by quantitative PCR and western blot analysis. Mmp2 is known to be critical for neuronal wiring, dendritic outgrowth, and spine remodeling. Primary cortical neuronal culture from p0 *Nr2f1+/-* pups shows decreased density of dendritic spine, particularly the stubby-shaped spines and mushroom-shaped spines, suggesting that Nr2f1 is important during the development of the central nervous system. In conclusion, we delineate the clinical spectrum of BBSOAS and characterize *Nr2f1+/-* mice as a potential disease model. The brain structural abnormalities, learning/memory deficit, and dendritic spine abnormalities in *Nr2f1+/-* mice provide insight into the disease mechanisms of BBSOAS.

2498T

Multiple human neuronal models of Lesch-Nyhan disease reveal adenosinergic and metabolic dysfunctions. *L. Crapper[1,2], S. Bell[1,2], H. Peng[1], H. Jinnah[4], J. Visser[5], G. Turecki[1,3], C. Ernst[1,3].* 1) McGill Group for Suicide Studies, McGill, Montreal, Quebec, Canada; 2) Neuroscience, McGill, Montreal, Quebec, Canada; 3) Psychiatry, McGill, Montreal, Quebec, Canada; 4) Neurology, Emory University, Atlanta, Georgia; 5) Neurology, Radboud University, Nijmegen, Netherlands.

   Background: Lesch-Nyhan Disease (LND) is a severe neurodevelopmental disorder caused by mutations in the gene *HPRT1*, characterized by hyperuricemia, dystonia, intellectual disability, and compulsive self-injurious behaviour. Although the disease was first identified in 1963, there is still no treatment for the neurological symptoms, nor has a functional association been established between *HPRT1* mutations and self-injury. Patients with LND have very low levels of dopamine, and it is believed that this contributes to the neurological symptoms, but again the connection of *HPRT1* to dopamine is unknown. **Objectives**: To develop novel *in vitro* models of LND, and determine the molecular impacts of HPRT deficiency in disease relevant cell types. **Methods & Results:** HPRT was reduced in immortalized midbrain neural progenitors using an shRNA mediated knockdown. RNA sequencing in these cell lines revealed large deficits in many genes important to adenosine signalling, and increases in genes related to energy metabolism and protein translation. We then generated induced pluripotent stem cells (iPSCs) from fibroblasts collected from LND patients and controls, which were used to generate forebrain and midbrain neural progenitor cells. These Patient derived neuronal cells showed similar transcriptional changes to the knockdown cell lines, and exhibited related functional changes. Finally, we performed RNA sequencing in the striatum and VTA of HPRT knockout mice, which do not show any of the behavioural deficits associated with LND. HPRT deficient mice do not show any major transcriptional alterations in the adenosine or energy metabolism pathways, and show opposite changes in protein translation. **Conclusions:** Two independent human derived neuronal models of LND reveal cell autonomous phenotypes that are not present in animal models, and may help to explain their lack of phenotype. Together these data present strong evidence for a role of energy metabolism, and adenosine signalling in Lesch-Nyhan Disease.

2499F

Multifaceted approach for understanding phenotypic variability in subjects with 16p11.2 duplication. *J. Dastan[1], C. Chijiwa[2], F. Tang[3], S. Martell[3], Y. Qiao[3], E. Rajcan-Separovic[3,4], M.E.S. Lewis[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 2) Children's and Women's Health Center of BC, University of British Columbia, Vancouver, Canada; 3) BC Child and Family Research Institute, University of British Columbia, Vancouver, Canada; 4) Department of Pathology, University of British Columbia, Vancouver, Canada.

   Background: Microduplication of 16p11.2 (dup16p11.2) is associated with a broad spectrum of neurodevelopmental disorders (NDD) confounded by variable expressivity. We hypothesized that while some unique features reported in individuals with dup16p11.2 may be explained by the over-expression of its integral genes, co-occurrence of other genetic alterations in the genome may account for the variability in their clinical phenotypes. This hypothesis was explored in two unrelated subjects with NDD who each inherited the dup16p11.2 from an apparently healthy carrier parent. **Results:** Detailed phenotyping analysis was used to find consistent and recognizable features among carriers of dup16p11.2 (published and current study); yet we did not find such evidence among these individuals. *KTCD13* has been identified as a major driver of a mirror brain phenotype of 16p11.2 CNV in zebrafish. However, the data presented here demonstrated that dup16p11.2 did not result in increased protein expression of KCTD13 in either probands or the healthy carrier parent, indicating that *KCTD13* is not the sole cause of microcephaly in cases with dup16p11.2. Finally, whole exome sequencing (WES) was used to investigate the presence of genomic sequence changes in dup16p11.2 carriers that could explain such clinical variability. Compound heterozygous variants of *VPS13B* in proband A and missense variants of *SYNE2* in proband B were identified. Mutations of *VPS13B* cause Cohen syndrome in keeping with proband A's phenotype (ID, microcephaly, facial gestalt, retinal dystrophy, joint hypermobility and episodic neutropenia), mRNA sequence aberration and low RNA expression. The protein encoded by *SYNE2*, Nesprin 2, plays critical roles in neurogenesis in mice. Over-expression of Nesprin 2 identified in proband B and his mother, who did not carry the dup16p112, may partly explain NDD phenotypes segregating in this family via a dominant-negative effect. **Conclusion:** Our study showed that other pathogenic variants in the genome could contribute to inter- and inter- familial clinical variability among dup16p11.2 carriers. This suggests discordance in phenotype of dup16p11.2 carriers warrants further study by WES and individualized genetic assessment and counselling.

2500W

Diagnostic exome sequencing identifies a novel homozygous alteration in *DEAF1* further delineating the phenotypic spectrum. *C. Gund[1], W. Alcaraz[1], A. Sajan[1], Z. Powis[1], A. Vashist[2], P. Suwannarat[2].* 1) Ambry Genetics, Aliso Viejo , CA; 2) Mid-Atlantic Permanente Medical Group, Rockville, MD.

In the last three years, three individuals from two families have been reported with a homozygous *DEAF1* mutation (c.676C>T) who all shared a phenotype characterized by microcephaly, brain abnormalities, and intellectual disability. The affected individuals also demonstrated a variety of other features including hypotonia, feeding difficulties, and behavioral abnormalities. Additionally, an intronic mutation causing exon skipping and which generated a premature stop codon in *DEAF1* was reported to cause disease in three siblings all affected by intellectual disability, autism, dyskinesia, abnormal cranial MRI and epilepsy. Here we report an additional patient with homozygous alterations in *DEAF1*. The patient is a 4 month old Afghani girl with developmental delay, delayed visual maturation, strabismus, hydronephrosis, vesicoureteral reflux and microcephaly as well as dysmorphic features including low posterior hairline, mild synophrys, thick and bushy eyebrows, long eyelashes, downslanting palpebral fissures, low set ears slightly posteriorly rotated with prominent antihelix bilaterally and Darwinian tubercle on the right, micrognathia, depressed nasal bridge, broad nasal tip, bilateral inverted nipples, and overlapping toes. Her head circumference was 38.1 cm (11th percentile). A MRI showed absence of splenium, hypoplastic putamina and other basal ganglia abnormalities, and wide and underdeveloped sylvian fissures. The family history is significant for multiple instances of consanguinity and an older brother who died at 6 months who had an abnormal CT of the brain. Family centered diagnostic exome sequencing (DES) on the proband and her healthy parents revealed a homozygous c.576C>A (p.Y192*) nonsense alteration in *DEAF1*. This alteration was not reported in healthy populations. This patient provides another example of a syndrome associated with homozygous mutations in *DEAF1*. While the patient's phenotype contains elements of previously reported patients, including developmental delay, relative microcephaly, truncal hypotonia, and brain malformations on MRI, the patient is notable for having a number of dysmorphic features not previously reported in homozygous *DEAF1* patients. This could represent an expansion of the phenotypic spectrum or could be the result of another cause such as the effect of an unidentified pathogenic alteration. This is not direct endorsement of Ambry Genetics by Kaiser Permanente or the Mid-Atlantic Permanente Medical Group.

2501T

Refining genotype-phenotype correlations in Iranian patients with AP4 deficiency syndrome. *K. Kahrizi[1], T. Akhtarkhavari[1], Z. Fattahi[1], M. Hosseini[1], S. Abedini[1], M. Mohseni[1], H. Hu[2], L. Mutsant[2], T. Wienker[2], H. Najmabadi[1], H.H. Ropers[2,3].* 1) Genetics Research Center, Tehran, Iran; 2) Max Planck Institute for Molecular Genetics, Berlin, Germany; 3) Institute of Human Genetics, University Medicine, Mainz, Germany.

AP4 deficiency syndrome is characterized by severe intellectual disability, microcephaly, ventriculomegaly, growth retardation and spastic tetraplegia. Up to now, twenty six families have been reported for this syndrome. Since 2011, eleven AP4 deficiency families were identified through studying a large Iranian cohort in more than five hundred intellectually disabled families by conducting whole exome sequencing and for repetitive mutations in unrelated families we applied haplotype analysis. So far, we identified six novel mutations among nine families, including: g.1:114441425T>C (p.372Acceptor), g.1:114442649del (p.331fs) and g.1:114441377_114441378del (p.387fs) in *AP4B1*, g.7:99700490del (p.86fs) and g.7:99703887A>C (p.H333P) in *AP4M1* and g.14:31542174C>T (p.R97X) in *AP4S1*. The novel mutation in *AP4S1* was observed in two unrelated families living in two different cities with small geographic distance; moreover, g.7:99701748G>A (p.E193K) in *AP4M1* was also detected in three unrelated families from the same ethnicity. Haplotype analyses in these families confirmed their identity by descent and eliminated the possibility of a mutational hotspot. Apparently unrelated families with identical mutations reflect the tradition to marry within families in this country. *In this study we* compared phenotype-genotype profiles of eleven Iranian families with 25 patients with previously 17 reported families to show the phenotype similarity and differences in 4 subunits of AP4 complex.

2502F

High risk Northern Finnish population aids in discovery of rare recessive and dominant variants in neurodevelopmental disorders. *M.I. Kurki[1,2], E. Saarentaus[4], O. Pietilainen[2,3], E. Hamalainen[4], J.S. Moilanen[5], J. Körkkö[5], O. Kuismin[5], M. Daly[2,6], A. Palotie[1,2,4], Sequencing Initiative Suomi consortium.* 1) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 2) Stanley Center, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Department of Stem Cell and Regenerative Biology, University of Harvard, Cambridge, Massachusetts, USA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Department of Clinical Genetics, Oulu University Hospital, Medical Research Center Oulu and PEDEGO Research Group, University of Oulu, Oulu, Finland; 6) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

Genetic variants with strong reproductive disadvantages are evolutionary constrained and remain generally rare in population. However, these variants can still exist at higher frequencies in young populations, such as Finns, where the negative selection hasn't had time to counteract the effect of genetic drift on rare alleles. Thus, population isolates provide a valuable study design to explore the role of rare genetic variants in complex traits. In Finland, the youngest settlement is in the Northern and Eastern parts of the country, dating back to a small number of founder families only a few centuries ago. In addition, this region has a higher prevalence of both schizophrenia and intellectual disability (ID). We tested this hypothesis by producing whole exome sequence (WES) and GWAS data from 499 patients from Northern Finland with ICD-10 diagnosis of ID of unknown etiology, and their 429 family members (including 164 full trios). The Northern Finland Intellectual Disability Project (NFID) exomes were combined with 13,165 Finnish exomes sequenced in the Sequencing Initiative Suomi project (SISu, http://sisuproject.fi/). As expected, we observed comparable amount of large CNVs and *de novo* mutations as reported in similar patient collections, both of these categories being enriched in the NFID patients. Given the genetic origin of NFID, we expected to observe variants enriched in Finland that are 1) strong acting recessive variants, that seem Mendelian but account for ~1% of a 1% phenotype rather than all of a 1/10000 phenotype, and 2) dominant alleles, with odds ratios in the range of 2-5. As per our hypothesis, we discovered a Finnish-specific recessive cause of ID in 4 cases, homozygosity of a variant in *CRADD* (p=3.5e-07). The variant is not observed in homozygous state in 60,706 individuals worldwide (http://exac.broadinstitute.org/) or in 13,165 SISu. We also recruited multiplex families (2-5 affected) and searched for shared loss-of-function (LoF) variants in LoF intolerant genes (ExAC pLI metric). We observed LoF in 5 affected siblings and their mother in the *INTS6* gene. No *INTS6* LoF's have been observed in 60,706 ExAC individuals or in 13,165 Finnish exomes. We show that an enrichment of deleterious alleles increases power to detect causal and disease associated variants that would require very large sample sizes in more diverse populations.

2503W

The prevalence and architecture of dominant developmental disorders. *J. McRae, Deciphering Developmental Disorders Study.* Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

Individuals with severe, undiagnosed developmental disorders (DDs) are enriched for damaging *de novo* mutations in developmentally important genes. We exome sequenced 4,293 families with individuals with DDs, and meta-analysed these data with published data on 3,287 individuals with similar disorders. We identified 94 genes enriched for damaging *de novo* mutation at genome-wide significance (P < 7 x 10[-7]), including 14 genes for which compelling data for causation was previously lacking. We estimated that we have statistical power to detect ~45% of all haploinsufficient genes, and that most haploinsufficient genes causing DDs have already been discovered. Our large number of genome-wide significant findings allow us to compare empirically the power to detect novel DD-associated genes using exome or genome sequencing. We find that, at current cost differentials, exome sequencing has much greater power for novel gene discovery for genetically heterogeneous disorders. Finally, we estimate that 43% of our cohort carry pathogenic de novo mutations in coding sequences, with approximately half operating by a loss-of-function mechanism, and the remainder being gain-of-function or dominant negative in action. By extrapolating from the DDD cohort to the general population, we estimate that de novo dominant developmental disorders have an average birth prevalence of 1 in 213-448 (0.22-0.47% of live births), depending on parental age. Subsequent to these analyses in 4,293 families, we have sequenced another 3,663 families, nearly doubling our cohort to 7,956 families and will describe updated analyses within this larger dataset.

2504T

Hepatosplenomegaly as the initial manifestation of Coffin-Siris syndrome caused by an *ARID1B* **mutation.** *K. Takano[1], M. Motobayashi[2], T. Yamaguchi[1], K. Wakui[1], Y. Inaba[2], Y. Fukushima[1], T. Kosho[1].* 1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 2) Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan.

Coffin-Siris syndrome (CSS, MIM# 135900) is a rare genetic disorder characterized by developmental delay or intellectual disability (ID), distinctive facial features, hypertrichosis, aplasia or hypoplasia of the distal phalanx or nail of the fifth fingers, and feeding difficulties. CSS is caused by mutations in 5 genes encoding subunits of the BAF complex (*SMARCB1*, *SMARCE1*, *SMARCA4*, *ARID1A*, *ARID1B*) ;and 2 genes related to BAF complex (*PHF6* and *SOX11*) (Kosho et al., 2014). Haploinsufficiency or loss-of-function mutations in *ARID1B* are reported to be the most frequent cause of CSS, and also the most frequent cause of ID through recent exome-based studies on a large cohort of developmental disorders. However, a wide spectrum of phenotypic features has been described in patients with *ARID1B* abnormalities, and it would be difficult to recognize patients with *ARID1B* abnormalities especially in early childhood. Here, we report a 1.7-year-old boy with *ARID1B*-related CSS, presenting with hepatosplenomegaly. The patient was the third child of unrelated Japanese parents and born by normal delivery at term. He was found to have abdominal distension at age 1 month. He entered hospital several times due to recurrent infections since he was 2-months-old. He was fed by nasogastric tube, because of severe feeding difficulties and failure to thrive at age 4 months. When seen by us at 7 months, he presented with hepatosplenomegaly, hypertrichosis, short stature, developmental delay and generalized hypotonia. Screening for mucopolysaccharidoses and lysosomal diseases were negative. G-banded karyotype and chromosomal microarray analysis were normal. He developed acute encephalopathy with biphasic seizures and late reduced diffusion associated with hand, foot and mouth disease at age 10 months. Broad and short fingers were evident at age 1.3 year, and we performed genetic screening using a custom-made NGS-based panel including 49 genes known to cause syndromic or non-syndromic ID. A novel splice-donor site mutation (c.4071+1G>A) in *ARID1B* was detected heterozygously. Previously, a patient with *ARID1B*-related CSS was reported to have hepatomegaly (Vals et al., 2014). Hepatomegaly, therefore, could be a rare but recognizable feature of *ARID1B*-related CSS.

2505F

Expanding phenotypic spectrum of Takenouchi-Kosaki syndrome. *T. Takenouchi[1,2], A. Ahmad[3], B. O'Conner[3], T. Uehara[1], W. Gahl[4], K. Kosaki[1].* 1) Center for Medical Genetics, Keio University, Tokyo, Japan; 2) Department of Pediatrics, Keio University, Tokyo, Japan; 3) Department of Pediatrics, University of Michigan Medical School, IL; 4) National Human Genome Research Institute, MD.

Introduction: We recently established a new syndromic form of thrombocytopenia caused by mutation in *CDC42*, which is a small GTPase of the Rho family that plays pivotal roles in diverse biological activities including the cell cycle and formation of the actin cytoskeleton. This new entity characterized by intellectual disability and macrothrombopenia is now referred to as *Takenouchi-Kosaki syndrome,* OMIM #616737. The exact phenotypic spectrum of this syndrome still needs to be explored.**Clinical Report:** Three (2 Japanese female and 1 Caucasian male) patients presented with severe developmental delay and thrombocytopenia with increased platelet size on peripheral smear. They all had sensorineural hearing loss and structural brain abnormalities. Two patients had camptodactyly, and lymphedema in lower extremities. One patient had congenital hypothyroidism.**Molecular analysis:** All three patients underwent whole exome sequencing in *trio, which* identified a *de novo* mutation in *CDC42* (p.Tyr64Cys).**Discussion: W**e have identified the three patients carrying the same *de novo* amino acid substitution, p.Tyr64Cys, in *CDC42*. The cardinal features in this new syndrome include macrothrombocytopenia, intellectual disability, sensorineural deafness, and structural brain abnormalities. Associated features included camptodactyly, lymphedema in lower extremities, and congenital hypothyroidism. Examination of platelet count and peripheral blood smear is recommended when seeing patients presenting with intellectual disability/developmental delay.

2506W

Expanding phenotypic spectrum of patients with *WAC* mutations. *T. Uehara[1], T. Takenouchi[1,2], T. Ishige[3], H. Yoshihashi[4], K. Kosaki[1].* 1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 3) Department of Pediatrics, Gunma University, Gunma, Japan; 4) Department of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

The *WAC* genes maps to chromosome 10p12.1-p11.2 and encoded protein localizes to nuclear speckles and has various functions including regulattion of P53-dependent P21 transcription. Recently, DeSanto et al. identified 5 different heterozygous truncating mutations in the *WAC* gene (Desanto-Shinawi syndrome, MIM 616708) in 6 children. Besides intellectual disabilities and gastrointestinal problems, phenotypic spectrum of *WAC* haploinsufficiency is yet to be delineated. Here, we report two unrelated patients with *WAC* nonsense mutation who were identified through the Japanese nation-wide undiagnosed disease program, Initiative on Rare and Undiagnosed Diseases. Patient 1: A 22-year-old male who had *de novo* heterozygousd *WAC* p.Glu406* mutation on trio exome analysis and severe intellectual disability and rectal and bladder disturbance. Dysmorphic features included unilateral ptosis, universal hypertrichosis, synophrys, downslanting palpebral fissures, high-arched palate, low-hair live, narrow forehead, prominent jaw, and short tapering fingers. He exhibited pancytopenia with hypoplastic bone marrow. Patient 2: A 5-year-old male who had *de novo* heterozygous *WAC* p. Ser560* mutation on trio exome analysis and mild intellectual disability and no gastrointestinal problems. Dysmorphic features included upturned nose, thick lower lip, right accessory auricle, and downslanting palpebral fissures. Here we confirmed that haploinsufficiency in the *WAC* gene leads to syndromic intellectual disability. We suggest that downslanting palpebral fissures and gastrointestinal problems may point to the diagnosis. From a clinical standpoint, it is critical to note that Patient 1 developed pancytopenia with hypoplastic bone marrow because somatic recurrent *WAC* mutations have been shown among patients with acute myeloid leukemia. *WAC* may represent an emerging class of genes which is associated with intellectual disability as germline mutation and hematologic malignancies as somatic mutations. It is to be investigated whether patients with germline *WAC* mutations may be at risk for hematological problems.

2507T

Whole exome sequencing of families with intellectual disability. *J. Inlora[1], R. Sailani[1], J.L. Lynch[1], Z. Zolfaghari[2], O. Aryani[2], M.P. Snyder[1].* 1) Genetics Department, Stanford University, Palo Alto, CA; 2) Special Medical Center for Rare Disorders, Tehran, Iran.

Intellectual disabilities are extensively heterogeneous genetic disorders. Here we performed whole exome sequencing (WES) in three Iranian families with history of mental retardation. The pattern of inheritance is likely autosomal recessive or compound heterozygous. We applied WES to seven affected and seven normal individuals across the three families with a mean coverage of 70x. The genetic variants were analyzed using the Varseq software and filtered against the 1000 Genomes Project, dbSNP, ExAC, the UK10K Project as well as the Iranian Genome Project database. To date, we identified missense mutations in *DUX4L4, RLTPR, FANCI, MCM8* and *FAM83F* as the likely candidates for the disease. To our knowledge, none of these genes were previously reported to be associated with mental retardation. Further studies are being performed to validate the relation of the candidate genes with the phenotype of these families.

2508F

The new phenotypes observed in Coffin-Siris syndrome patients with
ARID1B **mutation.** *Y. Tsurusaki[1], Y. Enomoto[1], T. Yokoi[2], M. Minatogawa[2], C. Hatano-Abe[2], K. Ida[2], T. Naruto[3], K. Kurosawa[2].* 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Human Genetics, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan.

   Coffin-Siris syndrome (CSS; MIM 135900) is a congenital disorder characterized by intellectual disability, growth deficiency, microcephaly, coarse facial features, and hypoplastic or absent fifth fingernails and/or toenails. CSS is genetically caused by mutations in *SMARCB1*, *SMARCA4*, *SMARCE1*, *ARID1A*, *ARID1B*, that are part of the Brahma-associated factor (BAF) ATP-dependent chromatin-remodeling complex and *SOX11*. We examined three CSS-suspected patients by copy number analysis and targeted resequencing. Clinical information and peripheral blood were obtained from the family members after written informed consent. This study was approved by the Institutional Review Board of Kanagawa Children's Medical Center. Copy number analysis were studied by SurePrint G3 Human CGH 8 x 60K microarray (Agilent Technologies) according to the manufacturer's instructions. DNA microarray analysis revealed a 1.0-Mb deletion involving *ARID1B* at 6q25.3 in a CSS patient with vitiligo vulgaris. Targeted resequencing by HaloPlex Target Enrichment System (Agilent Technologies) with oligonucleotide probes targeting 437 genes related to epigenetics. DNA libraries were sequenced using MiSeq with 151-bp paired-end reads. Mapping to human genome hg19 was performed using Burrows-Wheeler Alignment. Aligned reads were processed by Picard to remove PCR duplicates. Variants were called using the Genome Analysis Toolkit, and annotated by ANNOVAR. We identified that *ARID1B* were mutated in remaining two patients: c.3731_3734del (p.S1244Ifs*10) in CSS patient with biliary atresia, and c.5547dup (p.S1851Kfs*5) in CSS patient with chiari malformation. Except for CSS phenotype, the clinical features of our three patients with *ARID1B* mutation showed vitiligo vulgaris, biliary atresia, and chiari malformation, respectively. To date, these phenotypes have not been reported. Our data further support that CSS with *ARID1B* mutations found variable phenotypes.

2509W

Gynecologic and reproductive health issues in patients with Smith-Magenis syndrome. *M.A. Merideth[1], W.J. Introne[2], W.A. Gahl[1,2], A.C.M. Smith[2].* 1) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 2) Office of the Clinical Director, National Human Genome Research Institute, NIH, Bethesda, MD.

   Objective: To evaluate the gynecologic and reproductive health issues in females with Smith-Magenis syndrome.   **Background:** Smith-Magenis syndrome (SMS, OMIM#182290) is a complex multisystem neurodevelopmental disorder characterized by minor craniofacial abnormalities, otolaryngologic abnormalities, developmental and speech/language delay, cognitive and functional impairment, circadian sleep disorder, and neurobehavioral abnormalities. SMS is caused by heterozygous deletion or mutation of the RAI1 gene on chromosome 17p11.2. Genitourinary anomalies are less frequent but include infantile cervix and/or hypoplastic uterus in females. Detailed information about the gynecologic(GYN) issues in this patient population is sparse.   **Methods:** Nineteen females with SMS were evaluated under an NIH IRB-approved protocol (01-HG-0109) from 2003 to 2016. Evaluation included GYN history, lab testing, review of outside records and physical exam.   **Results:** The patients ranged in age from 5 to 29 years (median 15y). Age at menarche had a median of 12.5 years (range 6-16y); one had precocious puberty. Ten patients reported heavy menstrual periods, 7 of whom were treated with oral contraceptive pills (OCPs) and 3 with progesterone IUDs (1 of whom had problems due to uterine anomaly). Menstrual management issues included reluctance to use pads, discomfort with blood on pads and refusal to take daily ocps. Menstrual hygiene issues led to suppression of menses by continuous OCPs in 3 patients, IUD in 1 patient and hysterectomy in 1 patient. Increased emotional/behavioral issues prior to or at onset of menses were reported in 10 patients. Two patients had increased seizures right before menses. Though no patients had issues with vaginal infections, 3 patients had problems with vaginal insertion of objects. Laboratory testing included normal estradiol, FSH and LH levels in all 7 patients in whom they were checked.   **Conclusions:** Our results indicate a possible increased risk of menorrhagia, catamenial seizures and premenstrual behavioral/emotional issues in females with SMS and warrants further investigation. Management of menses is also an issue in this patient population. Gaining a better understanding of the range of GYN problems in SMS helps to expand the phenotype and is the first step toward developing GYN therapeutic strategies for this potentially vulnerable patient population.

2510T

A variant in *MYH11* **is a cause of familial hiatal hernia.** *M.A. Gilbert[1], R. Rajagopalan[1], B.J. Wilkins[1,2], C.M. Grochowski[1], S.E. Sivilich[3], R.S. Moreland[3], M. Pack[2], D. Piccoli[1], N.B. Spinner[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 3) Drexel University, Philadelphia, PA.

Congenital diseases of the esophagus represent a diverse group of developmental disorders that are collectively detected in 1 in 3,500-5,000 live births. Hiatal hernia (MIM 142400) is one such disorder and is characterized by localization of both the distal esophagus and proximal stomach above the diaphragm, within the thoracic cavity, rather than below the diaphragm. This disease disrupts the smooth muscle function of the lower esophageal sphincter and predisposes patients to gastroesophageal reflux. Hiatal hernia occurs frequently in the adult population, but is observed rarely in children. Despite its rare occurrence in children, it shows a strong autosomal dominant pattern of inheritance in multiple reports of familial cases suggesting a possible genetic component to the disease. We identified a family of 7 with 5 affected individuals including a father and 4 children, and performed whole exome sequencing to identify a causal gene for this disease. We discovered an insertion of 2 base pairs found at the end of the myosin heavy chain 11 (*MYH11* [MIM 160745]) gene that was shared by all affected family members. MYH11 functions within smooth muscle and is generally involved in contraction. The variant [NM_022844.2(MYH11):c.5798_5799insCA(p.Gln1934Asnfs*90)] results in a frameshift and produces an elongated protein product containing 90 unique amino acids. A previously identified 1 base pair deletion at this same genetic locus also causes a frameshift resulting in an elongated protein containing 90 amino acids, with 89 of them being identical between the two variants. This deletion was shown to alter the function of MYH11 by leading to unregulated motor activity. We therefore believe that the variant we identified is likely to be pathogenic in this family. Further supporting a role for MYH11 in esophageal dysfunction, a mouse model of a zebrafish mutation that disrupts MYH11 regulation [NM_001161775.1(MYH11):c.1516T>C(p.Trp506Arg)] results in esophageal atresia. Esophageal atresia is characterized by an esophageal gap that prevents children from ingesting food and liquids and is fatal if left untreated. In *Myh11* mutant mice, we are able to show smooth muscle hypercontractility, which likely contributes to the development of hiatal hernia. We propose a genetic link between these two seemingly unrelated congenital diseases of the esophagus and suggest a broader role for *MYH11* variants in more common sporadic esophageal and gastrointestinal motility disorders.

2511F

A novel autosomal recessive alteration in the *RYR1* **gene in a patient with profound hypotonia.** *J. Gilfert[1], B. Marr[2], S. Groskin[2], A. Chamberlin[3], S. Tang[3], Z. Powis[3], M. Byler[1], R. Lebel[1].* 1) Section of Medical Genetics, Department of Pediatrics, SUNY Upstate Medical University, Syracuse, NY; 2) Crouse Hospital, Syracuse, NY; 3) Ambry Genetics, Aliso Viejo, CA.

Homozygous and compound heterozygous mutations in the *RYR1* gene have been reported to cause hypotonia, facial weakness, nemaline myopathy, respiratory insufficiency, swallowing disturbances, and ophthalmoplegia. Heterozygous mutations in *RYR1* have been associated with malignant hyperthermia sensitivity, central core disease, multiminicore disease, and spondylocostal dysarthrosis. We report a female neonate homozygous for c.14203C>T (p. R4735W) in exon 98 of the *RYR1* gene, a novel alteration. The phenotype was notable for decreased fetal movement, profound hypotonia, cerebral dysfunction (by EEG), and cerebral atrophy (by MRI). She did not have any skeletal anomalies appreciated on several chest and abdomen radiographs (normal spine, ribs and appendicular girdles, in AP views; humerus 64 mm, normal with normal metaphyses). She died at age 17 days, when artificial support was discontinued at the parents' request. Arginine at amino acid position 4735 is located in the Inter S2-S3 domain, a highly conserved 114-amino acid loop (positions 4,666 – 4779) important for *RYR1* gating. This amino acid is highly conserved in available vertebrate species and is present in the pH-regulated prokaryotic potassium channel KcsA . Structural remodeling based on homology to the known rabbit RYR1 structure revealed that the residue R4735 lies in the VSC (cytoplasmic subdomain of the voltage-sensing like domain, VSL) and at the interface between monomers and contact residues in the EF-hand subdomain of the Central domain, and lies at the end of bundles of helices forming the VSL domain. The residue appears to form contacts with residues on adjacent EF-hand domain and with a network of charged residues lying in the alpha-helices of the VSL domain. p. R4735W is an extremely rare allele in the control population (not in ESP and 1000 genomes and observed in 2 out of 121,158 total alleles studied in ExAC). This alteration is predicted to be probably damaging and tolerated by PolyPhen and SIFT *in silico* algorithms, respectively. An allelic change p.R4735Q, as well as some neighboring missense changes (p.Y4733D, p.G4734E, p.R4737Q, p.R4737W), have been reported in patients with malignant hyperthermia and are currently considered variants of uncertain significance. .

2512W

Variable clinical features in Pompe disease associated with novel mutations. *M. Wencel[1], K. Hall[1], J. Alany-dy[1], J. Simon[1], E. Valenti[1], C. Shambaugh[1], A. Surampalli[1], J. Yang[2], N. Goyal[3], T. Mozaffar[3], V. Kimonis[1].* 1) Division of Genetics and Genomic Medicine, Department of Pediatrics, University of California- Irvine, Orange, CA; 2) Division of Neuromuscular Neurology, Department of Neurology, University of California- Irvine, Orange, CA; 3) Dietitian & Nutrition Services, University of California- Irvine, Orange, CA.

   Pompe disease is a lysosomal storage disorder caused by the deficiency of enzyme acid alpha-glucosidase (GAA) which results in accumulation of glycogen particularly in the skeletal, cardiac, and smooth muscles. The late-onset form with symptoms presenting in childhood through adulthood, is characterized by proximal muscle weakness, respiratory insufficiency, and unlike classic or infantile-onset form typically with no cardiac involvement.   We report our experience with 17 adult patients (4 F/13 M) with Pompe disease at one center, several of whom had unique findings and novel mutations. Patients ranged in ages from 18-69 y. (mean 51 y.) and were diagnosed at a range of 11-65 y (mean 37 y.) often after a history of progressive muscle disease of several years duration. Genetic sequencing revealed that 15/17 individuals had the common c.-32-13T>G adult onset mutation, and eight individuals had 6 novel mutations: c.1594G>A, c.2431delC, c.2655_2656delCG, c. 1951-1952delG-GinsT, c.525_526delTG, and c.1134C>G. A male with the c.1594G>A mutation developed an intracerebral aneurysm at the age of 43 y. and was treated with surgery. Another male with the c.525_526delTG developed testicular cancer and is currently in remission. Cardiomyopathy was noted in an adult with the c.525_526 delTG mutation, and peripheral neuropathy in a male with the c. 1951-1952delGGinsT. Two siblings with the c.2655_2656delCG developed very high antibody titers, one of whom developed a severe infusion reaction. Other clinical features included scoliosis and cardiomyopathy in an adolescent, BiPAP requirement in eleven, tinnitus in five, and one individual was born with partially developed hip and clubfoot. All patients currently receive alglucosidase alfa wi th different response rates in their muscle weakness, pulmonary function, dynamometry, and functional studies. Our patient cohort illustrates the variable range of clinical features, and alert us to the importance of careful monitoring and early management of these complications. Possible genotype-phenotype associations with these novel mutation will emerge with larger studies.

2513T

Continued identification of new genes, pathways and animal models for rare infantile forms of myopathies and neuromuscular disorders. *L.L. Baumbach-Reardon[1], J.M. Hunter[2], M.E. Ahearn[1], C. Balak[1], M. Huentelman[1], S.M. Bernes[3].* 1) TGen, Phoenix, AZ; 2) Ambry Genetics, Aliso Viejo, CA; 3) Phoenix Childrens Hospital, Phoenix, AZ.

   Neuromuscular disorders and myopathies account for a significant proportion of infant and childhood mortality and chronic disease. Phenotypes and genotypes of these disorders cover a broad spectrum but have many overlapping features. As we learn more about the etiology of human disease, we have found that most severe disease causing variants are very rare, only occurring in a single or a few individuals. One of the main goals of our most recent studies was to use next generation sequencing to identify causal mutations for cases in which prior genetic testing and histology did not reveal the disease etiology. Over the past 3 years, we have completed whole exome sequencing (WES) on 203 individuals including 73 affected individuals and additional family members. From these cases, we have identified novel probable pathogenic mutations in genes known to cause disease, as well as in genes not previously associated with disease, in ~30% of cases. These pathogenic variants include de novo, recessive, X-linked, and dominant inheritance patterns. We were able to detect distinct pathogenic mutations in several genes in multiple families (i.e., Nebulin, Titin, and RYR1). However, the majority of cases represent one-time occurrences of a mutation in either a previously known disease-causing gene, such as PLA2G6, or in previously unknown neuromuscular disease (CACNAS1 and COL6A3). Variants in novel genes will require further validation through functional studies. One of our most interesting findings is the detection of three novel pathogenic mutations in families which very much resembled XL-SMA, but instead, had mutations in another X-linked gene, SCLM2, an important gene in transcriptional repression that had not been associated with human disease. The clinical features of these families and our collaborative development of an SCML2 knock-out mouse will be presented, as well as our recent success in creating an XL-SMA targeted knock-in mouse. In summary, we applied WES using the latest NGS instrumentation and technologies to identify the genetic cause of disease in many of these cases. The genetic diversity revealed in these studies demonstrates the continued important clinical utility of WES for undiagnosed cases of infantile and childhood NMD. We continue to evaluate functionality of novel variants through *in vitro* studies, cell culture and animal models.

2514F

Attenuation of cardiomyopathy in Duchenne Muscular Dystrophy mice expressing retinal dystrophin in heart muscle. *A. Vince, M. Woods, A. Wiggins-McDaniel, J. Beckler, S. Myer, B. Wakefield, J. Harkless, R.A. White.* College of Biosciences, Kansas City University of Medicine & Biosciences, Kansas City , MO.

Duchenne Muscular Dystrophy (DMD) is a common neuromuscular disease characterized by progressive muscle degeneration and caused by mutations resulting in the absence of dystrophin. The 260 kDa human retinal dystrophin isoform (Dp260), plays a role similar to the mechanical support mechanism for muscle dystrophin. A human Dp260 transgene (Tg+) has been produced and tested in dystrophin and utrophin deficient double mutant mice (DM) which are excellent clinical models of DMD. We have previously shown that muscle-specific expression of Dp260 in transgenic double mutant DMD-Tg+ animals has striking palliative effects on spinal curvature, pathology and electrophysiology. Cardiomyopathy is also present in over 90% of DMD patients and is responsible for 40% of their deaths. DM mice also exhibit cardiomyopathy. Structural and functional analyses were carried out on the effect of expressing human Dp260 in the hearts of DMD mice. Expression of human Dp260 from the transgene in cardiac muscle was determined using western blotting. Detection of human retinal dystrophin was observed in transgenic mice and DM, Tg+ mice were the only animals that solely expressed the Dp260 and not full length dystrophin. Histopathology analysis was used to determine the effects of Dp260 expression on cardiac pathology in double mutant DMD mice, with and without the retinal dystrophin Tg+. DM mice can have thinning of the left ventricular wall with extensive fibrosis typical of dilated cardiomyopathy whereas DM, Tg+ hearts only showed normal phenotypes. DM mice heart cross-sections showed variability in fibrosis and dilated cardiomyopathy, while all DM, Tg+ mice displayed results more consistent with control mice. Analysis of Dp260 expression on heart function was also conducted on DM-Tg+, DM, and control mice which included measurements of the myocardium size, cardiac output, and fractional shortening (heart function during systole and diastole). EKG data showed a significant decrease in heart rate and left ventricular mass in DM mice while DM, Tg+ mice are normal. Fractional shortening data were not significantly different. The impact of Dp260 expression on lifespan in DM mice was significant. DM, Tg+ mice survive a normal lifespan while DM mice are deceased by 5 months of age. The effect of Dp260 expression in double mutant DMD mice has been examined and positive effects on lifespan are observed along with improvement of the heart ultrastructure and function.

2515W

Human patients and mouse modeling identifies striated preferentially expressed gene (SPEG) as a key triad muscle protein potentially involved in excitation-contraction coupling. *P.B. Agrawal, S. Cao, C. Lee, A.H. Beggs, J.J. Widrick.* Department of Medicine, Boston Children's Hospital Boston, MA.

We recently described several patients with centronuclear myopathy and cardiomyopathy caused by mutations in the *SPEG* gene. We now report the generation of a conditional knockout mouse model with inducible *Speg* deletion driven by crossing with various tissue-specific cre-expressing mice. Constitutive knockouts, and those bred with *Acta1*-cre died *in utero* at E19 while those bred with *MCK*-cre lived beyond 6 months of age. *MCK*-cre expression starts at postnatal day 1 (P1), peaks at P10 and stays high after that. *MCK*-cre+ *Speg* KO mice start to grow slower compared to their littermates at around a month of age with a significant weight difference by 3 months. The mean force generated by the *Speg*-KO tibialis anterior muscle at 3 months of age was about one-third of that generated by the WT muscle (386 versus 1209 mN in KO versus control). This force deficit remained after correcting for differences in muscle cross sectional area (128 versus 262 kilopascals in KO versus control) indicating a reduction in the intrinsic contractility of the *Speg* KO muscles. The force-frequency curves of TA muscles from the SPEG mice were shifted downward and to the right, indicating a preferential loss of force at low- to mid-frequencies of stimulation. Electron microscopy of muscles from affected mice demonstrates demonstrated abnormal triad morphology which is consistent with our immunhistochemistry studies showing that SPEG is localized there. Abnormal triads and preferential loss of low-frequency force suggest excitation-contraction failure in SPEG muscles, and work is ongoing to test this hypothesis. Further ongoing investigations include physiological and histopathological analysis, caffeine and single fiber studies . These data suggest that poor muscle function seen in SPEG-deficient mice is consistent with the human myopathy observed in patients with *SPEG* mutations.

2516T

Exploring the diverse genetic landscape of limb-girdle muscular dystrophies from a large cohort of myopathy patients. *B.R. Nallamilli, A. Tanner, Z. Whitt, C. da Silva, S. Mehta, M. Hegde.* Human Genetics, Emory University, Atlanta, GA.

Limb-girdle muscular dystrophies (LGMD) are a group of heterogeneous genetic disorders predominantly involved in proximal muscle weakness with an autosomal recessive or dominant mode of inheritance. To date, more than 30 genes have been shown to be associated with different subtypes of LGMD. The clinical and genetic heterogeneities of LGMD make disease diagnosis complicated and expensive. Recently, MDA and Jain Foundation jointly launched a LGMD sequencing program consisting of a 35 gene NGS panel through Emory Genetics Laboratory (EGL), through which over 2000 patients have been sequenced. Definitive molecular diagnosis has been established in 40% of the patients with the majority having mutations in either of 5 genes *CAPN3*, *DYSF*, *ANO5*, *FKRP* and *COL6A1*, indicating that these are the major contributors to LGMD. Late onset Pompe disease is a rare, but potentially treatable lysosomal storage disorder characterized by progressive glycogen accumulation and muscle weakness, with often a limb-girdle pattern. Therefore it should not be overlooked. We successfully identified more than 21 late onset Pompe patients with pathogenic variants in *GAA* gene. We also identified multiple patients with two pathogenic variants in *TTN* gene indicating the need for further studies in understanding the titin pathophysiological role in recessive muscle disorders. More interestingly, in at least 10% of the patients, pathogenic variants were detected in more than one LGMD genes. For example, one patient clinically diagnosed with LGMD2L but with an unexpected severe progression had homozygous pathogenic variants in both *ANO5* and *SGCA* genes. Similar multigenic combinations of pathogenic variants were detected in other individuals suggesting a possible role of synergestic heterozygosity and digenic contribution to disease presentation and progression. Around 25% of patients have been identified with variants of unknown significance (VOUS) which clearly suggesting the immediate need for further functional analysis to prove pathogenicity and to take the advantage of enrolling in numerous available clinical trials. Increased diagnostic yield observed in LGMD gene panel testing when compared to exome sequencing. Overall, current large scale LGMD sequencing project has tremendously improved our knowledge in understanding the relative prevalence of different subtypes of LGMD, allowing timely management and participation of affected individuals in clinical trials. .

2517F

AAV capsid evolution for therapeutic transduction of human skeletal muscle for use in delivery of passive vaccines and muscle gene therapies. *N.K. Paulk[1], G.W. Charville[2], K. Maguire-Nguyen[2], M. Gardner[3], K. Pekrun[1], Y. Zhang[1], B. Yoo[2], M. Tiffany[1], J.G. Vilches-Moure[4], G.K. Lee[5], J.B. Shrager[6], G. Gao[7], M. Farzan[3], T.A. Rando[2], M.A. Kay[1].* 1) Pediatrics and Genetics Depts, Stanford University, Stanford, CA; 2) Neurology and Neurological Science Dept, Stanford University, Stanford, CA; 3) Immunology and Microbial Science Dept, The Scripps Research Institute, Jupiter, FL; 4) Comparative Medicine Dept, Stanford University, Stanford, CA; 5) Surgery Dept, Stanford University, Stanford, CA; 6) Cardiothoracic Surgery Dept, Stanford University, Stanford, CA; 7) Microbiology and Physiological Systems Dept, University of Massachusetts, Worcester, MA.

Our goal is to develop an AAV vector that can transduce human skeletal muscle after intramuscular administration at levels sufficient to express therapeutic levels of antibodies in passive vaccines or transgenes in muscle gene therapies. Yet greater human skeletal muscle transduction is needed than has been achieved with existing serotypes. To engineer novel capsids with unparalleled human skeletal muscle transduction, we utilized primary human skeletal muscle cells from surgical resections to screen diverse libraries of shuffled AAV capsids. Two screens were performed in pools of primary muscle stem cells or myotubes from 6 patients. Six rounds of replicating selection were carried out and the 5 most highly selected variants from each screen were vectorized and validated against existing serotypes with muscle tropism (AAV1, 6 and 8). In primary human muscle stem cells, variants NP22, NP66 and NP94 had significantly increased transduction that ranged from a 10 to 284-fold improvement over AAV1, 6 and 8. In primary human myotubes, NP22, NP66 and NP94 again showed significantly increased transduction that ranged from a 13 to 871-fold improvement over AAV1, 6 and 8. To assess human transduction *in vivo*, we injected xenografted humanized muscle mice with variants expressing luciferase by IM injection and assessed transduction weekly by live imaging and variant NP66 demonstrated the highest sustained transduction over controls. To better predict eventual muscle transduction in humans, we transduced human muscle explants from surgical resections *ex vivo*. Four adults (2 male and 2 female) had healthy skeletal muscle tissue removed for *ex vivo* transduction with luciferase. In all 4 patients, NP22 and NP66 had significantly increased transduction by live fiber imaging as well as luciferase assays on lysed fibers that ranged from a 4 to 116-fold improvement over control serotypes. To support preclinical testing, we assessed transduction in rhesus macaque skeletal muscle explants *ex vivo*. Here again NP22 and NP66 showed a 30 to 57-fold improvement in transduction over control serotypes. Our results show that capsids NP22 and NP66 have highly significant increased human skeletal muscle fiber transduction when assessed *in vitro*, *in vivo* and *ex vivo*. These capsids represent powerful tools to express therapeutic quantities of gene products in humans. .

2518W

Identification of disease causing genes via exome sequencing and expression profiling in patients with mitochondrial disorder. *D.M. Bader[1,2], L.S. Kremer[3], G. Pichler[4], T. Schwarzmayr[3], E. Holzerova[3], R. Kopajtich[3], T. Wieland[3], T.M. Strom[3,5], M. Mann[4], J. Gagneur[1,2], H. Prokisch[3,5].* 1) Computational Biology, Departement of Bioinformatics, Technical University of Munich, Arcisstrasse 21, 80333 München, Germany; 2) Quantitative Biosciences Munich, Gene Center, Department of Biochemistry, Ludwig Maximilian University of Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany; 3) Institute for Human Genetics, Helmholtz Zentrum Munich, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany; 4) Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; 5) Institute of Human Genetics, Klinikum rechts der Isar, Technical University of Munich, Trogerstrasse 32, 81675 München, Germany.

So far gene expression was mostly studied at the RNA level, however mass-spectrometry (MS) has become more comprehensive. Therefore a proteome-scale quantification is now possible using MS. Here, we are performing an integrative analysis of transcriptome and proteome profiles of 31 fibroblast cell lines of patients affected by mitochondrial disease as a systematic expansion and follow-up of our earlier exome sequencing based studies. We can show that controlling for hidden technical biases rescues the similarity of biological replicates. When correlating the corrected transcriptome and proteome data, proteins that are members of the same complex tend to have widespread post-transcriptional co-regulation. The genes encoding the mitochondrial respiratory chain complex I also have significantly lower RNA to protein correlation than genes of other complexes. Accordingly, we can improve the detection of causal gene candidates compared to exome based diagnosis by screening for protein complex degradation and expression outliers. Starting from twelve (39%) patients with validated causal genes from exome analysis, we could already validate three new causal genes that are found with the help of transcriptome and proteome data. However, further causal gene candidates are still under validation which could boost the diagnosis rate.

2519T

Mitochondrial studies in a novel disorder associated with NUBPL associated mitochondrial complex I deficiency in patients with global developmental delays, ataxia, cerebellar and pons hypoplasia. *V. Kimonis[1], P.S. Eis[2], S. Parikh[3], D. Scott[4], M.K. Koenig[5], A. Maclean[6], S. Tang[7], A.N. Hasso[8], M. Wydro[6], J. Balk[6], E. Chao[1,7], E. Hatchwell[2].* 1) Div Gen, Dept Pediatrics, Univ CA Irvine, Orange, CA, Children's Hospital of Orange County, Orange, CA; 2) Population Diagnostics, Inc., Melville, NY; 3) Center for Pediatric Neurology, Cleveland Clinic, Cleveland, OH; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Departments of Pediatrics and Neurology, The University of Texas Medical School at Houston, Houston, TX; 6) John Innes Centre, Colney Lane, Norwich, UK; 7) Ambry Genetics, Aliso Viejo, CA; 8) Radiological Sciences, School of Medicine, University of California, Irvine, CA.

The *NUBPL* gene (MIM 613621) was first reported as a cause of mitochondrial Complex I deficiency syndrome (MIM 252010) in 2010, and only seven families have been reported worldwide. We now report clinical features of 5 patients in 4 families, representing the largest number of patients diagnosed worldwide with *NUBPL* Complex I deficiency (CID). Patients were diagnosed using Whole Exome Sequencing (WES) and all were found to be compound heterozygous for a *NUBPL* (NM_025152) splicing mutation (c.815-27T>C) that is common in the population (~1% frequency), plus a rarer non-synonymous or splicing mutation (in *trans*) as follows: c.311T>C (p.L104P) in families 1 (cases 1 and 2) and 4 (case 5), c.693+1G>A in family 2 (case 3), and c.545T>C (p.V182A) in family 3 (case 4). We note that in 3 of 4 families (all except family 1), the c.815-27T>C mutation is present in *cis* with c.166G>A (p.G56R), whose pathogenicity remains in question. Primary clinical features of the CID patients, who are presently ages 3-19 years, include onset of neurological symptoms at ages 3-18 months, global developmental delay, ataxia, nystagmus, speech articulation difficulties, and cerebellar dysfunction (brain MRIs show progressive cerebellar and pons hypoplasia) and Leigh like phenotype in one individual. Importantly, despite having a diagnosis of CID, most of the patients tested normal in an electron transport chain assay (fibroblast or muscle), including for Complex I activity. Functional studies in yeast (*Yarrowia lipolytica*) were performed and confirmed the pathogenicity of the non-synonymous *NUBPL* (yeast homolog is Ind1) mutations (L104P and V182A). In addition efficacy of the coenzyme Q10 component of current therapy (e.g., mitochondrial cocktail components) was suggested by improvement in CI activity. Biochemical assays in patient fibroblasts showed mitochondrial functional differences, and differentiated neurons are being utilized to screen for and identify novel therapies for these patients. In summary we present clinical features of a unique new NUBPL associated mitochondrial complex 1 disorder identified by WES and promising treatment studies in yeast and patient fibroblasts. .

2520F

Novel mutation in *ATP5D* **as a cause of mitochondrial ATP synthase deficiency.** *J. Kohler[1,3,4], D. Fisk[2,4], M. Grove[2,3,4], D. Zastrow[1,4], E. Fernandez[1,4], A. Dries[1,4], P. Zornio[1,4], S. Schelley[5], G. Enns[5], U.D.N. Members[8], E. Ashley[1,2,3,4,6], P. Fisher[1,7], J. Bernstein[1,5], M. Wheeler[1,3,4].* 1) Stanford Center for Undiagnosed Diseases, Stanford, CA; 2) Stanford Center for Genomics and Personalized Medicine, Stanford, CA; 3) Stanford Center for Inherited Cardiovascular Disease, Stanford, CA; 4) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 5) Department of Pediatrics, Stanford University, Stanford, CA; 6) Department of Genetics, Stanford University, Stanford, CA; 7) Department of Neurology, Stanford University, Stanford, CA; 8) National Institutes of Health, Bethesda, MD.

Mitochondrial diseases are challenging to diagnose given their wide clinical variability and nonspecific clinical findings.   Here we describe a 7-year old Mexican female who presented at 2 days of age with hypoglycemia (glucose 28), lactic acidosis (lactate 34mm/L), and hyperammonemia (ammonia 200umol/L) worsened by dextrose infusion. Extensive metabolic evaluation was nonspecific. Laboratory testing including fibroblast analysis for electron transport chain enzyme assembly, pyruvate carboxylase/pyruvate dehydrogenase deficiency, and mitochondrial DNA sequence analysis were normal. Brain MRI with MR spectroscopy showed no specific abnormalities. Through childhood she had several hospitalizations for metabolic decompensation, lactic acidosis and rhabdomyolysis. She was treated with metabolic cocktail. She is developmentally normal between episodes and has a history of resolved dilated cardiomyopathy. Further genetic workup including clinical whole exome sequencing were uninformative.   Exome reanalysis at the Stanford Center for Undiagnosed Diseases filtered variants by quality, frequency, predicted functional effect, and inheritance pattern, resulting in 95 rare homozygous variants, 6 of which were in mitochondrial genes. Final manual curation identified a variant [c.245C>T(p.P82L)] in the *ATP5D* gene, Sanger-confirmed homozygous in the proband and heterozygous in each parent. *ATP5D* encodes the δ subunit of mitochondrial ATP synthase, one of five subunits in its F1 complex, that combines with the γ and ε subunits to form the central stalk. The P82 residue is conserved to yeast; the variant is predicted damaging by in-silico models. Native gel electrophoresis was performed to assess if the *ATP5D* variant impacted F1 complex, and confirmed decreased activity and normal assembly of mitochondrial ATP synthase in mitochondria from patient fibroblasts.   *ATP5D* variants have not been previously associated with disease. However, pathogenic missense variants in genes encoding the α and ε subunits of the F1 complex have been described in patients with ATP synthase deficiency (ATPsd), a clinically heterogeneous mitochondrial disorder involving neonatal-onset hypotonia, respiratory distress, encephalopathy, lactic acidosis, hyperammonemia, and cardiomyopathy. Our patient's initial presentation and subsequent episodes of decompensation are consistent with ATPsd. Given the overlapping phenotype and pathophysiology, we report a compelling novel gene-disease association.

2521W

KIF5A **mutations cause an infantile onset phenotype including severe myoclonus with evidence of mitochondrial dysfunction.** *J. Duis[1], S. Dean[2], C. Applegate[1], A. Harper[3], R. Xiao[4], W. He[4], J.D. Dollar[5], L.R. Sun[2], M. Biderman Waberski[6], T.O. Crawford[2], A. Hamosh[1], C. Stafstrom[2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Pediatric Neurology Johns Hopkins School of Medicine Baltimore, MD; 3) Carolinas Pediatric Neurology Care Charlotte, NC; 4) Department of Molecular and Human Genetics Baylor Miraca Genetics Laboratories Baylor College of Medicine Houston, TX; 5) Carolinas Pathology Group, Charlotte, NC; 6) National Institutes of Health Bethesda, MD.

Objective: Missense mutations in kinesin family member 5A (*KIF5A*) are known to cause spastic paraplegia 10 characterized by adult onset spastic paraparesis. We report two patients with novel *de novo KIF5A* frameshift mutations resulting in severe infantile onset myoclonus, with associated negative EEG findings, hypotonia, optic nerve abnormalities, oral motor dysfunction, dysphagia, apnea, and early developmental arrest. Methods: Genetic testing to identify DNA variants included whole exome sequencing, mitochondrial DNA sequencing and solution-based hybrid capture of targeted genes coupled with next-generation sequencing using the Illumina® HiSeq platform. Results: Sequencing analysis revealed de novo c.2854delC and c.2934delG variants in *KIF5A,* both of which create a frameshift mutation that results in a stop-loss with read through the normal termination codon to produce an elongated protein. Enzymatic testing of muscle from the second case suggested complex IV deficiency supportive of mitochondrial dysfunction. EEG studies revealed no epileptiform correlate to clinical myoclonus. Conclusions: Our two cases expand the phenotypic spectrum of *KIF5A* mutations to include a neonatal-onset myoclonic disorder associated with presumed dominant negative stop-loss frameshift mutations with predicted alteration and elongation of the carboxy-terminus (C-terminus) of the protein. Biochemical evidence of mitochondrial dysfunction in the setting of an abnormal kinesin "motor" suggests a role of impaired mitochondrial trafficking. Additionally, we hypothesize that interrupted binding of the GABA receptor plays a role in the severe myoclonus. These results highlight the role of expanded testing and whole exome sequencing in critically ill infants and emphasize the importance of accurate test interpretation.

2522T

Increased survival and improvement of neurological phenotype by anaplerotic therapy in a fish model of propionic acidemia. *V.M. Ginocchio[1], E. De Felice[1], S.E. Hofherr[2], F.T. Barrows[3], R.M. Sepe[4], P. Sordino[4], F.G. Salierno[1], I. Conte[1], N. Brunetti-Pierri[1,5].* 1) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 2) Department of Lab Medicine, Children's National Medical Center, Washington, DC, USA; 3) USDA Agricultural Research Service, Bozeman, MT, USA; 4) Stazione Zoologica Anton Dohrn, Naples, Italy; 5) Department of Translational Medicine, Federico II University of Naples, Italy.

Propionic acidemia (PA) is an autosomal recessive inborn error of metabolism caused by deficiency of the mitochondrial propionyl-CoA carboxylase (PCC). The disease presents with acute, recurrent and life-threatening crises of metabolic decompensation starting from the newborn period. Patients also suffer from multi-organ complications, neurological dysfunction, and cardiomyopathy. Despite available treatments, disease mortality and morbidity remain elevated and investigation of novel and more effective therapies is highly needed. We generated a PA model in medaka fishes (*Oryzias latipes*) using TALENs targeting the *pccb* medaka gene, encoding one of the subunits of PCC. The *pccb-/-* medaka showed defective hatching, developmental delay, reduced locomotor activity, and early lethality. Livers and hearts of *pccb-/-* larvae exhibited increased lipid droplets, likely as a consequence of impaired mitochondrial beta-oxidation. Importantly, *pccb-/-* larvae showed a significant increase in the levels of C3 (propionylcarnitine), the biochemical hallmark of the disease. Non-protein energy sources and carnitine improved the survival of *pccb-/-* larvae. Because PCC deficiency induces depletion of Krebs cycle intermediates, we investigated the efficacy of an anaplerotic therapy in *pccb-/-* medaka. The anaplerotic therapy comprising sodium citrate, ornithine alpha-ketoglutarate, and glutamine resulted in significant improvements of both survival and locomotor activity in *pccb-/-* larvae. In summary, we generated a fish model of PA recapitulating the clinical and biochemical features of human patients. Anaplerotic therapy supplying intermediates of the Krebs cycle significantly improved the fish phenotype and has high potential for clinical translation. Given its small size, easy handling and large number of progeny, the PA medaka model is suitable for large-scale drug screening and rapid investigation of novel drugs.

2523F

Phenotypic heterogeneity of 2,4-Dienoyl-CoA Reductase deficiency caused by start loss NADK2 variant detected in Undiagnosed Disease Network subject. *J. Phillips, J. Cogan, T. Reimschisel, J. Gauthier, A. Robertson, A. Bican, T. McGregor, J. Newman, R. Hamid.* Dept Ped, Path, Micro, Immun, & Med, Vanderbilt Univ Sch Med, Nashville, TN.

Dienoyl-CoA reductase deficiency (DECRD) is a rare autosomal recessive disorder of mitochondrial β-oxidation of unsaturated fatty acids (OMIM 616034). DECR is crucial for the oxidation of polyunsaturated fatty acids and reduces conjugated Δ2, Δ4-dienoyl-CoAs to a Δ3-enoyl-CoA in an NADPH-dependent reaction. Until now DECR deficiency has been reported in only two patients whose disease courses were severe and unremitting; one patient died at 4 months and the other at 5 years of age. The first had failure to thrive, hypotonia, microcephaly and delayed milestones. The second developed metabolic acidosis, severe encephalopathy and leukodystrophy. Both had hyperlysinemia and DECR activities were 17% in muscle and 10% in fibroblasts, respectively. Here, we report an affected 15 year old female (third known case) with normal intelligence and a mild clinical and biochemical phenotype. She was referred to the undiagnosed diseases network (UDN) for a diagnostic workup. Her symptoms started at age 9 years and consisted of decreased visual acuity, bilateral optic atrophy, nystagmus, lower extremity weakness, numbness and difficulty walking. Limb weakness was episodic in nature and is accompanied with elevations of liver enzymes and creatinine kinase of up to 18k. Her neurological symptoms are currently stable. MRI showed stable optic pathway atrophy and volume loss. EMG showed chronic axonal sensorimotor peripheral neuropathy and mild proximal neuropathy. Urine organic acid showed highly elevated lactic and pyruvic acid. Plasma amino acids were within normal limits other than mean lysine and proline levels that were 3.7 and 2.5 times the upper limits of normal. Negative molecular studies included mitochondrial DNA sequencing and comprehensive muscular dystrophy panel. Whole exome sequencing showed homozygosity for a c.1A>G; p. Met1Val start loss mutation in the primary transcript (NM_001085411.1) of NADK2 which encodes the 442 amino acid isoform. This variant alters the start codon, has not been previously reported, is absent from external reference databases including 1000G/EVS/ExAC, and was heterozygous both of the patient's parents. This patient's near normal intelligence and stable disease expands the phenotypic heterogeneity and pleiotropism, as well as the prognosis associated with DECRD. Our results further suggest that attention should be paid to cases of hyperlysinemia especially those with visual loss and neurological phenotypes as some may also have DECRD.

2524W

Hyperammonemia induces hepatic autophagy and its enhancement improves the clearance of ammonia *in vivo*. *L.R. Soria[1], N. Pastore[2,3], P. Annunziata[1], E. Polishchuk[1], R. Polishchuk[1], E. Nusco[1], A. Ballabio[1,2,3,4], N. Brunetti-Pierri[1,4].* 1) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 2) Department of Molecular and Human Genetics; Baylor College of Medicine; Houston, TX, USA; 3) Jan and Dan Duncan Neurological Research Institute; Texas Children's Hospital; Houston, TX, USA; 4) Department of Translational Medicine, Federico II University, Naples, Italy.

The urea cycle is the main pathway in mammals for nitrogen detoxification. Systemic ammonia is elevated in patients with an inherited or acquired impairment of hepatic ammonia detoxification. Hyperammonemia may result in irreversible brain damage if not treated early and thoroughly and available treatments are often inadequate. Ammonia has been found in cell cultures to stimulate autophagy, a conserved catabolic process that degrades cytoplasmic components and organelles in the lysosome. Therefore, we investigated whether autophagy is also induced *in vivo* and whether modulation of autophagy plays a role in the severity of hyperammonemia. We confirmed that ammonia induced autophagy in cells and we found that it stimulates nuclear translocation of TFEB, a master regulator of autophagy and lysosomal function. Liver autophagy was found to be activated in C57BL/6 wild-type mice during acute hyperammonemia induced by intraperitoneal (i.p.) injections of ammonium chloride. Increased hepatic autophagy was also observed in ornithine transcarbamylase-deficient mice, a mouse model of the most common and severe inborn error of urea synthesis. Suppression of autophagy by colchicine or vinblastine severely impaired the clearance of ammonia following an i.p. ammonia load. Furthermore, TFEB liver-specific knockout mice showed a defect of ammonia handling compared to control mice, suggesting that hepatic autophagy is an important mechanism for ammonia detoxification. Moreover, enhancement of autophagy by liver-directed gene transfer of TFEB mediated by intravenous injection of a helper-dependent adenoviral vector protected against acute hyperammonemia. In addition, ammonia detoxification was ameliorated by autophagy-inducing small molecules. Autophagy activation was associated with increased ureagenesis and higher levels of N-acetylglutamate precursors, ATP, and aspartate that activate and fuel the urea cycle. Importantly, pharmacological enhancement of autophagy reduced blood ammonia in models of both acute and chronic hyperammonemia, i.e. thioacetamide-induced liver failure and diet supplemented with ammonium acetate, respectively. In summary, our data show that hepatic autophagy is activated by hyperammonemia, deficiency of autophagy impairs ammonia handling, and its enhancement can be exploited therapeutically for both acute and chronic hyperammonemia.

2525T

***NAXE* mutations disrupt the cellular NAD(P)HX repair system and cause a lethal neurometabolic disorder of early childhood.** *E. Koňaříková[1,2], L.S. Kremer[1,2], K. Danhauser[3], D. Herebian[3], D. Petkovic Ramadža[4], D. Piekutowska-Abramczuk[5], A. Seibt[3], W. Müller-Felber[6], T.B. Haack[1,2], R. Płoski[7], K. Lohmeier[3], D. Schneider[8], D. Klee[9], D. Rokicki[10], E. Mayatepek[3], T.M. Strom[1,2], T. Meitinger[1,2], T. Klopstock[11,12,13], E. Pronicka[5,10], J.A. Mayr[14], I. Baric[4], F. Distelmaier[3], H. Prokisch[1,2].* 1) Institute of Human Genetics, Technische Universität München, Munich, Germany; 2) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 3) Department of General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Heinrich-Heine-University Düsseldorf, Germany; 4) Department of Pediatrics, University Hospital Center Zagreb and University of Zagreb, School of Medicine, Zagreb, Croatia; 5) Department of Medical Genetics, The Children's Memorial Health Institute, Warsaw, Poland; 6) Department of Pediatric Neurology and Developmental Medicine, Ludwig-Maximilians-University of Munich, Munich, Germany; 7) Department of Medical Genetics, Warsaw Medical University, Warsaw, Poland; 8) Clinic of Pediatrics, Dortmund, Germany; 9) Medical Faculty, Department of Diagnostic and Interventional Radiology, University Düsseldorf, Düsseldorf, Germany; 10) Department of Pediatrics, Nutrition and Metabolic Diseases, Department of Medical Genetics, The Children's Memorial Health Institute, Warsaw, Poland; 11) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilians-University, Munich, Germany; 12) German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; 13) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 14) Department of Pediatrics, Paracelsus Medical University, Salzburg, Austria.

To safeguard the cell from the accumulation of potentially harmful metabolic intermediates, specific repair mechanisms have evolved. *APOA1BP,* now renamed *NAXE,* encodes an epimerase essential in the cellular metabolite repair for NADHX and NADPHX. The enzyme catalyzes the epimerization of NAD(P)HX, thereby avoiding the accumulation of toxic metabolites. The clinical importance of the NAD(P)HX repair system has been unknown. Exome sequencing revealed pathogenic biallelic mutations in *NAXE* in children from four families with (sub-) acute onset ataxia, cerebellar edema, spinal myelopathy, and skin lesions. Lactate was elevated in cerebrospinal fluid of all affected individuals. Disease onset was during the second year of life and clinical signs as well as episodes of deterioration were triggered by febrile infections. Disease course was rapidly progressive, leading to coma and finally to death in all affected individuals. NAXE protein levels were undetectable in fibroblasts from affected individuals of three families. In these fibroblasts we measured highly elevated concentrations of the toxic metabolite cyclic-NADHX, confirming a deficiency of the mitochondrial NAD(P)HX repair system. Finally, NAD or nicotinic acid (vitamin B3) supplementation might have important therapeutic implications for this fatal disorder.

2526F

A case of Leigh syndrome due to hemizygous *NDUFA1* **mutation with atypical spinal cord involvement.** *A. Miyauchi[1], H. Osaka[1], M. Nagashima[1], M. Kuwajima[1], Y. Monden[1], M. Kohda[2], Y. Kishita[3], Y. Okazaki[2,3], K. Murayama[4], A. Ohtake[5], T. Yamagata[1].* 1) Pediatrics, Jichi Medical University, Tochigi, Japan; 2) Division of Translational Research, Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan; 3) Division of Functional Genomics & Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan; 4) Pediatrics, Chiba Children's Hospital, Chiba, Japan; 5) Pediatrics, Saitama Medical University, Saitama, Japan.

   [;Introduction]; We report a case of Leigh syndrome possessing a c.55C>T, p(P19S) mutation in *NDUFA1*. *NDUFA1* encodes MWFE, which plays an important role in complex 1 assembly, function, and stability. We review the clinical phenotype induced by *NDUFA1* mutations. [;Case Report]; The patient was an 11-year-old boy. He had axial hypotonia and horizontal nystagmus at 5 months. At 10 months, afebrile generalized seizure was observed and the plasma lactate (44.9 mg/dL) and pyruvate levels (2.07 mg/dL) were elevated. At 17 months, brain magnetic resonance imaging (MRI) revealed bilateral lesions in the substantia nigra and globus pallidus; moderate intellectual disability and hearing difficulty were noted. He walked independently at 2 years 6 months. At 5 years, a muscle biopsy showed mild type 2B fiber atrophy. Fibroblast complex I enzymatic activities were 27.6% of control. A single missense mutation (P19S) in the X-linked *NDUFA1* gene was detected. Oral administration of vitamin B_1 and coenzyme Q_{10} began at age 3; consequently, the lactate and pyruvate levels decreased and the MRI lesions diminished. At 7 years, his ataxia was aggravated, resulting in difficulty sitting and walking; brain MRI showed new lesions in the bilateral putamen and medulla. Although his ataxia and muscle weakness gradually improved, at 10 years, he presented central apnea, intention tremor, and dysphagia. Brain MRI showed cerebellar atrophy and symmetric T2 hyperintensities in the cervical cord. After 1 year, his weakness mildly improved. [;Conclusion]; *NDUFA1* is reportedly involved in the assembly and function of complex☐by interacting with other subunits encoded by mitochondrial DNA. Four previously reported patients and our case with *NDUFA1* mutations presented with Leigh syndrome and prominent infratentorial lesions, cerebellum and brainstem. Moreover, our case showed spinal cord involvement.

2527W

Complex I assembly factor *TMEM126B* **mutations result in muscle weakness and isolated complex I deficiency.** *B. Ruzzenente[1], L. Sanchez-Caballero[2], L. Bianchi[1], Z. Assouline[3], G. Barcia[3], M.D. Metodiev[1], M. Rio[3], B. Funalot[4], M.A.M. van der Brand[2], S. Guerrero–Castillo[2], J.P. Molenaar[5], D. Koolen[6], U. Brandt[2], R.J. Rodenburg[2], L.G. Nijtmans[2], A. Rotig[1].* 1) INSERM U1163, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, 75015 Paris, France; 2) Radboud Center for Mitochondrial Medicine, Department of Pediatrics, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands; 3) Departments of Pediatric and Genetics, Hôpital Necker-Enfants-Malades, 75015 Paris, France; 4) Department of Genetics, Hôpital Henri Mondor, Créteil, France; 5) Department of Neurology, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands; 6) Department of Human Genetics, Radboud University Medical Center, 6500 HB Nijmegen, the Netherlands; Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6500 HB Nijmegen, the Netherlands.

   Mitochondrial complex I deficiency results in a plethora of often severe clinical phenotypes presenting already in early childhood. Here we report three complex I deficient adult subjects with relatively mild clinical symptoms, including isolated, progressive exercise induced myalgia and exercise intolerance but with normal later development. Exome sequencing and targeted exome sequencing revealed compound heterozygous mutations in *TMEM126B*, encoding for a complex I assembly factor. Further biochemical analysis of subject fibroblasts revealed a severe complex I deficiency caused by a defective assembly. Lentiviral complementation with the wild-type cDNA restored the complex I deficiency, demonstrating the pathogenic nature of these mutations. Further complexome analysis of one subject indicated that the complex I assembly defect occurred at the stage of its membrane module assembly. Our results show that TMEM126B defects can lead to complex I deficiencies and interestingly, symptoms could only occur after exercise.

2528T

Biallelic mutations in *TMEM126B* **cause severe complex I deficiency with a variable clinical phenotype.** *R.W. Taylor[1], C.L. Alston[1], A.G. Compton[2,3], L.E. Formosa[4], V. Strecker[5], M. Oláhová[1], T.B. Haack[6,7], J. Smet[8], K. Stouffs[9], P. Diakumis[10], E. Ciara[11], D. Cassiman[12], N. Romaine[13], L. He[1], B. De Paepe[8], A.V. Vanlander[8], S. Seneca[9], R.K. Feichtinger[14], R. Płoski[15], D. Rokicki[16], E. Pronicka[11,16], R.G. Haller[13,17], J.L.K. Van Hove[12,18], M. Bahlo[10,19], J.A. Mayr[14], R. Van Coster[8], H. Prokisch[6,7], I. Wittig[5,20,21], M.T. Ryan[4], D.R. Thorburn[2,3].* 1) Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, United Kingdom; 2) Murdoch Childrens Research Institute and Victorian Clinical Genetic Services, Royal Children's Hospital, Melbourne, VIC 3052, Australia; 3) Department of Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia; 4) Department of Biochemistry and Molecular Biology, Monash Biomedicine Discovery Institute, Monash University, Clayton Campus, Melbourne, VIC, 3800, Australia; 5) Functional Proteomics, SFB 815 Core Unit, Goethe-Universität, Theodor-Stern-kai 7, Haus 26, Frankfurt am Main, 60590 Frankfurt am Main, Germany; 6) Institute of Human Genetics, Technische Universität München, 81675 München, Germany; 7) Institute of Human Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany; 8) Department of Pediatrics, Division of Pediatric Neurology and Metabolism, Ghent University Hospital, De Pintelaan 185, B-9000 Ghent, Belgium; 9) Center for Medical Genetics, UZ Brussel, Research Group Reproduction and Genetics, Vrije Universiteit Brussel (VUB), B-1090 Brussels, Belgium; 10) Population Health & Immunity Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC 3052, Australia; 11) Department of Medical Genetics, The Children's Memorial Health Institute, 04-730, Warsaw, Poland; 12) Metabolic Center, University Hospitals Leuven, 3000, Leuven, Belgium; 13) Neuromuscular Center, Institute for Exercise and Environmental Medicine, Texas Health Presbyterian Hospital, Dallas, TX 75231, USA; 14) Department of Pediatrics, University Hospital Salzburg, Paracelsus Medical University, 5020, Salzburg, Austria; 15) Department of Medical Genetics, Warsaw Medical University, 02-106, Warsaw, Poland; 16) Department of Pediatrics, Nutrition and Metabolic Diseases, The Children's Memorial Health Institute, 04-730, Warsaw, Poland; 17) Department of Neurology and Neurotherapeutics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA; 18) Department of Pediatrics, University of Colorado, Aurora, Colorado 80045, USA; 19) Department of Medical Biology, University of Melbourne, VIC 3052, Melbourne, Australia; 20) Cluster of Excellence "Macromolecular Complexes", Goethe-Universität, 60590 Frankfurt am Main, Germany; 21) German Center for Cardiovascular Research (DZHK), Partner site RheinMain, Frankfurt, Germany.

Complex I deficiency is the most common biochemical phenotype observed in individuals with mitochondrial disease. With 44 structural subunits and over 10 assembly factors, it is unsurprising that complex I deficiency is associated with clinical and genetic heterogeneity. Massively parallel sequencing (MPS) technologies including custom, targeted gene panels or unbiased whole exome sequencing (WES) are hugely powerful in identifying the underlying genetic defect in a clinical diagnostic setting, yet many individuals remain without a genetic diagnosis. These cases may harbor mutations in poorly understood or uncharacterized genes, and their diagnosis relies upon characterization of these orphan genes. Complexome profiling recently identified TMEM126B as a component of the mitochondrial complex I assembly (MCIA) complex alongside the ACAD9, ECSIT, NDUFAF1 and TIMMDC1 proteins. We describe the clinical, biochemical and molecular findings in six cases of mitochondrial disease from four unrelated families with biallelic *TMEM126B* variants at c.635G>T, p.(Gly212Val) or c.401delA, p.(Asn134Ilefs*2). We provide functional evidence to support the pathogenicity of these *TMEM126B* variants, with evidence of founder effects for both variants, and establish defects within this gene as a cause of complex I deficiency in association with either pure adult-onset myopathy, or in one case, a severe multisystem presentation (chronic renal failure and cardiomyopathy) in infancy. Functional experimentation including viral rescue and complexome profiling of subject cell lines confirms TMEM126B as the tenth complex I assembly factor and validates the importance of both genome wide sequencing and proteomic approaches in characterizing novel disease-associated genes whose physiological roles were previously undetermined.

2529F

DNM1L-**related mitochondrial fission defect presenting as spastic paraparesis and optic atrophy.** *M. Tesarova[1], T. Danhelovska[1], M. Rodinova[1], V. Stranecky[2], J. Sladkova[1], A. Vondrackova[1], J. Krizova[1], J. Zamecnik[3], T. Honzik[1], H. Hansikova[1], J. Zeman[1].* 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; 3) Institute of Pathology and Molecular Medicine, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic.

An increasing group of mitochondrial disorders is caused by mutations in nuclear genes affecting mitochondrial dynamics and function. *DNM1L* gene encodes the dynamin 1- like protein DNM1L (Drp1) which is crucial in the mitochondrial fusion process as well as fusion of peroxisomes. We describe a patient with novel heterozygous *de novo DNM1L* mutation c. 176C>T (p.Thr59Ile), which further expands the associated clinical spectrum. The patient exhibits axial hypotonia, spastic paraparesis, neo- and paleo-cerebellar syndrome and optic atrophy. Muscle biopsy revealed mild decrease in CI+III activity, increased SDH and decreased COX activity staining in 5% of muscle fibers. Furthermore, regular occurrence of "mega-mitochondria" along elongated mitochondrial network was found in cultured myoblasts which confirms impaired mitochondrial dynamic. Novel mutation c. 176C>T (p.Thr59Ile) in *DNM1L* gene was identified by exome sequencing and affects highly conserved Thr59 in the GTPase domain of the protein. Trp59 has been previously shown to be indispensable for GTPase reaction since it is involved in the positioning of the catalytic water molecule and Mg^{2+} coordination (Wenger et al 2013). All previously described missense mutations in 3 patients with varied phenotype were localized in the middle domain that is important for the self-assembly of the protein and its oligomerization. To conclude, our results further expand clinical phenotypes (including optic atrophy) associated with *DNM1L* mutations that is typical for mitochondrial disorders due to altered mitochondrial dynamics. *Supported by research projects GAČR 14-36804G, RVO-VFN64165/2012 and AZV 16-32341A.*

2530W

Recurrent *de novo* dominant mutations in *SLC25A4* cause severe early-onset mitochondrial disease and loss of mitochondrial DNA copy number. *K. Thompson[1], H. Majd[2], C. Dallabona[3], K. Reinson[4,5], M.S. King[2], C.L. Alston[1], L. He[1], T. Lodi[3], S. Jones[6], A. Fattal-Valevski[7], N. Fraenkel[8], A. Saada[8], P. Isohanni[9,10], R. Vara[11], I.A. Barbosa[12], M.A. Simpson[12], C. Deshpande[13], P.E. Bonnen[14], R.J. Rodenburg[15], A. Suomalainen[9], K. Ounap[4,5], O. Elpeleg[8], I. Ferrero[3], R. McFarland[1], E.R.S. Kunji[2], R.W. Taylor[1].* 1) Wellcome Trust Centre for Mitochondrial Research, Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK; 2) The Medical Research Council, Mitochondrial Biology Unit, Cambridge Biomedical Campus, Wellcome Trust / MRC Building, Hills Road, Cambridge, CB2 0XY, UK; 3) Department of Life Sciences, University of Parma, Parco Area delle Scienze 11A, Parma 43124, Italy; 4) Department of Pediatrics, University of Tartu, Tartu, Estonia; 5) Department of Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia; 6) Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, St Marys Hospital, Oxford Road, Manchester, M13 9WL, UK; 7) Paediatric Neurology Unit, "Dana" Children Hospital, Tel Aviv Sourasky Medical Centre, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 8) Monique and Jacques Roboh Department of Genetic Research, Hadassah, Hebrew University Medical Center, Jerusalem, 91120, Israel; 9) Research Program of Molecular Neurology, Biomedicum-Helsinki, University of Helsinki, Helsinki, Finland; 10) Department of Child Neurology, Children's hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; 11) Department of Paediatric Inherited Metabolic Diseases, Evelina Children's Hospital, London, UK; 12) Division of Genetics and Molecular Medicine, King's College London School of Medicine, London, UK; 13) Clinical Genetics Unit, Guys and St Thomas' NHS Foundation Trust, London, UK; 14) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 15) Nijmegen Centre for Mitochondrial Disorders, Department of Paediatrics, Translational Metabolic Laboratory, Radboud University Medical Centre, Nijmegen, the Netherlands.

Mutations in *SLC25A4* encoding the mitochondrial ADP/ATP carrier AAC1 are a well-recognized cause of mitochondrial disease. Several heterozygous *SLC25A4* mutations cause adult-onset autosomal dominant progressive external ophthalmoplegia associated with multiple mitochondrial DNA deletions, whereas recessive *SLC25A4* mutations cause childhood-onset mitochondrial myopathy and cardiomyopathy. Here, we describe the identification by whole exome sequencing of 7 probands harboring dominant, *de novo SLC25A4* mutations. All affected individuals presented at birth, were ventilator-dependent and, where tested, revealed severe combined mitochondrial respiratory chain deficiencies associated with a marked loss of mitochondrial DNA copy number in skeletal muscle. Strikingly, an identical c.239G>A, p.(Arg80His) mutation was present in 4 of the 7 subjects, whilst the other 3 cases harbored the same c.703C>G, p.(Arg235Gly) mutation. Western blot analysis of skeletal muscle revealed a marked decrease of AAC1 protein levels and loss of respiratory chain components. By expressing recombinant mutant AAC1 in *Lactococcus lactis* we showed that both mutations severely impair ADP/ATP transport. Structural analysis suggests the mutations most likely affect the substrate binding and mechanics of the carrier, respectively. Both affected residues point inwards and thus cannot be involved in dimerization of AAC1, excluding this as an explanation of the dominant phenotype. The decreased capacity for transport likely affects mitochondrial DNA maintenance and in turn respiration, causing a severe energy crisis. The confirmation of the pathogenicity of these *de novo SLC25A4* mutations highlights a third distinct clinical phenotype associated with mutation of this gene and demonstrates that early-onset mitochondrial disease can be caused by recurrent *de novo* dominant mutations, which has significant implications for the application and analysis of whole exome sequencing data in mitochondrial disease.

2531T

The results of the *GJB2* gene copy-number analysis in the patients "homozygous" for the *GJB2* mutation. *E.A. Bliznetz, A.V. Polyakov.* Federal State Budgetary Institution "Research Centre for Medical Genetics", 115478 Russian Federation, Moscow, Moskvorechie street, 1.

Although the mutations in the *GJB2* gene sequence make up the majority of variants causing autosomal recessive congenital nonsyndromic hearing loss, few large deletions have been shown to contribute to DFNB1 deafness. Currently, clinical DFNB1 testing for nonsyndromic hearing loss patients includes *GJB2* sequencing and DFNB1 deletion analysis for two common large deletions (*GJB6*-D13S1830 and *GJB6*-D13S1854) in the patients, who carry one or no mutation in the *GJB2* gene sequence, by using ALFP analysis with deletion-specific breakpoint primers or copy-number analysis. While the *GJB6*-D13S1830 and *GJB6*-D13S1854 deletions lie upstream of *GJB2* and don`t affect *GJB2*, two mutations of 920 kb and 101 kb in size, published by Feldmann et al. (2009) and Bliznetz et al. (2014) respectively, directly affect the *GJB2* gene by deletion of the entire *GJB2* gene. These deletions have been observed to cause a false homozygosity of the *GJB2* mutation in compound heterozygous patients or the lack of the *GJB2* gene sequence in patients homozygous for the large deletion. We proposed that there are other large deletions of the *GJB2* gene, concealed by "homozygosity" of the *GJB2* point mutation. Here we investigated copy number changes of sequences in the *GJB2* and *GJB6* genes in 182 unrelated Russian patients, homozygous for the *GJB2* mutation. DNA from their parents was not tested. No copy number changes were detected by using the SALSA® MLPA® test (MRC-Holland). We assume that unidentified large deletions of the *GJB2* gene are extremely rare.

2532F

TMTC2 **as a candidate gene responsible for sensorineural hearing loss in children.** *H. Guillen Ahlers[1], M.J. Montoya[1], F. Chevalier[1], C. Erbe[2], C.L. Runge[2], M. Olivier[1].* 1) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Otolaryngology and Communication Sciences, Medical College of Wisconsin, Milwaukee, WI.

Hearing loss affects 1 in 1000 newborns, and 5 in 1000 children 3-17 years old. Even though it is estimated that half of those cases arise from genetic causes, only a limited number of genes have been identified, and most characterized mutations affect very few individuals. We recently investigated a large family of Northern European descent with sensorineural hearing loss. The hearing loss in all affected family members is characterized as bilateral, symmetric, progressive SNHL reaching severe to profound loss in childhood. Audiometric configurations demonstrated a characteristic dip at 1000-2000 Hz. All affected family members wear hearing aids or have undergone cochlear implantation. We discovered the fully penetrant variant rs35725509 in the *TMTC2* gene [MIM 615856] as the likely cause for hearing loss in this family. The nonsynonymous mutation leads to a Val to Ile substitution at position 381 of *TMTC2*, a gene that potentially affects the normal nerve function in the inner ear. The rs35725509 variant is found in the general population (0.8%), suggesting that it may affect hearing in individuals beyond the initial family. To determine whether rs35725509 or additional mutations in the *TMTC2* gene can be identified in other children affected by sensorineural hearing loss, DNA from 7 children with severe sensorineural hearing loss and their immediate family members was collected for exome sequencing. The rs35725509 variant was found in both affected members in one of the families, further suggesting the relevance of this mutation. The rs35725509 variant was the only one identified within the *TMTC2* gene, no additional mutations in the gene were uncovered in these families. A better understanding of genetic causes of hearing loss in children is essential to devise novel treatment options for this debilitating disorder.

2533W

Exome sequencing and bioinformatic analysis of Louisiana Acadian families with extensive family history of Usher syndrome, isolated retinitis pigmentosa, and multigenic intrafamilial mutations. *D. Mercer[1], A. Umrigar[1], A. Musso[2], A. Hurley[2], F. Tsien[1].* 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Dept. of Communication Disorders, School of Allied Health Professions, Louisiana State University Health Sciences Center, New Orleans, LA.

We are conducting a study of Louisiana Acadian families with extensive history of hearing and vision loss. Usher syndrome is an autosomal recessive condition characterized by hearing loss and progressive blindness due to retinitis pigmentosa (RP). It accounts for approximately half of all cases of concurrent deafness and blindness in adults. Though rare, Usher syndrome is more common in certain ethnic populations, including the Louisiana Acadians. The most common mutation in the Acadian population is the USH1C 216G>A mutation. Isolated cases of RP can also occur independently from Usher syndrome. We present two families with variable degrees of hearing loss and retinitis pigmentosa. 1.) Family A is affected by both autosomal recessive Usher syndrome and autosomal dominant isolated RP. The proband is clinically diagnosed with isolated RP and is a heterozygous carrier for the common USH1C mutation. The proband's brother is also affected with isolated RP with no USH1C mutation detected. The proband's two maternal half-brothers are homozygous for the USH1C G>A mutation revealed by PCR and targeted sequencing. An extensive family history spanning five generations revealed eight Usher syndrome affected individuals on the maternal side of the family, while most members of the paternal side of the family were clinically diagnosed with isolated RP. Exome sequencing and bioinformatic analysis revealed mutations of the PRPF3 and RHO genes responsible for the RP in the affected individuals. 2.) The proband from family B is homozygous for the common USH1C G>A mutation. At the current age of 18 years, he does not yet exhibit significant vision loss. Maternal family history spanning six generations identified 3 family members with Usher syndrome showing signs of RP in their 30s, which is a much later age than typically seen for Usher syndrome Type 1; and 8 family members in their 20s with deafness only. Exome sequencing and bioinformatic analysis revealed the presence of both USH1C and USH2A mutations in the proband. The present study demonstrates the prevalence of these rare mutations in isolated communities such as the Louisiana Acadians. In addition, we found multigenic intrafamilial mutations in the Louisiana Acadian families from our study. Genetic awareness in these communities and diagnosis are essential in the prognosis, genetic counseling, and treatment of these families.

2534T

The subtleties of CNVs and the power of targeted genomic enrichment in their detection. *A.O. Taylor[1], D.L. Kolbe[1], C. Sloan-Heggen[1,2], C.J. Nishimura[1], K.L. Frees[1], E.A Black-Ziegelbein[1], K.T. Booth[1], S.S. Ephraim[1], P.T. Ranum[1], D. Wang[1], A.E. Weaver[1], H. Azaiez[1], R.J.H. Smith[1,2].* 1) Otolaryngology, The University of Iowa, Iowa City, IA; 2) Molecular Physiology & Biophysics, The University of Iowa College of Medicine, Iowa City, IA.

Over 90 genes have been implicated in non-syndromic hearing loss (NSHL), making targeted genomic enrichment and massively parallel sequencing (TGE+MPS) an ideal technology by which to provide comprehensive genetic testing as a clinical tool in the evaluation of deaf and hard-of-hearing persons. Here, we illustrate the power of a TGE+MPS panel called OtoSCOPE to detect copy number variation (CNV) using five representative families.   **Case 1:** A 5-year-old Hispanic/Caucasian male with congenital moderate sensorineural hearing loss (SNHL) and hypermetropia underwent computed tomography of the temporal bones, mutation screening of *GJB2/6*, and array comparative genomic hybridization, all of which were uninformative. OtoSCOPE identified homozygosity for the tandem deletion of *STRC-CATSPER2* and a diagnosis of Deafness-Infertility Syndrome was made. **Case 2:** A 7-year-old male was diagnosed with congenital bilateral mild conductive hearing loss. The family history was remarkable for pre-auricular pits in the father, which the child also had, in addition to a second branchial arch sinus tract. OtoSCOPE identified a four-exon deletion in *EYA1* consistent with the diagnosis of branchiootorenal syndrome. **Case 3:** A 5-year-old Irish female was evaluated for congenital severe-to-profound SNHL and a family history remarkable for dominantly inherited maternal hearing loss; the mother also had a white forelock. Prior testing included a Waardenburg panel and whole exome sequencing both of which were non-diagnostic. A small, two-exon, deletion in *SOX10* was detected using OtoSCOPE and Waardenburg syndrome was diagnosed. **Case 4:** A 2-year-old male with congenital bilateral hearing loss was diagnosed with a single exon deletion and a pathogenic missense change in *CDH23* consistent with the diagnosis of USH1D. **Case 5:** A 5-year-old female with congenital profound hearing loss was diagnosed with a two-exon deletion and pathogenic missense variant in *TMC1* consistent with a DFNB7/11 diagnosis.   These cases illustrate the importance of high-resolution CNV detection in deafness. Because CNVs are implicated in up to 20% of deafness, analysis for this type of genetic change must be included in all clinical diagnostics panels used for deafness.

2535F

Identification of *CYP1B1* mutations in families with primary congenital glaucoma. *M. Ansar, B. Gul, F. Hashmi, A. Mehmood, N. Ullah, S. Sajid, E. Ul-lah.* Department of Biochemistry, Faculty of Biological Sciences, Quaid-I-Azam University, Islamabad, 45320, Pakistan.

Primary congenital glaucoma (PCG) is a rare autosomal recessive disorder which generally present during early infancy. The signs and symptoms of PCG include reduced visual acuity, enlargement of globe, blepharospasm, edema, corneal opacification and photophobia. Currently mutations in six genes/loci are known to cause PCG but among these CYP1B1 is the major contributor. For this study we identified seven multiplex Pakistani families which comprised of 24 patients presenting typical clinical features of PCG. Sanger sequencing of CYP1B1 gene in these families identified potentially pathogenic variants in four families, including two novel and two known variants. The identified variants were sequenced in all available samples of each family to confirm the segregation with PCG phenotype. Pathogenicity analysis by bioinformatics tools predicted that all identified variants are disease causing, but one of these (p.Glu229Lys) showed high frequency (0.01423; 41 homozygotes) in the ExAC data. Further sequencing of control DNA samples from various ethnic groups of Pakistan also revealed the high frequency of mutant allele and thus cast doubts on the pathogenic nature of this variant. Our findings indicate that CYP1B1 is the major contributor of PCG in Pakistan and it further highlights the importance of the detailed evaluations before declaring variants as pathogenic.

2536W

Next-generation sequencing-based deafness panel deciphers the genetic etiology of non-syndromic auditory neuropathy spectrum disorder in Han Chinese from Taiwan. *Y.H. Lin[1,2,3], Y.H. Chan[1,3], Y.H. Lin[1,4], C.J. Hsu[1,3], P.L. Chen[2,4,5,6], C.C. Wu[1,5].* 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Otolaryngology, Taichung Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Taichung, Taiwan; 4) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 5) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 6) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Non-syndromic auditory neuropathy spectrum disorder (ANSD) is a common cause of hereditary hearing impairment. Non-syndromic ANSD can be caused by damage to the inner hair cells, synapses, spiral ganglion neurons, auditory nerve, or brainstem auditory nuclei. Corresponding to the heterogeneity in pathology, the rehabilitation outcomes with hearing aids or cochlear implants vary significantly among the patients. Recent studies have revealed that genetic diagnosis is of great help in determining the site of pathology in hereditary hearing impairment. In this study, we performed comprehensive genetic analyses in 40 unrelated ANSD patients of Han Taiwanese ethnicity. All patients underwent a first-tier genetic screening for 3 common deafness genes (*GJB2*, *SLC26A4* and the mitochondrial 12S rRNA gene) and 1 ANSD hotspot in Han Taiwanese (*OTOF* p.E1700Q) using Sanger sequencing. Patients with non-diagnostic genotypes were then subjected to a next-generation sequencing (NGS)-based panel which targets 161 deafness genes, including the 7 documented non-syndromic ANSD genes: *OTOF*, *DFNB59*, *GJB2*, *PCDH9*, *DIAPH3*, *AIFM1*, and the mitochondrial 12S rRNA gene. In total, we confirmed genetic diagnosis in 22 (55%) patients, including 19 with bi-allelic *OTOF* mutations, 2 with bi-allelic *GJB2* mutations, and 1 with homoplasmic m.1555A>G mutation. Of the *OTOF* variants, p.E1700Q was the most prevalent, being homozygous in 8 patients and compound heterozygous with other *OTOF* mutations in 11 patients. We identified 3 novel *OTOF* mutations which had not been reported: p.E841K, c.4961-1G>A, and p.R1735W. Of note, we identified an *OTOF* variant, p.K1992R, which was linked with p.E1700Q on the same haplotype in all the chromosomes of the homozygous and heterozygous patients. As the allele frequency of p.K1992R (0.0006938) in the East Asian population of ExAC was much lower than that of p.E1700Q (0.00743), and p.K1992R was predicted as "probably damaging" by several softwares, we speculated that p.K1992R might be the true mutation leading to ANSD instead of p.E1700Q. We are currently performing immunoprecipitation experiments to validate the pathogenicity of *OTOF* p.K1992R, p.E1700Q, as well as the three novel *OTOF* mutations. In summary, our results demonstrated that *OTOF* mutations are the predominant cause of non-syndromic ANSD in the Taiwanese population, and that targeted NGS is useful in deciphering the genetic etiology of non-syndromic ANSD.

2537T

Resolving the pseudogene mapping complexities of conventional exome sequencing in hearing impairment using 10XGenomics approach. *R. Pellegrino[1], C. Kao[1], F. Mafra[1], J. Garifallou[1], C. Kaminski[1], L. Tian[1], S. Garcia[2], B. Vona[3], T. Haaf[3], H. Hakonarson[1].* 1) The Children´s Hospital of Philadelphia, Philadelphia, PA; 2) 10XGenomics; 3) Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.

Hearing deficiency is a particularly heterogeneous disorder affecting approximately 1 of 1000 newborns every year. Non-syndromic autosomal recessive deafness maps to a number of different non syndromic Hearing Loss and Deafness (DFNB) loci. For example, the DFNB1 locus contains *GJB2* (MIM: 121011) and *GJB6* (MIM: 604418), The stereocilin (*STRC*) gene (MIM: 606440), located in chromosome 15q15.3 within the DFNB16 locus, presents a major mapping challenge for standard next-generation sequencing (STRC) pipelines due to a closely linked pseudogene with 99.6% coding sequence identity. *STRC* also is part of a larger segmental duplication of four genes, *HISPPD2A* (MIM: 610979), *CATSPER2* (MIM: 607249), *STRC*, and *CKMT1A* (MIM: 613415) all of which having mapping issues using current short-read (~200-400 bases) NGS technologies. In fact, ~10% (1,998) of genes in the human exome have highly homologous paralogs (pseudogenes) located elsewhere in the genome that are problematic to map using short read NGS. Genes such as *STRC* are clinically important but largely inaccessible to current NGS and require customized assays to characterize. We use the Chromium instrument (10X Genomics) to generate barcoded reads that identify segments that are immediately contiguous to each other in the genome. The barcoded reads then underwent exome target enrichment using a premium bait design (Agilent SureSelect) containing "bridging baits" spaced within intronic and intergenic regions to preserve long-range information following capture. We subsequently sequenced individuals with hearing loss using the Illumina Hiseq2500 at 85X coverage. We use linked read technology coupled to whole exome sequencing to demonstrate proof-of-principle on samples with clinically validated *STRC* variants/mutations. This approach was able to detect nucleotide variants, indels, and larger structural variants. Importantly, we were also able to accurately phase compound mutations. For example, a suspected compound heterozygous sample with two pathogenic *STRC* variants, c.2303_2313+1del12 and c.5125A>G, were found to be arranged in *cis* instead of *trans*. Thus, linked read sequencing provide an approach to accurately map short NGS reads in *STRC* from its pseudogene and can be used to detect/resolve variants that contribute to congenital hearing impairment.

2538F

MAB21L1 **mutation causes anophthalmia and microphthalmia.** *X. Wu[1,2], C. Liu[1,2], Z. Kang[1,2], Z. Li[1,2], Y. Wang[1,2], L. Li[1,2], L. Peng[1,2].* 1) Research Center for Translational Medicine, Shanghai East Hospital, Tongji University School of Medicine,Shanghai,China; 2) Division of Medical Genetics,Tongji University School of Medicine, Shanghai, China.

   Anophthalmia and microphthalmia (AM) are rare ocular anomalies, characterized by complete absence of the globe or short of a total axial length of a globe. Utilizing whole-exome sequencing, we identify a heterozygous mutation, C.151 C>T, (p.Arg51Trp) in *MAB21L1* gene from patients affected by AM in a dominant heritant family. The *MAB21L1* gene, located at 13q13, encodes a 359-amino acid protein similar to *mab21*, a cell fate-determining factor, in *C. elegans*. We further observed the effect of *mab21l1* on the development of eye in zebrafish. In situ hybridization experiments showed that the *mab21l1* mainly expressed in both eye and midbrain. When knockdown of *mab21l1* using specific morpholino, the development of whole oclular was impaired , but could be really rescued by forced-expression mRNA of either wild-type (WT) *mab21l1* or *mab21l2*. TUNEL staining was also performed to detect the level of apoptosis in zebrafish, and showed that the downregulation of *mab21l1* induced a significantly apoptosis. These results thus implicate a vital role of *MAB21L1* in the pathogenesis of anophthalmia and microphthalmia.

2539W

Genetic causes of hearing loss in Middle Eastern Arab families. *A.I. Abu Rayyan[1], N. Danial-Farran[2], Z. Brownstein[3], L.A. Kamal[1], S. Gulsuner[4], S. Casadei[4], T. Walsh[4], S. Allon-Shalev[2], M.C. King[4], K.B. Avraham[3], M.N. Kanaan[1].* 1) Biological Sciences , Bethlehem University , Bethlehem , Bethlehem , Palestinian Territory; 2) Genetics Institute, Ha'Emek Medical Center, Afula, Israel; Rappaport Faculty of Medicine, Technion - Israel Institute of Technology, Haifa, Israel; 3) Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 4) Department of Medical Genetics and Department of Genome Sciences, University of Washington, Seattle, WA, USA.

   The goal of our project is to discover and characterize genes for inherited hearing loss (HL) by studying highly informative families from the Middle East. We enroll and sample all informative members of families with at least two relatives with HL, which may be dominant or recessive, congenital or later onset, and non-syndromic or syndromic. This report focuses on Arab families from the Palestinian West Bank and northern Israel. For each family, audiologic analysis and primary clinical evaluations were carried out, and genomic DNA was evaluated for mutations in *GJB2* and for several common founder alleles. Families with no such mutations were evaluated by hybridization and multiplexed sequencing, using our panel of all known human and mouse deafness genes (PMID: 21917145). SNVs, indels, and CNVs were identified in exons and proximal regulatory regions. Mutations were interpreted as damaging only if they co-segregated with hearing loss in their families; and were truncating, led to loss of transcript in patient RNA, completely deleted the gene, or were missense mutations with experimentally confirmed functional effect. Of 85 Arab families evaluated by our panel thus far, including 484 participating relatives, 53 families (62%) carried a pathogenic or likely pathogenic mutation in a known deafness gene. A total of 44 different mutations, most not previously reported, were identified in 22 genes and submitted to LOVD and ClinVar. Genes responsible for non-syndromic HL in 41 families were CDH23 (3 families), CLDN14 (3), ESRRB (2), ADGRV1 (2), GPSM2 (1), LOXHD1 (3), MYO15A (7), MYO6 (2), MYO7A (6), OTOA (1), OTOG (1), PCDH15 (1), PTRH2 (1), TBC1D24 (1), TMC1 (1), TRIOBP (3), and USH1C (2); genes for syndromic or other HL phenotypes in 12 families were OTOF (auditory neuropathy, 2), PAX3 (Waardenburg, 3), SLC26A4 (Pendred signs, highly variable across families, 6), and USH1G (RP, 1). The 38% of families with no mutation in any known deafness gene represent a highly informative resource for identification both of additional HL genes and of distant non-coding regulatory mutations. We are evaluating these unresolved families with whole exome and/or whole genome sequencing. We undertake this complete pipeline for approximately 70 new families each year. Supported by NIH R01DC011835 to KBA and MNK.

2540T

Exome sequencing and subsequent haplotype analysis of a Scottish Cone-Rod patient, homozygous for a common "Spanish" *CERKL* mutation, points towards a distant common ancestry. *A.J. Cassidy[1], S.V. Brugnatelli[2], C. Ayuso[3], A. Avila-Fernandez[3], D. Zeller[1], D. Lester[2], The Tayside Centre for Genomic Analysis.* 1) Tayside Centre for Genomic Analysis, University of Dundee, Dundee, Tayside, Scotland; 2) School of Science, Engineering & Technology, Abertay University, Dundee, Scotland, UK; 3) Department of Genetics, Fundacion Jimenez Diaz-CIBERER, Madrid, Spain.

Cone–Rod Dystrophy (CRD) is a subset of a group of retinal dystrophies, known as Retinitis Pigmentosa (RP), which together make up the main cause of inherited adult blindness in the developed world. CRD patients generally experience loss of their cone rich central vision in early adulthood, before also losing their rod rich peripheral vision, leading to complete blindness in later life. The genetic cause of most (>50%) RP cases is as yet unknown, especially for those carrying a recessive mutation. Such information can help design future treatments, e.g. viral gene therapy and also in ~10% of cases, the developing field of premature termination codon therapy. In this study we initially identified a pathogenic variation in CERKL (p.Arg257ter) using the Illumina TruSight sequencing panel in an isolated non-familial 59 year old non-consanguineous, Scottish CRD patient. This particular CERKL mutation was first found in a large Spanish CRD family, which was subsequently shown by haplotype analysis of 4 SNP loci, flanking the CERKL gene, to be distantly related to 6 other Spanish CRD families. Following haplotype analysis with the same 4 SNP markers with the Scottish *CERKL* patient we found that 3 of the closest SNPs used to previously link the 7 Spanish families together were also homozygous and in common. One haplotyping SNP, the most distant from the *CERKL* gene, was however shown to be different from the 7 Spanish families and this occurrence is likely to have come from a recombination event in a distant common ancestor of the Scottish and Spanish patient(s). .

2541F

Functional analyses of novel mutations in HPS (Hermansky-Pudlak Syndrome)-3, HPS-5 and HPS-6 patients with ocular albinism. *W. Li[1,3], A. Wei[2], Y. Yuan[3], D. Bai[4], Z. Zhou[3], J. Ma[1], Z. Hao[1], J. Yu[3], Y. Zhang[2], L. Yang[3], X. Yang[2], L. Li[4].* 1) Center for Medical Genetics, Beijing Children's Hospital, Capital Medical University, Beijing, China; 2) Department of Dermatology, Beijing Tongren Hospital, Capital Medical University, Beijing, China; 3) Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing , China; 4) Department of Ophthalmology, Beijing Children's Hospital, Capital Medical University, Beijing, China.

Hermansky-Pudlak syndrome (HPS) is a rare recessive disorder characterized by hypopigmentation, bleeding diathesis and other symptoms attitutable to multiple defects in lysosome-related organelles (LROs). Currently, 10 HPS subtypes have been identified with mutations in *HPS1* to *HPS10*. Functional studies have revealed that the encoded proteins of the HPS genes are components of different protein trafficking complexes named HPS protein associated complexes (HPACs) such as BLOC-1, BLOC-2, BLOC-3, and AP-3. HPS3, HPS5 and HPS6 protein consist of the BLOC-2 complex. So far there is no case report of HPS-3, HPS-5 or HPS-6 in Chinese population. Using next-generation sequencing, we have identified 1 case of HPS-3, 1 case of HPS-5, and 3 cases of HPS-6 in Chinese HPS patients with typical ocular albinism together with other variable phenotypes, which have been verified by the absence of platelet dense granules under whole-mount platelet EM. The manifested hypopigmentation in the eyes and hair of these HPS patients was similar to their mouse counterparts such as cocoa (*Hps3*), ruby-eye 2 (*Hps5*) and ruby-eye (*Hps6*). All these patients were compound heterozygotes. We verified the mutations by Sanger sequencing and family transmission. Among these mutations, 8 allelles were previously unreported alleles (PUAs), i.e. 4 frameshift mutations (3 in *HPS6*, 1 in *HPS3*), 3 missense mutations (2 in *HPS5*, 1 in *HPS6*), and 1 nonsense mutation (in *HPS6*). To evaluate their effects on BLOC-2 complex stability, we detected HPS6 protein level in platelets. We found that except null in the HPS-3 patient, HPS6 protein was detectable in all the other HPS patients, suggesting that the effects of compromised BLOC-2 functions are tissue-specific and genotype-specific. Our results suggest that next-generation sequencing is a powerful tool for the identification of HPS patients with autosomal recessive ocular albinism, which may be overlooked in Chinese population. **Key Words:** Hermansky-Pudlak syndrome (HPS); biogenesis of lysosome-related organelles complex-2 (BLOC-2); previously unreported alleles (PUAs); lysosome-related organelles (LROs); genotype-phenotype relationship; albinism.

2542W

CRISPR-Cas9 based treatment of myocilin-associated glaucoma. *A. Jain[1], G. Zode[2], R.B. Kasetti[2], F.A. Rann[3], W. Yan[3], K. Bugge[1], C. Searby[1], V. Buffard[1], M.L. Humbert[1], F. Zhang[3], A.F. Clark[2], V.C. Sheffield[1].* 1) Pediatrics, University of Iowa, Iowa City, IA; 2) North Texas Eye Research Institute, University of North Texas Health Science Center, Fort Worth, TX; 3) McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA.

Glaucoma is a leading cause of irreversible blindness with elevated intraocular pressure (IOP) being the most important and the only modifiable risk factor. Mutations in myocilin (*MYOC*) lead to its misfolding and endoplasmic reticulum (ER) stress in the trabecular meshwork (TM), the tissue that maintains aqueous humor outflow and regulates IOP. ER stress and/or death of the TM lead to elevated IOP and glaucoma. Mutations in *MYOC* have been reported in both primary open angle glaucoma and juvenile open angle glaucoma. Since the phenotype seems to result from the mutant misfolded protein rather than resulting from loss of myocilin function or haplo-insufficiency, we propose to relieve the ER stress and glaucoma by targeting the *MYOC* gene using CRISPR-Cas9 genome editing technology. Stable cell lines expressing mutant *MYOC* and human primary TM cells, as well as h*MYOC*[Y437H] transgenic mice were treated with adenovirus Ad5-CRISPR-Cas9 virus with *MYOC* and control guide RNAs. Treated and control cells, and mice were evaluated for ER stress and glaucoma phenotypes. Treatment with CRISPR-Cas9 with gRNAs targeting the *MYOC* gene reduces levels of MYOC and ER stress in *MYOC* mutant cells. We will also report on the rescue of glaucoma, during both disease stage and prophylactically, in the mouse model of *MYOC* glaucoma by CRISPR-cas9. This is the first detailed study showing that genome editing targeting the *MYOC* gene can relieve ER stress in TM cells and in transgenic h*MYOC*[Y437H] mouse model of glaucoma. This approach would provide a long-term or a permanent relief to the stress as compared to no current treatment or proposed short-term strategies in MYOC-associated glaucoma patients.

2543T

Analysis of the *COL8A2* and *SLC4A11* genes and of the TGC intronic expansion in the *TCF4* gene in families with Fuchs Endothelial Corneal Distrophy (FECD). *M.B. Melo[1], C. Martins[1], M.B. Oliveira[2], S.M.S. Costa[1], J.P.C. Vasconcellos[2].* 1) CBMEG, Univ Campinas, Campinas SP, Brazil; 2) Department of Ophthalmology, Univ Campinas, Campinas SP, Brazil.

Background: Fuchs' endothelial corneal dystrophy (FECD) is a hereditary disease of the corneal endothelium, which is characterized by the progressive and slow loss of corneal endothelial cells, thickening of Descemet's membrane and extracellular matrix deposition (*guttae*). Due to the progression and consequent loss of endothelial cells, the removal of fluids is affected, leading to stromal edema and in some cases, to blindness. Several genes are associated with Fuchs' dystrophy, including *COL8A2*, *SLC4A11* and *TCF4*. The main aims of this study were to evaluate the segregation pattern of mutations in the *COL8A2* and *SLC4A11* genes and analyze the familial association between the TGC trinucleotide repeat expansion in the *TCF4* gene in three families with FECD. **Methods:** We analyzed 39 individuals belonging to 3 families with FECD and 100 controls subjects. The *COL8A2* and *SLC4A11* genes were evaluated through Sanger sequencing. The TGC microsatellite in the *TCF4* gene was genotyped using the Short Tandem Repeat-Assay (STR-Assay) and analyzed by prediction programs *in silico*. **Results:** We found a deletion in the *SLC4A11* gene, rs113526162 (c.1172-19_1172-18delTC), co-segregating with the trait in one family, which was absent in controls subjects. *In silico* predictions allowed us to show that this mutation may induce changes in the mutant transcript. STR analysis demonstrated that the expanded allele of the *TGC* repeat also co-segregated with the trait in one family. The penetrance of the expansion in the study was 100%. **Conclusions:** These findings suggest that the deletion has an effect on regulatory splicing factors and might generate an abnormal or non-functional protein, affecting specific tissues in which this protein is expressed. Despite the fact that the expansion is present in only one family, we showed a strong relationship between the *TGC* expansion with FECD. However, further studies are necessary to understand the exact contribution to the disorder.

2544F

Exploring the characteristics of hearing loss and its etiology in a unique subset of families with hereditary deafness. *A. Pandya[1], M. Tekin[2], A.H. O'Brien[1], S.H. Blanton[2], K.S. Arnos[3].* 1) Pediatric Genetics and Metabolism, University of North Carolina, Chapel Hill, Chapel Hill, NC; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 3) Department of Science, Technology, & Mathematics, Gallaudet University, Washington DC.

Hearing loss (HL) is one of the most common neurosensory defects affecting nearly 37 million individuals, with one in a thousand newborns identified with profound congenital HL. Despite marked genetic heterogeneity, studies in population isolates, inbred families and large cohorts have enabled mapping of at least 167 loci and identification of 99 genes for non-syndromic HL. The high frequency of assortative mating amongst the deaf and presence of a Deaf culture have resulted in a unique subgroup of families with deaf parents who have deaf or hearing offspring. This population offers an opportunity to explore the inheritance and clinical characteristics resulting from the presence of multiple rare deafness genes in a single individual due to the phenomenon of gametic phase disequilibrium. We have studied 375 deaf probands from 327 matings where both parents are deaf (DxD), ascertained through the North American repository of DNA from deaf individuals. One hundred ninety seven of these families were non complementary with only deaf offspring, and 79 were segregating with both deaf and hearing children. Screening for mutations in the *GJB2/GJB6* and *MT-RNR1* genes in all probands identified 185 probands to have bi-allelic pathogenic *GJB2* mutations. This represents 49.3% percent who have their HL explained by mutations in *GJB2,* which is significantly higher than the observation in the deaf population at large. Of these, the majority (68%) were homozygous for the common 35delG pathogenic sequence variant. Another 7.2% of the probands were identified with a *GJB6* deletion and a *GJB2* sequence variant as the etiology for their HL. Thirteen percent of the probands are heterozygous carriers for a single pathogenic *GJB2* variant as compared to the 3.5% carrier frequency in the general population. All probands have severe to profound HL. There remain 165 probands without an identified etiology for their hearing loss. We are screening these probands for mutations in known deafness genes, to be followed by analysis of the coding exome with the goal of identifying known and novel genes for HL. The unique structure of these families will allow identification of previously uncharacterized interactions and or modifiers of HL as was the case with *GJB2/GJB6* mutations.

2545W

North Carolina Macular Dystrophy (MCDR1/NCMD): Mutations found affecting PRDM13. *K. Small[1], A. DeLuca[2], R.A. Lewis[3], M. Leys[4], B. Bakall[5], V. Puech[6], B. Puech[6], K. Rohrschneider[7], R. Silva Garcia[1], N. Udar[1], F. Shaya[1], E. Heon[8], C.A. Garcia[9], T.A. Rice[10], G. Fishman[11], J.C. Folk[12], F. Cremers[13], S. Yelchits[14], E.L. Kennedy[2], M.A. Luse[2], E. Stone[2].* 1) Molecular Insight Research Foundation, 501 N. Orange St Suite 250 Glendale, CA 91203; 2) University of Iowa, 4111 MERF Iowa City, IA; 3) Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030; 4) WVU Eye Institute, PO Box 9193 Morgantown, WV; 5) University of Arizona College of Medicine Phoenix 15042 N 49th Way Phoenix, AZ; 6) Service d'Exploration de la vision et Neuro-ophtalmologie CHRU, Lille, France 2 Avenue Oscar Lambret, 59000 Lille, France; 7) University of Heidelberg Im Neuenheimer Feld 400 Heidelberg; 8) Hospital for Sick Children 555 University Ave Toronto, ON; 9) Univ of Texas Houston Med Sch 4704 Montrose Boulevard Houston, TX; 10) Byers Eye Institute, Stanford University School of Medicine, Palo Alto, California; 11) Chicago Lthouse for the Blind & Vis Impaired 1850 W Roosevelt Road Chicago, IL; 12) Univ of Iowa Hospitals & Clinics 200 Hawkins Dr Iowa City, IA; 13) Raboud university medical center Geert Grooteplein 10 Nijmegen; 14) Jules Stein Eye Institute, University of California, Los Angeles, California, USA.

Purpose: Small and co-workers (Ophthalmology, 2016) recently showed that mutations affecting the expression of *PRDM13* cause North Carolina macular dystrophy (NCMD, MCDR1). The purpose of this study was to investigate the PRDM locus in additional families with this phenotype **Methods:** 18 new families with the NCMD phenotype (a total of 46 affected individuals) were studied using a combination of Sanger sequencing of the DNase I hypersensitivity site upstream of *PRDM13* and a PCR-based assay for the detection of a previously described duplication of the PRDM13 gene. IRB approval was obtained and informed consent was obtained from all participants. **Results:** The original NCMD publication reported four different mutations (V1-V4) distributed among eleven different kindreds. To date, we have identified the disease-causing mutations in 10 new independent families the NMCD phenotype. Seven of these harbor the previously described V1 mutation (all from the USA) and Five families harbor the previously described V2 mutation (3 European, 1 USA). These two variants segregated perfectly in the 36 affected and 10 unaffected members of these 12 new NCMD families. No mutations have yet been identified in the other six families. We have an additional 7 families yet to be studied. **Conclusions:** We identified 12 new families with disease-causing mutations in the DNase I hypersensitivity site upstream of the *PRDM13* gene . Additional families with the NCMD phenotype continue to support that these mutations are causative of MCDR1 / NCMD. .

Family #	Family Location	Referring Physician	Origin of Mutation	Chromosomal Position of Mutation
704	W. Virginia	Leys	N Carolina	CHR6: 100040906
705	Texas	Lewis	N Carolina	CHR6: 100040906
715	West Virginia	Leys	N Carolina	CHR6: 100040906
720	Seattle	Schneiderman	N Carolina	CHR6: 100040906
773	Ohio	KWS	N Carolina	CHR6: 100040906
732	Arizona	Bakall	N Carolina	CHR6: 100040906
730	Mexico / NY	Agemi	N Carolina	CHR6: 100040906
714	Belgium	Cremers / Hoynig	French	CHR6:100040987
718	Germany	KWS	French	CHR6:100040987
771	W. Virginia	Leys	French	CHR6:100040987
775	Germany	Rohrschneider	French	CHR6:100040987
731	Riverside, CA	KWS	French	CHR6:100040987

2546T

Massively parallel sequencing identified *KCNQ1* **compound heterozygous mutations caused Jervell and Lange-Nielsen syndrome or nonsymdromic hearing loss in three Chinese families.** *H. Yuan, C. Wang, Y. Lu, J. Cheng.* Medical Genetics Center, Southwest Hospital, Third Military Medical University, Chongqing 400038, China.

   Jervell and Lange-Nielsen syndrome (JLNS) involves congenital profound hearing loss associated with prolongation of the QTc interval in the electrocardiogram (ECG), and predisposes to lethal cardiac arrhythmia, especially under sympathic stimulation during periods of stress, exerciseor fright. As the main responsible gene for JLNS, mutations in *KCNQ1* are also cause of another dominant forms of disease, long QT syndrome type 1 (LQT1), characterized by long QTc without hearing loss. We investigated the underlying genetic defects for 10000+ subjects with congenital profound hearing loss by massively parallel sequencing (MPS). Compound heterozygous mutations of *KCNQ1* were identified in 4 subjects from 3 Chinese families. Besides the common manifestation of congenital profound hearing loss, detailed clinical evaluation presented variations of phenotype among the 4 patients: two siblings showed typical JLNS cardiac phenotype of long QTc greater than 500ms and recurrent syncopal episodes; one subject showed extra-cardiac-and-ear symptoms of iron-deficiency anemia and epilepsy; one case without QTc prolongation and no cardiac complains. Sanger sequencing confirmed the compound mutations of *KCNQ1* inherited from both parents carrying heterozygous mutation. Among the 6 disease-causing mutations, 5 mutations (c.683+5G>A, c.905C>T, c.965C>T, c.1484_1485delCT and c.1831G>A) were previously reported, while c.546C>A mutation was first identified in this study. Besides expanding the spectrum of *KCNQ1* mutations causing JLNS, our results also showed the diversity of JLNS phenotypes, and showed the importance of molecular genetic analysis in confirming the clinical diagnosis and making pre-symptom diagnosis possible.

2547F

Comparison of bioinformatic prediction with molecular analyses of *FOXC1* **mutations from patients with Axenfeld-Rieger syndrome.** *M. Seifi, T. Footz, S.A Taylor, M.A Walter.* Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada.

   Mutations in the gene forkhead box C1 (*FOXC1* [MIM 601090]) cause Axenfeld-Rieger syndrome (ARS [MIM 602482]), an autosomal dominant disorder characterized by malformations of the ocular anterior segment and non-ocular abnormalities. Here, we investigated the effect of four missense variants (p.H128R, p.C135Y, p.M161V and p.T368N) on the structure and function of *FOXC1*, and tested the predictive value of SIFT (sorting intolerant from tolerant) and PolyPhen (polymorphism phenotyping) for all known *FOXC1* missense variants. Our results show that the *FOXC1* p.C135Y variant significantly reduced DNA binding, transactivation, nuclear localization and expression levels of the recombinant protein, and has a shortened protein half-life. The p.M161V reduced transactivation capacity without changing other *FOXC1* functions. The p.H128R variant reduced DNA binding, transactivation, and nuclear localization. Molecular modeling of the FOXC1 forkhead domain predicted that p.H128R changes FOXC1's structure. In contrast, *FOXC1* p.T368N was indistinguishable from wild type *FOXC1* in all tests, consistent with being a rare, non-disease causing variant. The molecular investigations of these 4 and the 27 previously studied *FOXC1* missense mutations were compared to predictions from the SIFT and PolyPhen prediction tools. SIFT and PolyPhen correctly identified as pathogenic only the *FOXC1* mutations with significant functional defects. Our study demonstrates that the p.H128R, p.C135Y and p.M161V *FOXC1* missense variants cause ARS via different mechanisms. The variant p.T368N is likely a non-disease causing substitution. Importantly, our results support the use of SIFT and PolyPhen as reliable tools to predict the pathogenicity of *FOXC1* forkhead domain missense variants.

2548W

Pharmacogenetic variants in *TPMT* **alter cellular responses to cisplatin** *in vitro*. A.P. Bhavsar[1,2], E.P. Gunaretnam[2,3], Y. Li[2,3], B.C. Carleton[2,3], C.J.D. Ross[1,2]. 1) Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, British Columbia, Canada; 2) Child & Family Research Institute, BC Children's Hospital, Vancouver, British Columbia, Canada; 3) Department of Pediatrics, Faculty of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada.

Background: Cisplatin is a highly-effective and widely-used chemotherapeutic agent that also has a high incidence of drug-induced ototoxicity, especially in children. Pharmacogenomic studies of drug absorption, distribution, metabolism and excretion genes identified low-activity variants in thiopurine methyltransferase (*TPMT*) as strong predictors of the development of cisplatin-induced ototoxicity. It has been suggested that low-activity TPMT variants may lead to elevated levels of its substrate *S*-adenosylmethionine, which has been shown to exacerbate cisplatin toxicity, however direct studies of cisplatin-*TPMT* interactions have not been reported. In this work we sought to establish this drug-gene interaction in an *in vitro* cell based system. Methods: Epitope-tagged *TPMT**3A, *3B and *3C mammalian expression clones were generated by site-directed mutagenesis from the *TPMT**1 wild type clone. Expression of TPMT variants was monitored by immunoblotting following transient transfection into HEK293T cultured cells. To examine how the TPMT*3A variant affected cellular cisplatin response, *TPMT* expression clones were transiently transfected into murine inner ear cell lines, HEI-OC1 or UB/OC-1, and cell viability was assessed using an MTT assay. In addition, the expression of a cisplatin-response gene was monitored by quantitative PCR (qPCR) following transient transfection of HeLa cells with *TPMT* expression clones. Where required, gene silencing was performed using specific siRNA reagents. Results: Quantification of TPMT variant levels in cell culture showed that TPMT*3A was present at >20-fold lower amounts than the wild type, in agreement with published reports. This correlated with a significantly increased susceptibility to cisplatin cytotoxicity as the cisplatin IC_{50} in both HEI-OC1 and UB/OC-1 cells decreased approximately 30% upon expression of *TPMT**3A compared to wild type *TPMT*. Similarly, a sensitive cisplatin-response biosensor showed significantly higher induction in the presence of TPMT*3A compared to wild type TPMT. Taken together, these data suggest that TPMT*3A significantly sensitizes cells to cisplatin response and toxicity. Conclusions: This study has identified cellular cisplatin responses that are affected by *TPMT* variant status. We demonstrated an *in vitro* interaction between cisplatin and *TPMT**3A that supports the previously reported pharmacogenetic association with cisplatin-induced ototoxicity.

2549T

Mutations in the *PCYT1A* **gene are responsible for isolated forms of early-onset retinitis pigmentosa.** S. Banfi[1,2], F. Testa[3], M. Filippelli[3], R. Brunetti-Pierri[3], G. Di Fruscio[3], V. Di Iorio[3], M. Pizzo[2], N. Brunetti-Pierri[4], F. Simonelli[3]. 1) TIGEM, Fondazione Telethon, Pozzuoli, Naples, Italy; 2) Department of Biochemistry, Biophysics and General Pathology, Second University of Naples, Naples, Italy; 3) 1Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy; 4) Department of Translational Medicine, Federico II University, Naples, Italy.

Background Mutations in *PTCY1A* have been recently linked to two different phenotypes: one characterized by spondylometaphyseal dysplasia and cone-rod dystrophy (SMD-CRD) and the other by congenital lipodystrophy, severe fatty liver disease, and reduced HDL cholesterol without any retinal or skeletal involvement. Isolated retinal involvement due to *PTCY1A* mutations has not been reported thus far. **Methods and Results** By next generation sequencing targeted to genes associated with inherited retinal diseases (IRD), we identified pathogenic mutations in the *PCYT1A* gene in three young patients with isolated retinal dystrophy from two different Italian families. Ophthalmological assessment (best corrected visual acuity, color vision testing, retinography, optical coherence tomography and standard electroretinogram) showed the presence of Early Onset Retinitis Pigmentosa (EORP). A thorough clinical evaluation of the patients, with whole skeleton X-ray, metabolic assessment and liver ultrasound failed to reveal signs of skeletal dysplasia, metabolic and hepatic alterations. **Conclusions** This is the first report showing that the *PCYT1A* gene can be responsible for isolated forms of EORP thus further expanding the phenotypic spectrum induced by mutations in this gene. .

2550F

***EYS* knockout in zebrafish leads to mislocalization of cone opsins and progressive retinal degeneration.** *M. Liu[1], Z. Lu[1], X. Hu[1], Z. Tang[1,2].* 1) College of Life Science and Technology, Huazhong University Science and Technology, Wuhan, Hubei, China; 2) Key Laboratory of Molecular Biophysics of Ministry of Education, Department of Genetics and Developmental Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, PR China.

Retinal degeneration (RD) is the leading cause of human vision loss. Based on different clinical symptoms, RD is divided into retinits pigmentosa (RP), cone rod dystrophy (CRD), cone dystrophy and macular degeneration. Eyes shut homolog (*EYS*) is the largest gene known to be expressed in the human eye, spanning more than 2Mb within the *RP25* locus. *EYS* mutations have been reported to cause both RP and CRD, but its physiological function in photoreceptors and how *EYS* mutations lead to progressive retinal degeneration is still unclear. To reveal the underlying roles of *EYS*, we generated an *EYS* mutant zebrafish line using TALEN technology. HE staining showed that the retina ONL in *EYS* mutant zebrafish became thinner from the 3rd month, and degenerated in an age dependent way. Furthermore, TEM results showed that the number of photoreceptors decreased at the 3rd month, and decreased more significantly at the 10th month. The red cone opsins and UV cone opsins appeared to be mislocalized on the 2nd month. The outer segment of blue cones, green cones and rods also showed progressive degeneration, but their opsin didn't show mislocalization. Taken together, we built a vertebrate *EYS* mutant model for the first time, and found that red cone opsins and UV cone opsins mislocalizeed in the photoreceptors, which may be one of the major causes of photoreceptor degeneration.

2551W

In-frame indels in inherited eye disease and how structural modelling can help estimate their relative pathogenicity. *P.I. Sergouniotis[1,2], S.J. Barton[3], S. Waller[3], R. Perveen[3], J.M. Ellingford[2,3], C. Campbell[3], G. Hall[3], R.L. Gillespie[2,3], S.S. Bhaskar[3], S.C. Ramsden[3], G.C. Black[2,3], S.C. Lovell[4].* 1) Manchester Royal Eye Hospital, Manchester, UK; 2) Centre for Ophthalmology & Vision Sciences, University of Manchester, Manchester, UK; 3) Centre for Genomic Medicine, University of Manchester, Manchester, UK; 4) Department of Computational & Evolutionary Biology, University of Manchester, Manchester, UK.

Purpose: Although small in-frame insertions/deletions (indels) are common, the molecular mechanisms and frequency by which they give rise to disease phenotypes remain largely unknown. The aim of this study is to provide insights into the role of in-frame indels (≤ 21 nucleotides) in two genetically heterogeneous eye disorders. **Methods**: 486 probands with retinal dystrophy and 181 probands with childhood cataracts underwent multigene panel testing in a clinical diagnostic laboratory. In-frame indels were collected and evaluated both clinically and *in silico*. Variants that could be modeled in the context of protein structure were identified and analysed using integrative structural modeling. **Results**: 55 small in-frame indels were detected; 17 of these changes were novel to this study and 18 variants were reported clinically. A reliable model of the corresponding protein sequence could be generated for 8 variants. Structural modeling indicated a diverse range of molecular mechanisms of disease including disruption of secondary and tertiary protein structure and alteration of protein-DNA binding sites. **Conclusions**: In retinal dystrophy or childhood cataract subjects, one small in-frame indel is clinically reported in every ~37 individuals tested. The clinical utility of computational tools evaluating these changes increases when the full complexity of the involved molecular mechanisms is embraced.

2552T

Expanding the RASopathies: Germline mutations in SYNGAP1. *K.A. Rauen, L. Bivina. Department of Pediatrics, MIND Institute, University of California Davis, Sacramento, CA.*

The RASopathies are a defined group of medical genetic syndromes caused by germline mutations in genes that encode components or regulators of the Ras/mitogen-activated protein kinase (MAPK) pathway. These disorders currently include neurofibromatosis type 1 (NF1), Noonan syndrome, Noonan syndrome with multiple lentigines, capillary malformation–arteriovenous malformation (CM-AVM) syndrome, Costello syndrome, cardio-facio-cutaneous syndrome and Legius syndrome. Because of the common underlying Ras/MAPK pathway dysregulation, the RASopathies exhibit numerous overlapping phenotypic features. The Ras/MAPK pathway plays an essential role in regulating the cell cycle and cellular growth, differentiation, apoptosis and senescence, all of which are critical to normal development. RasGAPs (GTPase-activating proteins) are negative regulators of the Ras/MAPK pathway. Haploinsufficiency in neurofibromin (encoded by the *NF1* gene) and p120 RasGAP (encoded by the *RASA1* gene) are both examples of RasGAPs and cause NF1 and CM-AVM syndromes, respectively. SYNGAP1 (synaptic Ras GTPase activating protein 1) is a more recently described RasGAP thought to be selectively expressed only in neurons. Germline mutations in the *SYNGAP1* gene have been described and associated with Autosomal Dominant Intellectual Disability Type 5, a non-syndromic intellectual disability that has not been thought of as part of the RASopathies. Here we present two individuals with pathogenic germline nonsense mutations in SYNGAP1 identified through exome sequencing and further expand upon the SYNGAP1 phenotype. We present this as a new RASopathy based on the identification of causative mutations in this new RasGAP gene with neuronal-restricted expression understanding that it has a unique neurologically-centered clinical phenotype which does overlap phenotypically with other RASopathies. .

2553F

Loss of the protein arginine methyltranserase PRMT7 causes severe intellectual disability, facial dysmorphism, microcephaly, short stature, brachydactyly and seizures. *D. Chitayat[1,2], K.D. Kernohan[3], A. McBride[3], Y. Xi[3], N. Martin[1], J. Schwartzentruber[4,5], D.A. Dyment[3,6], K.M. Boycott[3,6], S. Blaser[7]. 1) The Prenatal Diagnosis , Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Canada M5G 1X8; 3) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada K1H 8L1; 4) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada H3A 0G1; 5) Department of Human Genetics, McGill University, Montreal, Quebec, Canada H3A 1B1 5Department of Genetics, Children's Hospital of Eastern Ontario, 401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1; 6) Department of Genetics, Children's Hospital of Eastern Ontario, 401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1; 7) Department of Diagnostic Imaging, Hospital for Sick Children, University of Toronto, Toronto, Canada M5G1X8.*

The human proteome is incredibly complex and highly regulated to enable proper cellular functioning. Post translational modifications (PTMs) play a vital role in proteome regulation; to date, more than 150 PTMs have been described. The most abundant modifications include phosphorylation, glycosylation, ubiquitination, nitrosylation, acetylation, bromination and methylation. Each PTM has potential to affect three dimensional protein structure, enzymatic activity, protein stability, protein-protein interactions and protein-nucleic acid interactions. One such modification is the covalent addition of a methyl group to arginine or lysine residues, which is used to regulate a substantial proportion of the proteome. Arginine and lysine methylation are catalyzed by protein arginine methyltransferase (PRMTs) and protein lysine methyltransferase proteins (PKMTs), respectively; each methyltransferase has a specific set of target substrates. We report a male with severe intellectual disability, facial dysmorphism, microcephaly, short stature, brachydactyly, cryptorchidism and seizures who was found to have a homozygous 25,309 bp deletion encompassing the transcription start site of *PRMT7*, which we confirmed is functionally a null allele. We show that the patient's cells have decreased levels of protein arginine methylation, and that affected proteins include the essential histones, H2B and H4. Finally, we demonstrate that patient cells have altered Wnt signaling, which may have contributed to the skeletal abnormalities. Our findings confirm the recent disease association of *PRMT7*, inform the phenotypic understanding of this disorder and provide insight into the molecular pathogenesis of this new condition.

2554W

Youngest surviving patient diagnosed with *MAGEL2* mutation (Schaaf-Yang syndrome, OMIM #615547). *V. Neerukonda[1], B. Marr[2], I. Cherrick[3], Z. Powis[4], W. Alcaraz[4], R.R. Lebel[1].* 1) Development, Behavior, and Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Crouse Memorial Hospital, Neonatal Intensive Care Unit. Syracuse, NY; 3) Department of Pediatrics, SUNY Upstate Medical University, Syracuse, NY; 4) Ambry Genetics, Aliso Viejo, CA.

This male infant had congenital macrocephaly, delivered by Cesarean section at 38 weeks of gestation to a 21 year old G2P0010>1011 female whose pregnancy was complicated by third trimester maternal respiratory distress. Paternal age was 24. The neonatal period was notable for respiratory insufficiency requiring intubation, ventilator support, and an extended NICU stay. The genetics service was consulted on day three of life for multiple craniofacial dysmorphic features and Arthrogryposis Multiplex. Brain Magnetic Resonance Image (MRI), Echocardiogram, and karyotype were normal. Exam revealed macrocephaly, hypertelorism, down slanting palpebral fissures, micrognathia, webbed neck, micropenis, tapering digits, hypotonia, and Arthrogryposis Multiplex. He developed Diabetes Insipidus, Adrenal Insufficiency, Failure to thrive, and became tracheostomy and ventilator dependent. Brain MRI at fifteen months revealed a dysplastic corpus callosum, hypoplastic vermis, myelin dysplasia and global atrophy. The microarray was normal. Whole exome sequencing revealed a heterozygous *de novo* frame shift mutation c.3122delT 9 (p.V1041Afs*7) in the *MAGEL2* gene resulting in a truncated protein. *MAGEL2* is located inside the Prader Willi Syndrome (PWS) critical region of 15q11-13 along with four other genes (*MKRN2, NDN, NPAP1,* and *SNURF-SNRPN*) implicated in PWS phenotype. *MAGEL2* encodes the melanoma-antigen-subfamily-like-2 protein and is involved in cell cycle regulation, neuronal signal transduction and neurite growth, by preventing proteasome-mediated protein degradation. Many animal models note a high degree of homology in the coding region of *MAGEL2* between mice and humans. *MAGEL2* null mice develop a PWS phenotype. *MAGEL2* is expressed predominantly within the developing central nervous system parenchyma, especially in localized areas like the hypothalamus that are implicated in the PWS phenotype. We are aware of thirteen reported persons, nine of whom are alive, with a *MAGEL2* mutation (7 *de novo* mutations and 2 cases of gonadal mosaicism). The age range at diagnosis spans from *in utero* to early adulthood. We propose this is the youngest known patient to be both diagnosed and survive, after being investigated for a PWS like phenotype and neurologic deficits, most likely due to mutation in the *MAGEL2* gene. This patient confirms and extends the distinct phenotype of Schaaf-Yang syndrome (OMIM #615547).

2555T

Isolated hereditary congenital ptosis can be caused by defects of craniofacial development. *S.A. Di Gioia[1], W.M. Chan[1], B. Barry[1], A. Magli[2], T. de Berardinis[2], P. Pavone[3], E.C. Engle[1].* 1) Department of Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA; 2) Università degli Studi di Napoli, Dipartimento di Scienze Oftalmologiche, Naples, Italy; 3) Unit of Pediatrics, University Hospital OVE-Policlinico, Catania, Italy.

To uncover genetic causes of congenital ptosis, or drooping of the upper eyelid, we analyzed exome sequence of affected and selected unaffected members of 12 unrelated multiplex families segregating isolated hereditary congenital ptosis. Genome wide SNP data (Illumina OmniExpress array) was also generated for a subset of the families for both linkage and CNV analyses. Using these approaches, we identified three novel variants in three unrelated Italian families, none of which were annotated in the EXAC database. Remarkably, the variants were in three different genes, each of which had been previously identified as important for craniofacial development. Affected members of Pedigree UU, diagnosed with a mild form of unilateral ptosis, harbored a novel *FGFR2* variant, c.760C>T, p.H254Y. Mutations in *FGFR2* are typically associated with syndromic craniosynostosis. Affected members of Pedigree IQ, with variable expression of ptosis, harbored a novel missense variant, c.379G>A p.A127T, in the helix-loop-helix domain of *TWIST1*. *TWIST1* mutations are typically associated with Saethre-Chotzen syndrome, a second form of craniosynostosis. There were no signs or symptoms of craniosynostosis detected by the physicians who referred these families. Lastly, Pedigree TQ was diagnosed with severe congenital bilateral ptosis suggestive of blepharophimosis, ptosis and epicanthus inversus syndrome (BPES). CNV analysis revealed the presence of a heterozygous deletion on the short arm of chromosome 2 (chr2:45024863-43300029) that segregated with affection status, and that removed the entire *SIX2* gene. A deletion of *SIX2* with similar breakpoints was recently reported in a novel form of frontonasal dysplasia with clinical features similar to Pedigree TQ. Together, these data expand the phenotype of known craniosynostosis genes and confirm the *SIX2* phenotype. Moreover, they suggest that the presence of ptosis should alert clinicians to the possibility of variants in genes that may be more commonly associated with more severe forms of craniofacial developmental disorders.

2556F

Further delineation of Au-Kline Syndrome: A craniosynostosis and Kabuki-like syndrome due to loss of function mutations and deletions of *HNRNPK.* *P.Y. Au[1,2], C. Goedhart[1,2], M. Kerr[1,2], M. Ferguson[3], J. Breckpot[4], K. Devriendt[4], K.J. Wierenga[5], E.A. Fanning[5], U. Kini[6], H. Stewart[6], L. Lange[7], A. Pagnamenta[7], J. Taylor[7], D. Keays[8], D.K. Grange[9], S.L. Dugan[10], J.S. Parboosingh[1,2], A.D. Kline[3], A.M. Innes[1,2], Care for Rare Canada Consortium.* 1) Medical Genetics, University of Calgary, Cumming School of Medicine, Calgary, Alberta, Canada; 2) Alberta Children's Hospital Research Institute for Child and Maternal Health, University of Calgary, Alberta, Canada; 3) Harvey Institute for Human Genetics, Department of Pediatrics, Greater Baltimore Medical Center, Baltimore, Maryland; 4) Center for Human Genetics, KU Leuven, Leuven, Belgium; 5) Department of Pediatrics, Section of Genetics, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma, USA; 6) Department of Clinical Genetics, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; 7) National Institute for Health Research Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 8) Research Institute of Molecular Pathology, Vienna, Austria; 9) Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, Missouri, USA; 10) Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City, Utah.

We recently reported 2 patients with intellectual disability and dysmorphic features with *de novo* loss of function mutations in *HNRNPK* (heterogeneous nuclear ribonucleoprotein K). With 2 additional previously reported deletion cases, this is now known as Au-Kline syndrome (AKS). Lange et al. [2016] reported a further case with a Kabuki-like presentation and nodular heteropia. Through international collaboration, 4 additional patients with *HNRNPK* mutations, and a patient with 9q21.32 microdeletion encompassing *HNRNPK,* have now been identified, resulting in a current total of 9 known patients with AKS. Frameshift and splice mutations have been found in 6 patients identified by WES, and deletion encompassing *HNRNPK* in 3 patients identified by aCGH. Loss of function intolerance of *HNRNPK* is supported by a high pLI score of 1.0 [ExAC et al. 2015]. AKS is characterized by striking dysmorphism including long facies, long ears, a deeply grooved tongue, connective tissue anomalies and aortic dilation, and intellectual disability. Further clinical delineation is now possible with this larger cohort. Features of AKS are shared not only with TGFβ pathway related disorders, but also with Kabuki syndrome (KS). The AKS facial gestalt includes long palpebral fissures, ptosis and shallow orbits, resembling KS. Additionally, >50% of AKS patients have craniosynostosis with at least one suture affected, typically the metopic and sagittal sutures. AKS should therefore also be considered in the differential for syndromic craniosynostosis. Of note, in previously reported KS cohorts ascertained prior to availability of molecular testing, both craniosynostosis and bifid/fissured tongue have been reported [Armstrong et al. 2005], suggesting that some Kabuki-like patients may have *HNRNPK* mutations. The prenatal presentation of AKS includes increased nuchal translucency and/or fetal hydrops. Although haploinsufficiency of *Hnrnpk* in mice is associated with susceptibility to hematopoietic neoplasm [Gallardo et al. 2015], no AKS patients have had malignancy. However, the eldest patient is only 18 years of age. AKS is likely not as rare as initially expected, with several individuals rapidly identified through clinical WES (or reanalysis of WES data) since the publication of our original patients. As the features of AKS are striking, this syndrome is clinically recognizable. Therefore, Sanger sequencing of *HNRNPK* in patients suspected to have an AKS phenotype is currently underway.

2557W

Phenotypic and mutational spectrum of patients with overgrowth, macrocephaly and Weaver-like features. *A.S.A. Cohen[1,2], Y. Shen[3], J. Denny[4], S.J.M. Jones[1,3,5], W.T. Gibson[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Child and Family Research Institute, Vancouver, BC, Canada; 3) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada; 4) Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 5) Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada.

In late 2011, we and others discovered that constitutional mutations in *EZH2* cause Weaver syndrome (WS). WS is a rare overgrowth disorder characterized by tall stature, large head, advanced bone age, facial dysmorphism, intellectual disability and cancer susceptibility. Since then, we have collected samples on almost 70 patients with Weaver-like features from across the world. We have also collected detailed phenotypic information where available. We have tested these patients for mutations in *EZH2*, as well as *NSD1*, which when mutated causes a similar overgrowth disorder called Sotos syndrome. With this detailed phenotyping and targeted gene testing approach, we were able to reach a DNA-level diagnosis in close to 25% of our cohort. Furthermore, we were able to expand on the phenotypic spectrum of WS to include neuronal migration disorders. In addition, using whole-exome sequencing in a select subset of undiagnosed patients, we identified a rare *de novo* mutation in *EED* in one patient. Sanger sequencing was then carried out in the patients with the greatest phenotypic overlap, and a second patient was found to carry a different rare *de novo* mutation in *EED*. Interestingly, EED is the main partner of EZH2 within the Polycomb Repressive Complex 2, which maintains gene silencing. With two independent individuals having similar rare phenotypes associated with different *de novo* mutations in the coding sequence of the same (relatively small) gene, we believe the association to be causal and that we have identified a novel overgrowth gene. We will present an overview of this Weaver-like cohort, including frequencies of the main phenotypic traits observed and a list of all the mutations detected in *EZH2*, *NSD1* and *EED*.

2558T

Persistent thrombocytopenia, growth retardation and developmental delay in a three-year-old female: Possible Forsythe-Wakeling syndrome. *M. Hajianpour[1], M. Popescu[2].* 1) East Tennessee State University, Department of Pediatrics/ Medical Genetics, Quillen College of Medicine, Johnson City, TN; 2) East Tennessee State University, Department of Pediatrics, Hematology/ Oncology, St. Jude's Tricities Affiliate Clinic, Johnson City, TN.

Proband is a 3 y.o. female with persistent thrombocytopenia, developmental delay/ language deficit, growth retardation, mild body asymmetry, keratinized, purple, nodular skin rash on the face and exposed areas and dysmorphic features, such as narrow bifrontal diameter (mild), positional plagiocephaly (normal skull CT scan), frontal bossing, slightly posteriorly rotated left ear, low nasal bridge, bulbous nose, thin vermilions, micrognathia, pointed chin, hypoplastic nipples, and mild fifth fingers clinodactyly. She was born at 29 weeks of gestation by C-section delivery due to maternal pre-eclampsia/fetal distress/IUGR to a 20 y.o., G2, P1 mother. Her BW was 1050 g with a BL of 33 cm. Developmental milestones and speech were delayed. She developed thrombocytopenia at age 2 years. She had episodes of low bicarb with normal renal ultrasound. Chromosome analysis, microarray, breakage study, sister chromatid exchange, and molecular studies for Shwachman-Diamond anemia were negative. No ROH was detected on CMA. Two bone marrow biopsies showed normal histology. The Russell-Silver syndrome (RSS) was considered. But presence of thrombocytopenia and typical skin rash together with lack of cafe-au-lait spots do not support RSS. Growth retardation, dysmorphic features, typical skin rash, and thrombocytopenia were reported in a single Pakistani family (consanguineous marriage) with four affected siblings (Forsythe et al., 2009). They all had microcephaly (one with HC of 9%ile), postnatal growth retardation, and global developmental delay with language deficit and dysmorphic features, such as frontal bossing, sunken eyes, high nasal bridge, and large, prominent, low-set ears. Three siblings developed persistent thrombocytopenia and steroid- nonresponsive nephrotic syndrome (SUNS), resulting in death in two children. One child did not have nephrotic syndrome or thrombocytopenia by age 25 months. The oldest child with thrombocytopenia, developed SUNS at age 7 y, with a renal biopsy showing glomerulosclerosis, tubular atrophy and other abnormalities. None had photosensitivity, poor vision, or deafness, but two had keratinized, purple, nodular skin lesions. Fibroblasts showed intermediate recovery of RNA synthesis following UV irradiation. Dysmorphic features and fibroblasts findings were thought to be reminiscent of Cockayne syndrome. A 27-Mb interval of shared homozygosity on chromosome 1p33-p31.1 (rs2354462 to rs718883) in these patients may indicate a related gene locus.

2559F

Dubowitz syndrome: A literature review and prospective whole exome sequencing study confirm high clinical and genetic heterogeneity. *A. Innes[1,2], T. Hartley[3], B. McInnes[2], P.Y. Au[1,2], A.E. Chudley[4], C. Clericuzio[5], C.R. Greenberg[4], J.L. Lauzon[1], R.B. Lowry[1], J. Moilanen[6], M. Nowaczyk[7], L. Penney[8], S.L. Sawyer[9], K.M. Boycott[3,9], D.A. Dyment[3,9], Care4Rare Canada Consortium.* 1) Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Canada; 2) Alberta Children's Hospital Research Institute, Calgary, Canada; 3) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada; 4) Departments of Pediatrics and Child Health and Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Canada; 5) Pediatric Genetics, University of New Mexico, Albuquerque, NM; 6) Department of Clinical Genetics, Oulu University Hospital, Medical Research Center and PEDEGO Research Unit, University of Oulu, Oulu, Finland; 7) Departments of Molecular Medicine & Pathology and Pediatrics, McMaster University, Hamilton, Canada; 8) Department of Pediatrics, Dalhousie University, Halifax, Canada; 9) Department of Genetics, CHEO, Ottawa, Canada.

The Dubowitz Syndrome (DS OMIM 222370) was first reported in 1965. Including a previous definitive review (Tsukahara and Opitz, 1996), nearly 200 cases have been reported in the literature. While typical clinical features include microcephaly, sparse hair, eczema and mild craniofacial and developmental anomalies, a wide variety of clinical features have been reported. Autosomal recessive inheritance has been postulated, and some patients appear predisposed to malignancy. Despite the apparent frequency of this condition, no common recurrent genetic cause has been identified. Recent small studies have suggested clinical and genetic heterogeneity may explain this paradox. To further explore this possibility we embarked on a detailed literature review of the 50 patients published from 1996 to present and performed whole exome sequencing on 10 trios with probands diagnosed by a clinical geneticist with DS. The literature review underscored both clinical and genetic heterogeneity. For 39 of the 50 patients, photographs were available and a wide variety of facial features and clinical problems were noted. Consanguinity and/or familial recurrence were noted in 8 and 6 families respectively. Strikingly while zero of 19 patients reported from 1996-2005 had an underlying genetic abnormality identified, 16/31 patients (52%) reported since 2006 had a presumed causal variant identified on genetic testing (nearly all via aCGH or WES). While there are almost no recurrent abnormalities; causes that will be discussed include biallelic mutations in *LIGIV* and *NSUN2*, as well as certain CNVs. *De novo* dominant mutations in several genes were also seen. Four patients developed a solid tumor and others had bone marrow failure, increased chromosome breakage was frequent in this group. These literature findings were reproduced in our WES cohort. Of 10 patients, four had different definitive mutations identified (biallelic *BRCA1* and *SCL35A1* mutations, de novo *ARID1B* mutation and *HDAC8* deletion) and in another a *de novo* variant in *CREBBP* is being studied further. For the other five patients no clear causal variants were identified, although novel candidate genes are currently being pursued via global matchmakers. Nevertheless no recurrent genes were seen across multiple patients. The implications of these findings for clinicians considering a clinical diagnosis of DS will be reviewed, including recommendations for further investigation, management and counselling.

2560W

Craniosynostosis in RASopathies. *N. Okamoto[1], K. Ueda[1], M. Yaoita[2], T. Niihori[2], Y. Aoki[2].* 1) Dept Medical Genetics, Osaka Med Ctr/Res Inst, Osaka, Japan; 2) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan.

   Noonan syndrome (NS), Costello syndrome, and cardio-facio-cutaneous (CFC) syndrome are characterized by a distinctive facial appearance, heart defects, ectodermal abnormalities and intellectual disability. These syndromes are caused by a dysregulation of the RAS/mitogen activated protein kinase (MAPK) signaling pathways, and are collectively called "RASopathy". Fibroblast grout factor receptors (FGFR) is interacted with the RAS/MAPK signaling pathways. FGFR proteins are receptor tyrosine kinase upstream of the RAS/MAPK signaling pathway. Genomic mutations in FGFR are the most common causes of inherited craniosynostosis (CS), Apert, Crouzon, Pheiffer and Muenke syndrome. These syndromes are caused by the dysregulated FGFR signaling cascade. Recently CS with *KRAS* mutations was reported. Approximately 10% of published cases with *KRAS* mutations have CS [Addissie et al., AJMG 2015]. The review suggested that individuals with NS associated with *KRAS* mutations have greater risk for CS. In addition, Noonan syndrome caused by *SHOC2* mutation has been reported to have severe CS [Takenouchi, et al., AJMG 2014]. These clinical findings support the idea that RAS/MAPK pathway might be associated with the development of cranium and have interaction with *FGF*. We present 9 patients with CS who had *KRAS, BRAF*, and *PTPN11* mutations. This is the first report of CS with *BRAF* and *PTPN11* mutations. We suggest that patients with CFC syndrome and NS should be followed carefully for cranial deformation. In RASopathy patients with cranial deformities or neurological symptoms due to suspected increased intracranial pressure or microcephaly, further assessment, including 3D-CT of the head, might be necessary to evaluate for possible CS. Proper treatment including surgical intervention is necessary. .

2561T

Williams-Beuren syndrome: Case report and three-year follow-up.
G. Rodríguez-González[1,2], A. Marín-Medina[1,2], S.A. Alonso-Barragán[1,2], A. Ramírez-Velasco[1,2], L. Gutiérrez[2], M.G. Domínguez-Quezada[1,2], I.P. Dávalos-Rodríguez[1,2]. 1) Doctorado en Genética Humana, IGH, CUCS, Universidad de Guadalajara; 2) CIBO, UMAE, HP, CMNO; IMSS.

   Williams-Beuren Syndrome (WBS) (MIM # 194050) is caused by a hemizygous deletion of 1.5 to 1.8 Mb in 7q11.23 The SWB is characterized by short stature, dysmorphic facial features, mental retardation, cardiovascular disorders, infantile hypercalcemia and sociable personality. It has a frequency of 1/10000 to 1/20000 births. Objective: To present a male patient whose clinical data led us the diagnosis of Williams-Beuren syndrome. Case report: Male patient, 4 years old, product of second pregnancy, full-term, obtained by caesarean section. Mother of 27 years, father of 32 years, at birth, non consanguineous. Sister of 7 years old, apparently healthy. After 15 days of extra uterine life he was hospitalized due to pneumonia, at one year old left orquidopexy was done. Delay of psychomotor skills development. At 15 months, cardiological evaluation reported right branch pulmonic stenosis and at two years old ophthalmologic evaluation reported convergent strabismus. At physical examination revealed a sociable personality, short stature, normocephalic skull, epicantus, upward slanting palpebral fissures, strabismus, flat nasal bridge, large mouth, thick lips and low set ears. In chest presence of systolic murmur in pulmonary focus grade II, remaining unchanged. Previous clinical data allowed the diagnosis of SWB. Cytogenetic analysis: 46, XY [16]. ish (7) (q11.23q11.23) (ELN -) [5] Conclusions: In this case the presence of short stature, characteristic facies, pulmonary stenosis and social behavior allowed the clinical diagnosis of WBS, Cytogenetic study reported 7q11.23 deletion. Early clinical diagnosis of SWB enables a multidisciplinary monitoring and early treatment of complications.

2562F

Molecular genetics of focal facial dermal dysplasias types 3 and 4. *B. Lee[1,2], I. Nazarenko[1], L. Edelmann[1], F. Morice-Picard[3], A.M. Slavotinek[4], R.J. Desnick[1]. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 3) Centre de Référence des Maladies Rares de la Peau, Service de Dermatologie et Dermatologie Pédiatrique. Hôpital Pellegrin-Enfants, Bordeaux, France; 4) Division of Genetics, Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA.*

The Focal Facial Dermal Dysplasias (FFDDs) are a group of rare developmental disorders which have been classified into four subtypes based on: 1) the location of the congenital "scar-like" facial lesions in the bitemporal (FFDD Types 1-3) or in the peri-auricular (FFDD Type 4) regions, 2) presence of additional facial abnormalities, and 3) pattern of inheritance. Although the genetic etiologies for FFDD Types 1 and 2 have not been delineated, we recently identified homozygous recessive mutations in *TWIST2*, a transcriptional factor involved in craniofacial mesenchymal development, and a 1.3 Mb copy number duplication or triplication at chromosome 1p36.22p36.21 causing FFDD Type 3, (Tukel et al. *Am. J. Hum. Genet.* 2010), and homoallelic recessive *CYP26C1* mutations underlying FFDD Type 4 (Slavotinek et al. *Hum. Mol. Genet.* 2014). Here, we report the clinical spectrum and causative mutations in 22 unrelated FFDD Type 3 patients and six unrelated FFDD Type 4 patients. Among the FFDD Type 3 patients, *TWIST2* mutations and the 1p36 copy number duplication or triplication were each identified in five FFDD Type 3 patients, respectively, (combined total of 10/22 or 46%). Among the five FFDD Type 4 patients, all were found to have *CYP26C1* mutations including three novel mutations, IVS5+1G>T, IVS5+2insT, p.Arg77Pro. Currently, efforts are underway to identify the causative genes underlying the remaining 12 FFDD Type 3 patients. The interaction of TWIST2 with other genes, especially located in the 1p36.22p36.21 region, was assessed by *in silico* analysis. Two genes, *DHRS3* and *CYP26A1,* were suggested as candidate genes that interacted with TWIST2. Therefore, the 4.2 Mb *TWIST2* regulatory region and the coding regions of *DHRS3* and in *CYP26A1* were sequenced, and several variants with unknown significance were identified, and their potential pathogenic effects are being investigated. In addition, whole genome and exome sequencing has been undertaken in a consanguineous family and are now being analyzed. In conclusion, the results of our study indicate that FFDDs are a group of disorders that are genetically heterogeneous and additional causative genetic defects will be identified in the future.

2563W

Assessment and management of feeding difficulities in patients with trisomy 13 syndrome. *E. Nishi[1,2], N. Fueki[3], Y. Gomi[3], C. Seki[3], M. Arakawa[1], K. Takano[1,2], Y. Fukushima[2], T. Kosho[1,2]. 1) Department of Medical Genetics, Nagano Children's Hospital, Azumino, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Rehabilitation, Nagano Children's Hospital, Azumino, Japan.*

Trisomy 13 (T13) syndrome is a most common autosomal trisomy syndrome, characterized by multiple congenital anomalies, severe developmental delay, and a short life span with the 1-year survival rate as 5–10 % and the median survival time as 7 days according to a large-scale population-based study. Feeding difficulties in patients with T13 remain one of the challenges of their management. In one study, 53% of children with T13 had bottle or breast feeding during the first few weeks or months [Baty et al.,1994]. To evaluate the management of swallowing and feeding in patients with T13, we reviewed detailed clinical data of 4 patients with full T13 in the outpatient clinic specialized for patient with swallowing and feeding problems in Nagano Children's Hospital, from 2012 to 2016. Patients (3 boys, 1 girl), aged from 2 years old to 6 years old, were diagnosed with full T13. All had multiple congenital defects and medical complications including congenital heart defects, cleft lip or cleft palate, and holoprocencephaly. Three of them needed tracheostomy, and two required mechanical ventilation. They were discharged home at the age from two months to 1 year old. They all showed feeding difficulties, required for a nasal or oral gastric tube from infancy. Three needed gastrostomy, but the other accomplished bottle feeding and ceased to use tube feeding at age 7 month. Further evaluation through videofluoroscopic (VF) swallow studies revealed details of their feeding and swallowing problems, including abnormal oral sensation, uncoordinated tongue movement, dyscoordination between sucking and swallowing, inappropriate sizes and forms of the food, and presumed risk of aspiration. In conclusion, the current study, though the sample size is small, suggests a wide variation in oral feeding in patients with T13. Also, detailed and comprehensive evaluation including VF would contribute to safe oral feeding. The evidence is helpful for clinicians to offer the best information on management options to families of patients with T13.

2564T

Mutations in *LZTR1* can underlie rare cases of autosomal recessive Noonan syndrome. *J.J. van der Smagt, N.E. Verbeek, K.L.I. van Gassen.* clinical genetics, UMCU, Utrecht, Utrecht, Netherlands.

Noonan syndrome (NS) is a quite common autosomal dominant syndrome characterized by a recognizable constellation of dysmorphic features, heart defects and short stature. Activating mutations in many genes of the RAS-MAPK pathway have been demonstrated to be causative. In 2000 the existence of an autosomal recessive form of NS indistinguishable from classical autosomal dominant NS was suggested(1), but subsequently this has never been confirmed. In 2015 heterozygous LZTR1 mutations were shown to cause Noonan syndrome in a small autosomal dominant pedigrees (2). No role in the RAS-MAPK pathway had been known for LZTR1. Germline loss of function mutations in LZTR1 had been described in the disorder schwannomatosis (3). Two unrelated index patients with Noonan syndrome are described, that were mutation negative for known Noonan genes but were either homozygous for a missense variant or compound heterozygous for a splice mutation and a missense variant in LZTR1, on whole exome screening. In both cases carrier parents were normal on physical and cardiac evaluation. Both indexes had sibs with the same LZTR1 genotype who died either during pregnancy (fetal hydrops) or shortly after birth (AVSD+ hypertrophic cardiomyopathy +neural tube defect).So these sibs with an identical LZTR1 genotype had at least symptoms consistent with NS. In neither family schwannomatosis has occurred. Discussion: Albeit very rare autosomal recessive Noonan syndrome does occur an can be caused by homozygous or compound heterozygous LZTR1 mutations.The lztr1 protein excusively localizes to the Golgi apparatus (4), linking this subcellular structure to NS. The loss of function mutation in one of our cases suggests that LZTR1 may have a primarily inhibitory effect on RAS MAPK signaling, just like NF1. Interestingly loss of function mutations in both genes are also associated with nerve sheath tumors. Refs: 1.)Ineke vander Burgt and Han Brunner. Genetic Heterogeneity in Noonan syndrome.. Am J Med Genet 200;94:46-51 2.)GL Yamamoto, et al. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. J Med Genet 2015;52:413–421. 3.)A Piotrowski, et al. Germline loss-of-function mutations in *LZTR1* predispose to an inherited disorder of multiple schwannomas. Nat Genet. 2014 February ; 46(2): 182–187. 4.)TG Nacak, et al. The BTB-kelch Protein LZTR-1 Is a Novel Golgi Protein That IsDegraded upon Induction of Apoptosis. Journal of biological chemistry 2006;281:5065-71.

2565F

ARCN1 **Mutations cause a novel craniofacial syndrome due to COPI-mediated transport defects.** *K. Izumi[1,2,3], M. Brett[4], E. Nishi[3], S. Drunat[5,6], E.S. Tan[7], K. Fujiki[2], S. Lebon[5], B. Cham[7], K. Masuda[2], M. Arakawa[3], A. Jacquinet[8], S.T. Chen[2], A. Verloes[5,6,9], Y. Katou[2], P. Gressens[5,9,10], R. Foo[11], S. Passemard[5,6,9], E.C. Tan[4,12], V. Ghouzzi[5,9], K. Shirahige[2].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Research Center for Epigenetic Disease, Institute for Molecular and Cellular Biosciences, The University of Tokyo; 3) Division of Medical Genetics, Nagano Children's Hospital; 4) KK Research Centre, KK Women's and Children's Hospital; 5) INSERM UMR1141; 6) Département de Génétique, Hôpital Robert Debré; 7) Genetics Service, Department of Paediatrics, KK Women's and Children's Hospital; 8) Département de Génétique, CHU et Université de Liège; 9) Université Paris Diderot Paris 7, Hôpital Robert Debré; 10) Center for Developing Brain, King's College, St. Thomas' Campus; 11) Genome Institute of Singapore; 12) Paediatrics ACP, SingHealth Duke-NUS Medical School.

Cellular homeostasis is maintained by the highly organized cooperation of intracellular trafficking systems including COPI, COPII, and clathrin complexes. COPI is a coatomer protein complex responsible for intracellular protein transport between the endoplasmic reticulum (ER) and the Golgi apparatus. Using exome sequencing, we identified four individuals with heterozygous mutations in *ARCN1*, which encodes the coatomer subunit delta of COPI. All individuals with *ARCN1* mutations have a highly conserved clinical phenotype including facial dysmorphisms, severe micrognathia, rhizomelic shortening, short stature, joint laxity, microcephaly, and mild developmental delay. *ARCN1* mutant cell lines demonstrated an ER stress response, suggesting the involvement of ER stress response in the pathogenesis of the *ARCN1*-related syndrome. The clinical overlap between *ARCN1*-related syndrome and Stickler syndrome prompted us to evaluate the role of COPI-mediated transport in collagen transport, and defective intracellular collagen transport was demonstrated in cell lines with *ARCN1* dosage reduction. *ARCN1*–related syndrome represents a novel clinically recognizable syndrome. Our findings demonstrate the importance of COPI-mediated transport, which is a fundamental biological process, in human development.

2566W

Computer-assisted facial recognition of Angelman syndrome: A possible diagnostic aid. *L. Bird[1,2], L. Wolf[3].* 1) Dysmorphology Genetics, Rady Children's Hospital, San Diego, CA; 2) University of California San Diego; 3) Tel Aviv University, Tel Aviv, Israel.

Angelman syndrome (AS) is characterized by severe intellectual disability, seizures, limited speech and distinct behavioral profile. Absence of a functional copy of the paternally-imprinted *UBE3A* gene on chromosome 15 causes AS. 75% of AS is due to deletion of chromosome 15q11.2; the remainder is due to paternal uniparental disomy (UPD), imprinting defect (ID) or mutation of the maternally-inherited *UBE3A*. Subtle facial dysmorphism (midface recession, prognathism, broad mouth) can be appreciated in older AS individuals, but most AS children are generally non-dysmorphic. Automated facial recognition technology has the potential to assist in syndrome identification. Face2Gene (F2G) is a free tool that uses Facial Dysmorphology Novel Analysis (FDNA) technology to evaluate 2D facial photographs. We compared F2G's accuracy of AS recognition with that of clinical experts. The FDNA system was trained using frontal images from 60 individuals (mostly children) with molecularly diagnosed AS (46 with deletion, 9 with UPD or ID, 5 with *UBE3A* mutations). 20 experts were provided a compilation of frontal photographs of 25 children, some of whom had molecularly proven AS (n=10) and the rest having other syndromes (n=15), and were asked to designate which of the children had AS. (Experts did not know the actual number of AS images among the 25.) Frontal images of 117 additional unique molecularly-diagnosed AS patients were used to determine the accuracy of the FDNA technology and to determine which syndromes are most often erroneously identified. Clinical experts and F2G had comparable accuracy, sensitivity and specificity on the 25-image challenge set. Deletion (n=6) and non-deletion AS cases were equally recognized by experts and F2G. AccuracySensitivitySpecificityAccuracy FemalesAccuracy MalesAccuracy WhiteAccuracy non-WhiteExperts72%60%81%80%67%73%72%F2G76%50%*93%82%71%79%73%*sensitivity increased to 80% if AS appearing on the top 10 match list was considered a correct response F2G accuracy (AS #1 match on list) on the expanded set of images was 42/117 = 36%; if AS appeared in the top 5, 10 or 20 matches, sensitivity increased to 62%, 75% and 82%, respectively. The most common other condition on the top 10 match list was Pitt-Hopkins syndrome, which occurred in 33/117. These results show that automated facial recognition could assist clinicians in syndrome recognition. Sensitivity and specificity are likely to improve as the image collection increases.

	Accuracy	Sensitivity	Specificity	Accuracy Females	Accuracy Males	Accuracy White	Accuracy non-White
Experts	72%	60%	81%	80%	67%	73%	72%
F2G	76%	50%*	93%	82%	71%	79%	73%

2567T

Therapeutic trial with an HDAC inhibitor in Rubinstein-Taybi syndrome. *D. Lacombe[1], M.P. Baudier[1], E. Bestaven[2], S. Fraisse[1], E. Taupiac[1], E. Grech[1], I. Guillain[1], P. Fergelot[1], A. Gimbert[3], F. Salvo[3], G. Catheline[4], M. Allard[4], C. Bader[5], N. Moore[6], P. Perez[5], J.R. Cazalet[2], B. Arveiler[1].* 1) Service de Génétique Médicale, CHU Bordeaux; INSERM U1211, Université de Bordeaux, Bordeaux, France; 2) Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, Université de Bordeaux; 3) Direction de la recherche clinique, CHU Bordeaux; 4) Département d'Imagerie Médicale, CHU Bordeaux; 5) Service d'information médicale, USMR, CHU Bordeaux; 6) Service de Pharmacologie Médicale, CHU Bordeaux, Bordeaux, France.

The Rubinstein-Taybi syndrome (RTS) is a multiple congenital anomalies – intellectual deficiency syndrome characterized by growth retardation, microcephaly, characteristic facial features, and broad thumbs and first toes. RTS is genetically heterogeneous with mutations in the *CREBBP* or *EP300* genes encoding transcription cofactors involved in histone acetylation. RTS knock-out mouse studies showed an improvement of long-term memory using histone deacetyltransferase (HDAC) inhibitors. This suggests a therapeutic option for RTS. Among the identified HDAC inhibitors, sodium valproate has the marketing authorization for epilepsy and mood disorders and can go through the blood brain barrier. We conducted an exploratory phase 2 therapeutic trial («proof of concept »), with a primary endpoint to estimate the efficiency of sodium valproate after one year of treatment (30 mg/kg/d) on long term memory in RTS children. Secondary Outcomes included cognitive profiles, brain imaging, motor skills evaluation, and histone acetylation profile. It was a monocentric, double-blind, phase 2 trial, designed as a one-step Fleming with a control group, randomized 2:1 into two parallel groups. 41 RTS cases between ages 6 and 21 participated. Using subtests of neuropsychological battery test especially designed for memory evaluation (point location for CMS, and image recognition for RBMT), results were not significant between the treatment group and the placebo group for the primary outcome. However regarding motor skills for walking, it was noticed a slight decrease of cadence and an increase of the duration of the double step in the valproate group, showing an amelioration in the treatment group. Posturometry showed more oscillations in the placebo group during the task. This confirms the *CREBBP* and *EP300* roles in motor skill learning, and encourages to pursue therapeutic approaches with HDAC inhibitors in RTS.

2568F

Craniofacial phenotype segregates with Loeys-Dietz Syndrome mutation/type. *D.K. Liberton[1], R. Mishra[1], E. Akpa[1], H. Bolan[2], M. Rasooly[2], D. Cantave[3], P.A. Guerrerio[2], J.S. Lee[1].* 1) National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD; 2) National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; 3) Clinical Center, National Institutes of Health, Bethesda, MD.

   Loeys-Dietz Syndrome (LDS) is an autosomal dominant connective tissue disorder with five known types based on their underlying genetic mutation. The main presentation includes arterial tortuosity and aneurysms but there are also craniofacial findings such as hypertelorism, cleft palate, and craniosynostosis. However, there is a large degree of craniofacial phenotypic variability in affected individuals. In this study, we examined patients with Type I (*TGFBR1* mutation) and Type II (*TGFBR2* mutation) who were evaluated at the Craniofacial Dental Clinic at the National Institutes of Health Clinical Center as part of an IRB approved protocol. All patients underwent a cone-beam computed tomography scan (Planmeca Promax 3D) and 3D surface image (3dMD) as well as a comprehensive craniofacial and oral exam. Clinically, both groups demonstrated midface hypoplasia, mandibular retrognathia, and low-set ears. There was variable expression of exophthalmia, hypertelorism, craniosynostosis, drooping nasal tip, cleft palate/bifid uvula/narrow palate and enamel defects. We performed cephalometric analysis on five patients from the two types and compared their craniofacial findings. Both types demonstrated significantly short cranial base, short upper facial height, shortened mandibular base, mandibular retrognathia, and increased gonial-jaw angle. The major morphological difference between the two types was in maxillary and mandibular development. Type I individuals were mildly deficient for measures of maxillary and mandibular growth while Type II individuals were more than two standard deviations below the norm for both measures, indicating significant and severe hypoplasia of the mid- and lower- face in this group. This work expands upon the known craniofacial phenotypes associated with LDS and suggests that there is significant phenotypic variability among and as well as within different mutations. TGFβ signaling is critical for mediating epithelial-mesenchyme interactions, including those important for mandibular, calvarial, and tooth morphogenesis (ref 2). Deep phenotyping and morphometric analysis improves our understanding of the diagnostic features of this disorder and supports the critical role of TGFβ in human craniofacial development. Future research will investigate the developmental mechanisms of these mutations and also examine the craniofacial phenotypes of Type III (SMAD3 mutation), Type IV (TGFB2 mutation), and Type V (TGFB3 mutation) LDS patients.

2569W

Severe oral aversion in three children with rasopathies. *S. Mizuno[1], E. Nishi[2], M. Inaba[1], Y. Muramatsu[1], Y. Aoki[3], C. Hayakawa[1].* 1) Dept Pediatrics, Central Hospital, Aichi Human Service Ctr Pediatrics, Kasugai, Aichi, Japan; 2) Department of Clinical Genetics, Nagano Childrens Hospital, Azumino, Nagano, Japan; 3) Department of Medical Genetics, Tohoku University School of Medicine.

   Cardio-facio-cutaneous (CFC) syndrome (MIM #115150) is a multiple congenital anomaly/mental retardation syndrome characterized by cardiac abnormalities, a distinctive facial appearance, cutaneous abnormalities, and mental retardation. Its phenotype overlaps with both Noonan syndrome and Costello syndrome. These three syndromes belong to a class of genetic syndromes caused by the mutation of genes associated with the RAS/MAPK pathway.CFC syndrome is genetically distinguished by the detection of BRAF, MEK1/2, and KRAS mutation. However the phenotype-genotype correlation of this syndrome has not been fully clarified.

2570T

Application of whole exome sequencing to reverse dysmorphology: Case report of one patient with pathogenic variant in *HDAC8.* *A.B.A. Perez[1], M.P. Migliavacca[1], N.L. Sobreira[2]. 1) Medical Genetics Center, UNIFESP, Sao Paulo, Brazil; 2) Johns Hopkins University - IGM.*

The application of New Generation Sequencing (NGS) for Mendelian Diseases is enabling the delineation of new syndromes that have less distinctive clinical features by a process that can be termed "reverse dysmorphology." Here, the classical practice of identification of the phenotype first and analysis of the genotype second is reversed. In this way, cases are first recognized by their genotype and subsequently the phenotypes of the individuals are compared to the features of the disease assigned to that genotype. The Cornelia de Lange syndrome (CDLS) is a multisystem malformation syndrome recognized primarily on the basis of characteristic facial dysmorphism, including low anterior hairline, arched eyebrows, synophrys, anteverted nares, maxillary prognathism, long philtrum, thin lips, and 'carp' mouth, in association with prenatal and postnatal growth retardation, mental retardation and, in many cases, upper limb anomalies. However, there is wide clinical variability in this disorder, with milder phenotypes that may be difficult to ascertain on the basis of physical features. Up to now there are five different types of CDLS and 19 genes related to them. **Purpose:** To describe a patient with multiple dysmorphisms and a previous clinical diagnosis of Ohdo Syndrome that underwent Whole Exome Sequencing (WES) with detection of a pathogenic variant in the *HDAC8* gene and further acknowledgment of characteristic features of CDLS type 5. **Methods:** A 6 years old female patient with IUGR, neonatal seizures, cardiac malformation, horse-shoe kidney, developmental delay, hypothyroidism and bilateral deafness had her exome sequenced in the Illumina ™ platform and analysis performed with access through the PhenoDb website. **Results:** A rare missense variant was found in exon 4 of *HDAC8* (c.C356T; p.T119M) with a SIFT score of 0 and a GERP score of 4.89. **Discussion:** This new approach permits the recognition of common clinical features that were initially too subtle or too variable (when seen in patients of different age, sex, and ethnicity) to enable a new syndrome to be identified solely on clinical grounds. This approach may be particularly valuable for the investigation of disorders with high locus heterogeneity, like CDLS.

2571F

A case report of Bardet-Biedl Syndrome. *I.M. Salazar-Dávalos[1], M.A. Aceves-Aceves[1], S.A. Alonso-Barragán[1,2,3], A. Valdez-Haro[2,3], C. Enriquez-Barajas[3], M.F. Romero-Moran[1], D. García-Cruz[1,2], N.O. Dávalos[1,2], I.P. Dávalos[1,2,3]. 1) CUCS, Instituto de Genética, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, México; 3) CIBO, HE, IMSS, México.*

INTRODUCTION Bardet-Biedl Syndrome (BBS) MIM #209900, is an autosomal recessive and genetically heterogeneous ciliophaty. BBS clinically characterized by retinitis pigmentosa, polydactily, central obesity, behavioral dysfunction, mental retardation, hypogonadism and kidney dysfunction. *(Beales et al., 1999).* Eight proteins implicated in the disorder assemble to form BBSome. OBJECTIVE Present a male patient with clinical diagnosis of BBS. CASE REPORT A male aged 24 years old, presented congenital postaxial polydactily on hands and feet and bilateral cryptorchidism. Bilateral orchidopexy was performed at 8 years old. At 10 years, he presented decreased visual acuity. Later, he presented behavioral dysfunction and hypogonadism. At 16 years old, high blood pressure was diagnosed. At present he studies at a special school. Physical examination: height 1.67 m, weight 123kg, BMI 44.1 kg/m2, central obesity, postaxial polydactyly on hands and feet and hypogonadism. Ophthalmological evaluation revealed compatible data with retinitis pigmentosa. Abdominal ultrasound reported hepatic steatosis and multilobulated kidneys. CONCLUSIONS The diagnosis of Bardet-Biedl syndrome is established with the presence of four major clinical criteria or three major criteria and two minor. *(Beales et al 1999).* The major clinical criteria in the patient: retinitis pigmentosa, obesity, polydactyly, learning difficulty, hypogonadism and kidney abnormalities, led us to consider the clinical diagnosis of BBS. An early diagnosis of BBS allows an interdisciplinary approach and a suitable treatment of complications.

2572W

Surgical procedures and anesthetic complications in 27 Brazilian patients with Mucopolysaccharidosis (MPS) types I, II and VI. *C.A. Kim, D.C. Soares, L.M. Albano, R.S. Honjo, D.R. Bertola, J.F. Franco.* Medical Genetics Unit, Hospital das Clínicas da Faculdade de Medicina da USP, Sao Paulo, Brazil.

Background: Mucopolysaccharidosis (MPS) are multisystemic disorders caused by lysosomal enzyme deficiencies. MPS patients are generally considered high risk for anesthesia care due to a difficult airway, cervical spine disease and an increased prevalence of cardiovascular manifestations. **Objective:** To assess the incidence of anesthetic complications in patients with MPS I, II, and VI. **Methods**: We have studied 27 MPS patients (13 MPS I; 8 MPS II; 6 MPS VI) diagnosed by enzymatic and urinary GAGs analysis. **Results**: First symptoms appeared at about 6mo to 8y (MPS IH – mean 7mo; MPS IHS – mean 2y; MPS IS – mean 6y10mo; MPS II – mean 3y6mo; MPS VI – mean 1y). The mean ages at diagnosis were: MPS IH –1y6mo; MPS IHS – 4y8mo; MPS IS – 13y7mo; MPS II and VI – 5y. There were five familial cases. The age of onset of ERT ranged from 1y2mo to 31y9mo. The duration of ERT ranged from 40wk-556wk (mean 259wk). Before ERT, the main clinical complications were high blood pressure (25%), hypoacusis (37%), and hydrocephalus (15%). After ERT these same findings worsened: 37%, 59%, and 22%, respectively. Among 27 patients, 22 (81%) were submitted to surgical procedures (mean of 2.9 surgeries/patient). The most common procedures were: herniorrhaphy (16/27), adenotonsillectomy (13/27) and median nerve decompression (3/27). In contrast, other non-routine procedures, such as cardiac valvuloplasty (1/27), spinal decompression (1/27), ventriculoperitoneal shunt (5/27) and orthopedic surgery for scoliosis correction (1/27) were needed. Five patients presented anesthetic complications (23%) and two of them deceased during the procedures due to orotracheal intubation. **Conclusions**: Because of the high anaesthetic risk, the benefits of a procedure in patients with MPS should always be balanced against the associated risks. Therefore, careful evaluation of anaesthetic risk factors should be made before the procedure, involving evaluation of airways and cardiorespiratory and cervical spine problems. In addition, information on the specific type of MPS, prior history of anaesthesia, presence of cervical instability and range of motion of the temporomandibular joint are important and may be pivotal to prevent complications during anaesthesia.

2573T

Somatic mosaicism for a lethal *TRPV4* mutation results in non-lethal metatropic dysplasia. *M.M. Weinstein[1], T. Kang[2], R.S. Lachman[3], M. Bamshad[4,5,6,7], D.A. Nickerson[4,5], D. Krakow[3,8,9,10], D.H. Cohn[1,3,9,10], University of Washington Center for Mendelian Genomics.* 1) Molecular, Cell, and Developmental Biology, University of California Los Angeles, Los Angeles, California; 2) Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, Los Angeles, California; 3) International Skeletal Dysplasia Registry, University of California Los Angeles, Los Angeles, California 90095, USA; 4) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 5) University of Washington Center for Mendelian Genomics, University of Washington, Seattle, Washington, USA, 98195; 6) Department of Pediatrics, University of Washington, Seattle, WA 98195; 7) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA 98105, USA; 8) Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095; 9) Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095; 10) Orthopaedic Hospital Research Center, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095.

Somatic mosaicism for dominant mutations has previously been observed causing milder, and sometimes clinically undetectable, disease in patients. Dominant mutations in *TRPV4*, which encodes the Transient Receptor Potential Cation Channel Subfamily V Member 4 calcium channel, result in a series of musculoskeletal disorders that include a set of peripheral neuropathies and a broad phenotypic spectrum of skeletal dysplasias. The skeletal phenotypes range from brachyolmia, in which there is scoliosis with mild short stature, through perinatal lethal metatropic dysplasia. We describe a case with phenotypic findings consistent with metatropic dysplasia, but in whom no *TRPV4* mutation was detected by Sanger sequence analysis. Exome sequence analysis identified a known lethal metatropic dysplasia mutation, *TRPV4[L618P]*, which was observed in about 23% of reads covering the locus, a significantly lower frequency than would be expected for a heterozygous change. The affected individual was shown to be a somatic mosaic for the mutation, providing an explanation for the difficulty of observing the change by Sanger sequence analysis and for the milder than expected phenotype observed. The data identify a new genetic mechanism for metatropic dysplasia and illustrate the power of high-throughput sequencing to detect variants present in the mosaic state.

2574F

A double mutation including a novel mutation of *SLC9A9* **gene and mosaic mutation of** *PORCN* **gene in a male patient with multiple congenital anomalies, intellectual disability, and autism spectrum disorder.** *M. Minatogawa[1], C. Abe[1], T. Yokoi , Y. Enomoto , K. Ida , Y. Tsurusaki , N. Harada , T. Naruto[2], K. Kurosawa .* 1) Medical genetics, Kanagawa children medical center, Yokohama, Japan; 2) Biomedical sciences, Tokushima university graduate school, Tokushima, Japan.

   Background: There are some congenital anomalies and/or intellectual disabilities that are not enough to be explained by a single gene mutation. As a result of advances in clinical exome and genome sequencing technology, double or multiple causative genes mutation have recently been reported. We report on a male patient presented with congenital anomalies, intellectual disability, and autism spectrum disorder, who carries a novel mutation of *SLC9A9* responsible gene of the autism spectrum disorder and mosaic mutation of *PORCN* responsible gene of Goltz syndrome. Case report: We describe a 4-year-old boy with clinical features with multiple congenital anomalies including scalp aplasia, dermatrophia, hypoplastic nails, left central polydactyly, glaucoma, left blepharoptosis, and right cryptorchism, dyssomnia, failure to thrive, moderate intellectual disability, and autism spectrum disorder. Clinical whole exome sequencing for the diagnosis of mendelian disorders was performed. The case was found to be a de novo mosaic heterozygous nucleotide substitution for the nonsense mutation in exon 12: c.1077delC (p.Y359X) in *PORCN* gene, and a parental origin heterozygous nucleotide substitution for the nonsense mutation in exon 12: c.1462C>T (p.Q488X) in *SLC9A9* gene. Putting these mutations together, it permitted explanation for all of his clinical features. Conclusion: We often encounter patients with congenital anomalies and/or intellectual disabilities which do not enable us to explain by a single gene mutation. It is important to keep in mind that some of the patients carry multiple causative genes mutation when analyses are done. *COI:;None.*

2575W

PIK3C3 **as responsible gene for new neurocutaneous syndrome with learning disorder, epidermal nevus and cataract.** *A. Matsumoto[1], Y. Inaguma[2], M. Noda[2], D. Usui[4], M. Goto[1], E. F. Jimbo[1], H. Tanabe[2], A. Maeda[4], K. Kikkawa[3], M. Y. Momoi[1], H. Osaka[1], K. Nagata[2], T. Yamagata[1].* 1) Pediatrics, Jichi Medical University, Shimotsuke, Japan; 2) Department of Molecular Neurobiology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan; 3) Department of Pediatrics, Kochi Health Sciences Center, Kochi, Japan; 4) Department of Pediatrics, Hata Kenmin Hospital, Kochi, Japan.

   PIK3C3 encodes Class III phosphoinositide 3-kinase which regulates vesicular trafficking, autophagy and nutrient sensing. Classes of PI3Ks phosphorylate the different ring of phosphatidylinositol and have different functions. We detected the deletion of *PIK3C3* on the aCGH in the siblings with epidermal nevus, cataract and borderline intelligent quotient (IQ). (Patients and methods) ; The patient was a girl of 6 years of age. At six years and 2 months of age, the girl's full IQ, working memory index, verbal comprehension index, perceptual reasoning index, and processing speed index were 80, 69, 99, 87, 76, respectively, according to the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV). WISC-IV's profile is borderline IQ and diagnosed as specific learning disorder. She showed hypotonia, cataract and retinitis pigmentosa. Many epidermal nevus was detected in her body. Her 8-month-old-brother showed the epidermal nevi since seven months of age. His development was not delayed, so far and his ophthalmological evaluation was normal. Their gDNA was extracted from lymphocytes after obtaining informed consent from their parents. Array CGH analysis was performed using Agilent Human genome CGH 180K. The expression level of PIK3C3 among family members was analyzed by western blot. *In utero* electroporation, PIK3C3 was suppressed using RNAi that was induced to the fetal brain on embryonic day 14.5 and the brain was analyzed after birth. (Results) A heterogenous 107Kb deletion at 18q12.3 (nt. 39554147-39661206) was detected in the patient, her brother and mother by array aCGH. Exon5 to 23 of PIK3C3 located in this deleted region. Mutation on PIK3C3 was not detected by direct sequence. In the western blot analysis, PIK3C3 expression levels were 0.61-fold in patient, 0.91-fold in father, 0.38-fold in mother, and 0.41-fold in brother, compared to the control samples. Suppression of PIK3C3 in the embryonal brain resulted the delayed migration of cortical neurons and axon extension to the contralateral hemisphere was delayed. (Conclusion) *PIK3CA*, the class I of PIKs, has been shown to be a causative gene for epidermal nevi. The de novo *PIK3C3* deletion was reported in the patient with ID and agenesis left cerebellum. PIK3C3 knockdown mice have been reported to develop congenital cataract and microphthalmia. Taken together, deletion of *PIK3C3* appears account for epidermal nevi, cataract and also the ID in the patients.

2576T

Phenotypic spectrum in females carrying mutations in *ZC4H2*. *K.J. Buckingham[1], J.X. Chong[1], C.T. Marvin[1], A.G. Shankar[1], A.E. Tattersall[1], R. Cornejo Jr.[1], K.M. Shively[1], J.M. Graham Jr.[2], D.L. Earl[1,3], I.A. Glass[1], D.A. Nickerson[4], M.J. Bamshad[1,3,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA; 4) Department of Genome Sciences, University of Washington, Seattle, WA.

Arthrogryposis is a heterogeneous group of disorders characterized by non-progressive congenital joint contractures of two or more different body areas and is often associated with reduced fetal movements. Arthrogryposis is a clinical feature of more than 500 rare conditions and can be caused by abnormalities of the central or peripheral nervous system, neuromuscular junction, or muscle. The distal arthrogryposes (DA) are a phenotypically and genetically heterogeneous subset of arthrogryposis conditions. Persons with DA typically have a normal neurological exam and no cognitive delay, however these characteristics can be difficult to discern early in life. Analysis of exome sequence data from a large cohort of families putatively diagnosed with DA identified mutations in *ZC4H2* in seven kindreds. Six out of the seven kindreds were simplex female cases, each with a different mutation – four of which were confirmed to be *de novo*. Clinical characteristics of these cases include camptodactyly, contractures of the ankles, hip dysplasia, scoliosis, cleft lip/palate, seizures, and intellectual disability. The seventh case was an affected male. Five (four nonsense, one frameshift) of the six mutations identified in the simplex female cases are predicted to result in loss of function (LOF), while the sixth female case and the only male case both had missense mutations, p.(Gln194Arg) and p.(Cys206Tyr), affecting residues in the highly conserved zinc-finger domain of ZC4H2. None of these mutations have been previously reported. Hemizygous mutations in *ZC4H2* cause Wieacker-Wolff syndrome (WWS), a severe X-linked recessive condition characterized by arthrogryposis, hypotonia, motor and developmental delay, and sometimes early death. Some carrier females are unaffected, while others are reportedly "mildly affected" with features including intellectual disability, camptodactyly, and ankle contractures. The females in our cohort with predicted LOF mutations appear to be comparatively more severely affected than previously described females in WWS families and simplex female cases. Most previously reported WWS families have missense variants suggesting that LOF variants may cause a more severe phenotype in WWS females. However, two females with heterozygous deletions spanning *ZC4H2* and "mild" phenotypes have been described, raising the possibility that nonsense/frameshifts mutations cause WWS via a different mechanism (e.g., dominant negative) than that of full gene deletions.

2577F

Expanding the phenotypic spectrum of a single mutation: The same nonsense mutation in *HOXD13* can alternately cause brachydactyly type E or synpolydactyly in different families. *C.T. Marvin[1], J.X. Chong[1], K.J. Buckingham[1], K.M. Shively[1], T.M. Harrell[1], A.G. Shankar[1], R. Cornejo Jr.[1], A.E. Tattersall[1], D.A. Nickerson[2], M.J. Bamshad[1,2,3], University of Washington Center for Mendelian Genomics.* 1) Dept. of Pediatrics, University of Washington, Seattle, WA, USA; 2) Dept. of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Seattle Children's Hospital, Seattle, WA, USA.

Homeobox (HOX) genes are highly conserved and encode transcription factors that regulate the development of multiple organs and body areas. *HOXD13* regulates the skeletal development of the wrist, fingers, ankle, and toes, and mutations in *HOXD13* cause multiple distinct skeletal abnormalities of the digits. Putative loss of function mutations and expansions of the N-terminal polyalanine repeat cause synpolydactyly, while missense variants can cause synpolydactyly 1 as well as brachydactyly types D and E, syndactyly type V, and brachydactyly-oligodactyly. We studied a single affected male who presented with bilateral complex syndactyly of the third web space with an extra digit on each hand and no mobility of the proximal or distal interphalangeal joints on the left. Review of the family history found a pattern of finger and toe malformations including syndactyly and apparently isolated camptodactyly in the extended paternal family. Exome sequencing of the proband with subsequent Sanger validation and segregation studies revealed a paternally inherited heterozygous nonsense mutation in the homeodomain of *HOXD13* (c.820C>T, p.[R274*]). This variant is predicted to be damaging (CADD 28.8), highly conserved (GERP 4.05), and very rare (not present in ExAC, EVS, or 1000 Genomes). This same nucleotide change has been reported in a family which had isolated brachydactyly type E. Our finding of c.820C>T, p.[R274*] in a family with synpolydactyly suggests that this nonsense mutation has more variable effects on phenotype than previously suspected and that the phenotypic outcome of synpolydactyly versus brachydactyly, or a combination thereof, may be the product of modifier variants in *HOXD13* or other genes, epigenetic factors, or other environmental factors present during development.

2578W

Novel *RPL11* splicing variant associated with an atypical familial presentation of Diamond-Blackfan anemia-7. *C.M. Carlston[1], J.C. Palumbos[2], H. Bagley[3], C. Barbagelata[2], W.L. Wooderchak-Donahue[1], R. Mao[1], J.C. Carey[2].* 1) Pathology Dept, University of Utah/ ARUP, Salt Lake City, UT; 2) Pediatrics Dept, Division of Medical Genetics, University of Utah School of Medicine, Salt Lake City, UT; 3) Intermountain Healthcare, Salt Lake City, UT.

Diamond-Blackfan anemia (DBA) is a collection of clinically and genetically heterogeneous inherited red blood cell aplasia disorders with or without congenital anomalies. Mutations in genes encoding 16 ribosomal proteins, a ribosomal RNA, and the GATA1 transcription factor have all been associated with various forms of DBA. Diamond-Blackfan anemia-7 (DBA7) (OMIM #612562), caused by heterozygous mutations in the *ribosomal protein L11* (*RPL11*) gene, accounts for approximately 5% of DBA. DBA7 is usually characterized by severe anemia, often accompanied by malformations including thumb anomalies, short stature, heart defects, craniofacial anomalies, and urogenital defects. Variable expressivity and incomplete penetrance have been observed within families. A two-year-old boy presented with chronic mild anemia (requiring no treatment), proportionate short stature, bilateral underdevelopment of the thumbs (preaxial hypoplasia, altered thenar eminence, and one apparent triphalangeal thumb), fenestrated atrial septal defect at birth, and hypospadias. Hematological testing in the proband showed slightly decreased hematocrit and hemoglobin, normal fetal hemoglobin, and elevated adenosine deaminase activity. Family history included maternal relatives with thumb defects, but the mother's thumbs were normal. Whole exome sequencing analysis detected a maternally-inherited variant, c.396+3A>G, in the *RPL11* gene that is predicted to affect splicing. A family correlation study of the identified *RPL11* variant found that it segregated with thumb anomalies. RNA studies in this family suggest that the identified variant produces an alternative transcript that may be susceptible to nonsense-mediated decay. This report describes a non-classical familial presentation of DBA7 more associated with thumb anomalies than with severe anemia.

2579T

Functional characterization of a novel mutation in *EPHB4* that underlies familial lymphangiomatosis. *M.E. March, D. Li, R. Pellegrino, L. Tian, F. Mafra, J. Snyder, F. Wang, J. Garifallou, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Lymphangiomatosis is a rare condition characterized by proliferation of lymphatic vessels. The specific presentation of the condition, and its severity, are determined by the location of the hyperproliferating lymphatic tissue. With the exception of the central nervous system, which lacks lymphatic vessels, any part of the body can be affected. The molecular basis of lymphangiomatosis is largely not understood. Whole exome sequencing of a family of subjects with dominantly inherited lymphangiomatosis led to the identification of variant in *EPHB4* that is predicted to affect mRNA splicing. RT-PCR performed with RNA from EBV-transformed lymphoblastoid cell lines from family members confirmed the presence of an aberrantly spliced, yet in-frame, variant in an affected subject. The variant contains a small insertion of 12 base pairs, corresponding to an in-frame insertion of 4 amino acids in the protein sequence. In order to examine the functional consequences of this alteration of the EPHB4 protein, we have generated expression constructs representing the wild type and abnormally spliced variants of *EPHB4*, and expressed them exogenously in cell lines. Wild type and mutant proteins are expressed to similar levels following transfection, and stimulation of cells with pervanadate followed by Western blotting shows that both proteins can be phosphorylated on tyrosine residues. However, under normal culture conditions, despite similar expression levels, the wild type protein is much more heavily tyrosine phosphorylated than the mutant. This indicates that the small insertion in the protein sequence caused by altered mRNA splicing has significant impact on protein functionality. Experiments are currently underway to examine the impact of these biochemical changes on the behavior of cells expressing EPHB4 in adhesion and migration assays. Additionally, we continue to investigate the mechanism through which this mutation acts in a dominant manner.

2580F

Novel homozygous missense mutation in a SH3 binding motif of the STAMBP gene causing microcephaly-capillary malformation syndrome. *I. Hori[1], F. Miya[2,3], Y. Negishi[1], A. Hattori[1], N. Okamoto[4], M. Kato[5], T. Tsunoda[2,3], M. Yamasaki[6], Y. Kanemura[7,8], K. Kosaki[9], S. Saitoh[1].* 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 5) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 6) Department of Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 7) Division of Regenerative Medicine and Department of Neurosurgery, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 8) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 9) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

  Microcephaly-capillary malformation syndrome (MICCAP [MIM 614261]) is characterized by severe progressive microcephaly, intractable epilepsy, profound developmental delay and multiple small capillary malformations on the skin. MICCAP is caused by biallelic mutations in the STAM-binding protein gene (*STAMBP* [MIM 606247]) in chromosome 2p13, which has a key role in cell surface receptor-mediated endocytosis and sorting. McDonell et al reported that patient cell lines showed reduced STAMBP expression, elevated apoptosis and insensitive activation of the RAS-MAPK and PI3K-AKT-mTOR pathways. Here we report the first mutation located in the SH3 binding motif of the *STAMBP* gene. The patient is a 2-year-old Japanese boy with no family history of nonconsanguineous parents. The boy was born by vaginal delivery in 37th gestational weeks without any perinatal abnormalities. At birth he had a weight of 2680 g (-0.8SD) and OFC 33.0 cm (0SD). He showed progressive microcephaly, hypertelorism, ptosis, downturned mouth, and multiple capillary malformations involving the trunk and extremities. At 2 years he was not able to hold up his head yet. When he had infections, his feeding became poor necessitating tube feeding. He developed seizures at 7 months. Seizures were refractory to multiple anti-epileptic medications. Head MRI showed symmetric atrophy of the cerebral cortex. Karyotype and genomic microarray were normal. We performed whole exome sequencing on the proband and his parents, and identified a homozygous mutation (c.707C>T; p.S236F) in the *STAMBP* gene. His parents had the mutation in a heterozygous fashion. To interrogate RAS-MAPK pathway function, we performed western blot analysis of patient derived lymphoblastoid cell lines (LCLs) using antibodies against phosphorylated S6 ribosomal protein (pS6) in starvation condition. Serum starvation of LCLs significantly reduced the expression of the pS6 protein not only in WT LCLs but also in patient derived LCLs, indicating that insensitive activation of PI3K-AKT-mTOR pathway was not observed in patient derived LCLs. STAMBP interacts with the SH3 domain of STAM to transduce downstream signals from the Jaks-STAM complex. Because the substitution of S236F found in our patient is located in the SH3 binding motif, the mutation is likely to be pathogenic. Our western blot result is inconsistent with that in the previous report, suggesting that other mechanisms may underlie the MICCAP phenotype.

2581W

Contractures, rhabdomyolysis, metabolic crisis and encephalopathy caused by bi-allelic mutations in *TANGO2.* *L. Fernandez[1], P. Zornio[1], A. Dries[1], D. Zastrow[1], J. Kohler[1], S. Schelley[2], G. Enns[2], K. Van Haren[3], A. Myers[2], C. Eng[4], M. Walkiewicz[4], Members of the UDN[5], E. Ashley[1,6], P. Fisher[1,3], J. Bernstein[1,2], M. Wheeler[1,6].* 1) Undiagnosed Diseases Network, Stanford School of Medicine, Palo Alto, CA, USA; 2) Division of Medical Genetics, Stanford School of Medicine, Palo Alto, CA, USA; 3) Division of Neurology, Stanford School of Medicine, Palo Alto, CA, USA; 4) Baylor Miraca Genetics Laboratories, Houston, TX, USA; 5) Undiagnosed Diseases Network, National Institutes of Health, Bethesda, MD, USA; 6) Division of Cardiovascular Medicine, Stanford School of Medicine, Palo Alto, CA, USA.

  Transport and Golgi organization (TANGO) proteins play an important role in cargo loading of newly synthesized secretory proteins in the endoplasmic reticulum (ER). TANGO2 is localized to the Golgi and cytoplasm. Bi-allelic mutations in *TANGO2* have been recently reported in individuals with rhabdomyolysis, metabolic crisis, encephalopathy, and cardiac arrhythmia. We report on a 4 year-old, Hispanic male with global developmental delay, contractures preventing ambulation, and an episode of acute decompensation with encephalopathy, hypoglycemia, lactic acidosis, and rhabdomyolysis. Comprehensive metabolic workup including fatty acid oxidation probe on fibroblasts, mitochondrial genome, and intracellular acylcarnitine profile did not reveal a diagnosis. Trio whole exome sequencing performed through the Undiagnosed Diseases Network identified pathogenic variants in the *TANGO2* gene: a heterozygous point mutation c.460G>A (p.G154R) and heterozygous exon 3-9 deletion. Parental sequencing confirmed compound heterozygosity. Compared to the clinical features of other subjects reported with bi-allelic *TANGO2* mutations, this individual does not present with seizures, cardiac arrhythmias, hypothyroidism, optic atrophy, microcephaly or macrocephaly, or hearing loss. Interestingly, our patient has macrocytic anemia and contractures in all limbs, findings not reported in previous cases and not explained by exome sequence reanalysis. Functional studies performed on fibroblasts and lymphoblastoid cell lines of individuals with biallelic *TANGO2* mutations have shown evidence of involvement of perturbed vesicular Golgi/ER transport and increased ER stress; other studies favor a functional defect in mitochondrial beta-oxidation of fatty acids as contributing to disease. Ongoing work to characterize iPS-cell derived lineages from our patient may elucidate the mechanisms by which *TANGO2* mutations impact cellular function in multiple tissues.

2582T

Balanced X-autosome translocation suggests association between AMMECR1 and growth, bone and heart alterations. *M. Moyses-Oliveira[1,2], R. Fish[3], F. Petit[4], G. Giannuzzi[2], V. Meloni[1], M. Soares[5], L. Kulikowski[6], A. Di Battista[1], T. Liehr[7], N. Kozyakova[7], G. Carvalheira[1], J. Andrieux[8], M. Neerman-Arbez[3], M. Melaragno[1], A. Reymond[2].* 1) Department of Morphology and Genetics, Universidade Federal de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; 3) Department of Genetic Medicine and Development, 1211 University of Geneva Medical School, Geneva, Switzerland; 4) Clinique de Génétique, CHU Lille - Hôpital Jeanne de Flandre, 59037 Lille, France; 5) Psychobiology Department, Universidade Federal de São Paulo, 04023-062 São Paulo, Brazil; 6) Department of Pathology, Laboratório de Citogenômica, LIM 03, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, 05403-000 São Paulo, SP, Brazil; 7) Institute of Human Genetics, Jena University Hospital, Friedrich Schiller University, D-07743 Jena, Germany; 8) Institut de Génétique Médicale, CHU Lille - Hôpital Jeanne de Flandre, 59037 Lille, France.

Female balanced X-autosome translocations are usually associated with absence of functional copies of the gene mapping at the breakpoint through disruption of the derivative-chromosome copy and inactivation of the normal X-chromosome. We report on a nine year-old girl with karyotype 46,X,t(X;9)(q23;q12)dn, disproportionate short stature (<1st centile), septal atrial defect, scoliosis, bone dysplasia, and normal cognition. Array-CGH and breakpoint sequencing confirmed the full complement of genetic material, whereas EdU incorporation method and HUMARA assay showed preferential inactivation of the normal X-chromosome. The autosomal breakpoint affects a heterochromatic region, while the X-chromosomal breakpoint was mapped between the *AMMECR1* and *RGAG1* genes. Whereas expression of the latter was unmodified, RT-qPCR showed absence of *AMMECR1* expression in blood and lymphoblastoid cells of the proband. The encoded protein contains evolutionary conserved nucleic acid-binding RAGNYA domains and localizes to the nucleus. *AMMECR1* is co-expressed with genes implicated in cell cycle and translation regulation, five of which were previously associated with growth and bone alteration syndromes. Concordantly, we also identified an individual with bone dysplasia and ventricular septal defect and duplication of the *AMMECR1*. Our knockdown of the zebrafish orthologous gene resulted in animals with phenotypes such as shorter tails, thinner bodies, kinked tail-ends, poorly defined somites, pericardial edema, tachycardia and hydrocephaly, features reminiscent of the proband's phenotype. Additionally, a phenotype similar to the *ammecr1* knockdown zebrafish was previously observed in zebrafish ribosomal protein knockdown and *top2a* mutant. The knockdown phenotype was partially rescued by co-injection of mRNA from the human ortholog. Our results suggest that *AMMECR1* is potentially involved in cell cycle control and linked to a new syndrome(s) with growth, bone and heart alterations.

2583F

Recurrent *ZIC1* frameshift mutation in siblings causes syndromic cerebral and cerebellar dysgenesis, tethered chord and scoliosis, without craniosynostosis. *G.M. Mancini[1], R. Schot[1], M. Wilke[1], C. Lincke[2], J. Hoogeboom[1], R. de Coo[3], F.W. Verheijen[1], M. van Slegtenhorst[1].* 1) Clinical Gen, Erasmus Med Ctr, Rotterdam, ZH, Netherlands; 2) Pediatrics, Erasmus Med Ctr, Rotterdam, ZH. Netherlands; 3) Child Neurology, Erasmus Med Ctr, Rotterdam, ZH, Netherlands.

The ZIC family of zinc finger proteins are transcriptional regulators , which govern essential steps of brain development. Deletion of contiguous genes *ZIC1* and *ZIC4* on chromosome 3q24 has been associated with Dandy-Walker malformation (DWM) and cerebellar vermis dysplasia with reduced penetrance in patients and double K.O. mice (Grinberg et al, Nat Genet 2004, 36(10):1053). Isolated *ZIC4* mutations have not been reported, but inherited missense and de novo truncating gain of function *ZIC1* mutations have been associated with coronal craniosynostosis and intellectual disability (Twigg et al, AJHG 2015, 97:378) in 5 families. During screening of patients with neurodevelopmental disorders and brain malformations , we found by exome sequencing a *de novo* heterozygous frameshift mutation in *ZIC1* (c.1214_1215ins10, p.Pro406fs) in two siblings, a boy and a girl of Caucasian ancestry, suggesting germ line mosaicism They share an identical phenotype of severe developmental delay, no speech and no independent deambulation, congenital microcephaly (- 3.5 SD) and strabismus. Similar features were noted consisting of deep set eyes and prominent lips, sharp brachycephalic skull without evidence of premature suture closure at X-rays and scoliosis. MRI of brain and spinal cord showed in both children sacral lipoma, spina bifida occulta, agenesis of the corpus callosum, hypoplastic pons and brain stem, wide IVth ventricle communicating with the cisterna magna, hypoplastic vermis and cerebellar hemispheres, without upward rotation of the vermis, and patchy white matter lesions diffusely present in the white matter. In the boy the vermis hypoplasia is associated with incomplete separation of cerebellar hemispheres, compatible with a partial rhombencephalosynapsis. Retrospective comparative review of the MRI from the patient with the p.Gln389* mutation of Twigg et al (AJHG 2015), confirmed major differences with the present patients. Testing of *ZIC1* transcripts in skin fibroblasts from our patients is ongoing. It has been suggested that modifying factors contribute to the phenotypic expression of the *ZIC1* mutations. We conclude that, besides the contiguous *ZIC1* and *ZIC4* deletion associated with DWM, the phenotypic spectrum of *ZIC1* mutations, be it loss- or gain-of-function, is much wider than appreciated until now and that it includes both DWM, distinctive CNS dysgenesis patterns and syndromic craniosynostosis.

2584W

The pathogenic germline variant p.Arg882His in the DNA methyltrans-ferase 3A (*DNMT3A*) gene is associated with Tatton-Brown-Rahman syndrome. *J. Heeley[1], N. Shur[2], K. Miller[2], L. Manwaring[4], K. Levine[3], G. Douglas[3], M. Shinawi[4].* 1) Mercy Kids Genetics, Mercy Hospital St. Louis, St Louis, MO; 2) Department of Pediatrics, Section of Genetics & Metabolism, Albany Medical Center, Albany, NY; 3) GeneDx, Gaithersburg, MD; 4) Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, MO.

   Background: Somatic pathogenic variants in *DNMT3A* are found in 30% of acute myeloid leukemia cases and are typically present in the founding clones. Most pathogenic variants that cause AML do so via haploinsufficiency, but more than half of AML cases with *DNMT3A* variants are due to substitutions at amino acid R882. Variants at R882 have been shown to alter the catalytic domain resulting in a dominant negative effect. Recently, constitutional pathogenic variants in *DNMT3A* have been determined to cause an overgrowth syndrome with intellectual disability (Tatton-Brown-Rahman syndrome; TBRS). In this report we describe two patients with TBRS associated with novel *DNMT3A* germline variants. One of the patients has the R882H variant which is recurrently observed as a somatic variant in AML. **Methods:** Whole exome sequencing (WES) was performed at a diagnostic laboratory and revealed novel pathogenic variants in the *DNMT3A* gene in two patients. Their clinical findings were compared with previously reported cases. **Results:** Patient 1 is a 9-year-old male who presented with severe hypotonia, macrosomia, macrocephaly, dysmorphic features, and global developmental delay (formal IQ testing was ~50). WES revealed heterozygosity for a constitutional p.Arg882His pathogenic variant (c.2645G>A) from peripheral blood. Similar results were obtained from buccal swabs. Patient 2 is a 6-year-old female who presented with hypotonia, complex partial seizures, dysmorphic features, macrosomia and macrocephaly. Starting at age 2, patient exhibited regression with loss of motor and language milestones and was diagnosed with autism. Using WES, a heterozygous nonsense variant, W306X (c.918G>A), was identified. **Conclusion:** Both patients with pathogenic germline variants in *DNMT3A* had similar phenotypic features, which were consistent with previous reports of TBRS due to other *DNMT3A* variants. Although it was initially hypothesized that the p.Arg882His variant is limited to somatic tissues and cancer development, our data indicate that this variant can also cause TBRS. To our knowledge, this is the first reported patient with a constitutional variant at R882. Our patients have not shown evidence of leukemia or other malignancy to date. However, as there is a presumed risk for developing malignancies in patients with constitutional *DNMT3A* variants, optimal screening and surveillance protocols should be developed to monitor those patients.

2585T

Bi-allelic mutations in *PKD1L1* are associated with laterality defects in humans. *F. Vetrini[1,2], L.C.A. D'Alessandro[3,5], Z.C. Akdemir[1], A. Braxton[2], M. Azamian[1], M.K. Eldomery[1], K. Miller[5], C. Kois[6], V. Sack[6], N. Shur[6], A. Rijhsinghani[6], J. Chandarana[2], Y. Ding[4], S. Jhangiani[4], C.M. Eng[1,2], T. Harel[1], J.A. Rosenfeld[1], N. Hanchard[1,5], J.W. Belmont[1], J.R.L. Lupski[1,4,5], Y. Yang[1,2].* 1) Department of Molecular and Human Genetics, Houston, TX, USA; 2) Baylor Miraca Genetics Laboratories, Houston, TX, USA; 3) Division of Cardiology, Texas Children's Hospital, Houston, TX, USA; 4) Human Genome Center, Baylor College of Medicine, Houston, TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Albany Medical Center, Albany, NY, USA.

   Disruption of the establishment of left-right (L-R) asymmetry leads to *situs* anomalies ranging from *situs inversus totalis* (SIT) to *situs ambiguus* (heterotaxy). The genetic causes of laterality defects in humans are highly heterogeneous. Using whole-exome sequencing (WES), we identified homozygous mutations in *PKD1L1* from three affected individuals in two unrelated families. *PKD1L1* encodes a polycystin-1-like protein and its loss of function is known to cause laterality defects in mouse and medaka fish models. Family 1 had one fetus and one deceased child with heterotaxy and complex congenital heart malformations. WES identified a homozygous splicing mutation c.6473+2_6473+3delTG, which disrupts the invariant splice donor site in intron 42, in both affected individuals. In the second family, a homozygous c.5072G>C (p.Cys1691Ser) missense mutation was detected in an individual with SIT and congenital heart disease. *In silico* analysis shows that the p.Cys1691Ser substitution affects a highly conserved cysteine residue and is predicted by molecular modeling to disrupt a disulfide bridge essential for the proper folding of the G protein–coupled receptor proteolytic site (GPS) motif. Damaging effects associated with the substitutions of this conserved cysteine residue in the GPS motif have also been reported in other genes, namely *GPR56, BAI3* and *PKD1* in human and *lat-1* in *C. elegans,* further supporting the likely pathogenicity of p.Cys1691Ser in *PKD1L1*. The identification of bi-allelic *PKD1L1* mutations recapitulates the previous findings regarding phenotypic consequences of the orthologous genes in mice and medaka fish and further expands our understanding of genetic contributions to laterality defects in humans   .

2586F

De novo missense variants in *PPP1CB* are associated with features overlapping with Noonan syndrome. *H.M. McLaughlin[1], L. Ma[2], Y. Bayram[3], M.T. Cho[1], A.M. Lewis[3], J.J. White[3], S.N. Jhangiani[3,4], J.R. Lupski[3,4,5,6], K. Mayberry[7], W. Mu[8], V. Weinstein[9], A. Krokosky[10], Y. Shao[3], R.E. Schnur[1], K. Retterer[1], S. Bale[1], I. Wentzensen[1], J. Bodurtha[8], C. Bupp[7], D.A. Scott[3,11], C.E. Turner[10], K. Lindstrom[12], E.Y. Gulec[13], S.R. Lalani[3], W.K. Chung[1,2], J. Juusola[1]. 1) GeneDx, Gaithersburg, MD; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 5) Texas Children's Hospital, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 7) Spectrum Health, Grand Rapids, MI; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 9) Division of Genetics and Metabolism, Children's National Medical Center, Washington D.C; 10) Walter Reed National Military Medical Center, Bethesda, MD; 11) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 12) Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, AZ; 13) Medical Genetics Section, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey.*

Protein phosphatase-1 (PP1) is a serine/threonine-specific protein phosphatase involved in the dephosphorylation of a variety of proteins. PP1 has been shown to interact with *SHOC2*, a gene that is mutated in a Noonan syndrome disorder. The PP1 catalytic subunit beta (*PPP1CB*) gene encodes a subunit of PP1 that is important for the regulation of many cellular processes including adipogenesis, cytoskeletal organization, regulation of synaptic plasticity pathways, and muscle contraction and its expression has been inversely correlated with arterial stiffness. A de novo frameshift mutation in *PPP1CB* was previously reported in a patient with severe intellectual disability, short stature, dysmorphic features, and lymphangioma. Using whole exome sequencing, we identified five *de novo PPP1CB* missense variants (p.Pro49Arg, p.Glu183Ala, p.Glu183Val, p.Asp252Tyr p.Glu274Lys) in seven unrelated individuals who share many clinical features with Noonan syndrome. Overlapping clinical features in these individuals include facial dysmorphism, macrocephaly, developmental delay, intellectual disabilities, congenital heart defects, skeletal and connective tissue abnormalities, and short stature. The dysmorphic features present in these individuals resemble those associated with Noonan syndrome and include prominent forehead, short, broad nose, hypertelorism with short palpebral fissures, and webbed neck. Congenital heart defects were present in the majority (5/7) of the individuals in our series and were comprised of aortic anomalies including dilatation and coarctation, hypoplastic arch, and peripheral pulmonic stenosis. Some individuals also exhibited connective tissue anomalies including joint hypermobility, doughy skin, bifid uvula, high arched palate, and pectus excavatum. All five affected amino acid residues are highly conserved among the PP1 subunit family, and all are predicted to disrupt PP1 subunit binding and impair dephosphorylation, likely leading to perturbation of downstream cellular processes. Our data suggest that *de novo* missense variants in *PPP1CB* are associated with a novel syndromic disorder which may be considered within the differential diagnosis for individuals with suspected Noonan syndrome and related disorders.

2587W

CHIME syndrome with *PIGL* mutation: Description of a Brazilian case with mild elevation of alkaline phosphatase. *R.S. Honjo[1], J.R.M. Ceroni[1], M.R. Passos-Bueno[2], G.L. Yamamoto[1,2], C.A. Kim[1], D.R. Bertola[1,2]. 1) Genetics Unit, Instituto da Criança, São Paulo, SP, Brazil; 2) Instituto de Biociências da Universidade de São Paulo, São Paulo, SP, Brazil.*

CHIME syndrome (MIM 280000) is an extremely rare autosomal recessive disorder characterized by **C**oloboma, congenital **H**eart defects, migratory **I**chthyosiform dermatosis, **M**ental retardation, and **E**ar anomalies/**E**pilepsy. It is a glycosylation disorder caused by mutations in *PIGL*, which encodes for an enzyme responsible for the de-N-acetylation of N-acetylglucosaminyl-phosphatidylinositol in the glycosylphosphatidylinositol (GPI) biosynthesis. We report on a 4-year-old male patient, the first and only child of non-consanguineous parents, with neonatal respiratory complications. He presented with neuromotor delay, including severe speech delay; seizures; macrosomia; dysmorphic facial features; bilateral retinal coloboma; dry/hyperpigmented cutaneous lesions in thorax and abdomen; large hands, with mild brachydactyly, finger pads and palmar hyperkeratosis. Complementary exams showed bilateral pyelectasis; ALP of 456U/L (normal: 150-380U/L). Lymphocyte and fibroblast G-banded karyotype, as well as SNP-array were normal. Whole-exome sequencing revealed the presence of compound heterozygous mutations in *PIGL*: a missense recurrent mutation and an intragenic deletion. The latter was confirmed by bridging PCR. The patient here presented showed typical clinical features of CHIME syndrome including mild elevation of ALP. In the six cases so far described, five of them presented a missense mutation in one allele, associated with a loss-of-function mutation in the other one, as our patient. In the few individuals in which ALP serum levels were measured, they showed mild elevation. Only one atypical case, with two frameshift different mutations, was reported recently showing high serum elevation of ALP. This finding suggests that higher compromise of the protein function may explain the difference in ALP serum levels. CHIME syndrome is one of the inherited disorders of the formation and function of GPI and part of the growing group of congenital disorders of glycosylation (CDG). Several other *PIG* genes are responsible for disorders that share in common intellectual disability, epilepsy, coarse facial features and multiple organ anomalies. This group of disorders is still poorly recognized, requiring description of further typical and atypical cases, in order to give better knowledge of the whole phenotypic spectrum of the *PIG* gene family disorders. (FAPESP/CNPQ).

2588T

Phenotypic variability in rare disease: Using *KAT6A* syndrome to explore rare disease variation in the context of genetic background. *V. Arboleda[1], S. Nelson[1,2].* 1) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA.

Understanding natural history of rare disorders can provide prognostic insight for families and provide a method by which researchers can identify patients who are eligible for clinical trials. There are few well-done natural history studies for rare disorders and with the advent of clinical Exome sequencing there has been a huge increase in the number of monogenic disorders with an identified gene, with few avenues for providing systematic descriptions of patient cohorts. We recently reported the first description of *de novo* nonsense mutations in *KAT6A* (a.k.a. MOZ, MYST3) that were identified by clinical exome sequencing in 4 independent families. *De novo* truncating mutations have now been identified in more than 40 additional patients around the world, 16 of which have been reported in the literature. Through the KAT6A Foundation, we have identified 2 of the oldest individuals in the KAT6A cohort, who are 31 and 28 years old. Using in-depth interviews with parents of individuals' diagnosed based on mutations in *KAT6A*, we are able to identify features of the syndrome that are often overlooked and present only in a subset of individuals. Our early assessments indicate that defects in neurodevelopmental pathways, gastrointestinal motility, and cardiac development, are the most commonly observed. However, rare patients also express severe manifestations of autoimmunity, and it is unclear whether the autoimmunity is related to *KAT6A* mutation directly or secondary other factors. Defining a syndrome based on the totality of phenotypes expressed by individuals with *KAT6A* mutations reveals the degree of phenotypic variance, even among individuals who have the exact same mutation. We hypothesize that each individual's genetic background modulates expression of KAT6A syndrome but can also independently increase susceptibility to other diseases that are unrelated to the KAT6A diagnosis. To address this, we have begun a systematic effort to collect functional genomic data from KAT6A families and tissue samples. This, coupled with the growing knowledge of the genetic heritability of common disease can be used to disentangle phenotypes caused primarily by the *KAT6A* mutation from those inherited in the patient's collective genetic background of the patient.

2589F

Recurrent *de novo* missense mtuations in *PPP1CB* cause a novel rasopathy closely resembling Noonan syndrome with loose anagen hair. *K.W. Gripp[1], K.A. Aldinger[2], J.T. Bennett[2], L. Baker[1], J. Tusi[1], N. Powell-Hamilton[1], D. Stabley[3], K. Sol-Church[3], A.E. Timms[4], W.B. Dobyns[2], S.A. De Man[5], A. Larson[6], L. Spruijt[7], D.A. Koolen[7], S.J.C. Stevens[8], A.P.A. Stegmann[8], W.M. Nillesen[7], V. Noordhoek Hegt[9], M. Bollen[10], M. Nellist[11], F. Verheijen[11], P.J. Willems[12], B.H. Chung[13], T. Rinne[7], M.W. Wessels[11].* 1) Division of Medical Genetics, A. I. du Pont Hospital for Children, Wilmington, DE; 2) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 3) Center for Applied Clinical Genomics, A. I. du Pont Hospital for Children, Wilmington, DE; 4) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 5) Department of Pediatrics, Amphia Hospital, Breda, Netherlands; 6) Department of Pediatics, University of Colorado Anschutz Medical Center, Children's Hospital Colorado, Aurora, CO; 7) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 8) Department of Clinical Genetics, Maastricht University Medical Center, The Netherlands; 9) Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands; 10) Department of Cellular and Molecular Medicine, Laboratory of Biosignaling and Therapeutics, KU Leuven, Belgium; 11) Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 12) GENDIA, Genetic DIAgnostic Network, Antwerp, Belgium; 13) Department of Pediatrics and Adolescent Medicine, Li Queen Mary Hospital, University of Hong Kong, Hong Kong, China.

Background Germline mutations resulting in dysregulation of the RAS/MAPK pathway cause syndromic conditions collectively referred to as rasopathies. Rasopathies share phenotypic manifestations including short stature with macrocephaly, developmental delay, cardiac manifestations, skin hyperpigmentation and a malignancy predisposition. Noonan syndrome is a rasopathy caused by mutations in multiple genes encoding components of the RAS/MAPK pathway. Despite its variable phenotype, limited genotype-phenotype correlations exist. Noonan syndrome with loose anagen hair (NS-LAH) is characterized by distinctive hair anomalies, developmental differences and structural brain abnormalities and is caused by a single recurrent missense *SHOC2* mutation. SHOC2 forms a complex with protein phosphatase 1 (PP1). Protein phosphatases counterbalance kinases and control activation of signaling proteins, such as the mitogen activated protein kinases of the RAS/MAPK pathway. **Methods and Results** Using exome analysis, we identified nine unrelated patients with (de novo) missense mutations in PP1 catalytic subunit beta (*PPP1CB*), sharing a recognizable phenotype. These *PPP1CB* variants met the standardized ACMG criteria for pathogenicity. Four individuals had the recurrent *PPP1CB* c.146G>C, p.Pro49Arg mutation; three had *PPP1CB* c.548A>C, p.Glu183Ala; and one each had c.166G>C, p.Ala56Pro; or c.820G>A, p.Glu274Lys. Findings typical for rasopathies were present in all individuals: 5/9 had short stature; 6/9 had relative or absolute macrocephaly; 7/9 had feeding difficulties with long term feeding tube use in 3/9; 9/9 had delayed development. Facial features included hypertelorism in 6/9 and low-set or posteriorly angulated ears in 8/9. Eye abnormalities were seen in 7/9, hearing loss in 4/9. Notably slow growing, light colored or sparse hair suggestive of loose anagen hair was present in 6/9. Cardiac manifestations occurred in 5/9 (pulmonic or mitral valve anomaly, VSD, ASD or mild HCM). Short or webbed neck was present in 4/9 and pectus in 3/9. Brain imaging showed ventriculomegaly in 7/9 and cerebellar tonsillar ectopia or Chiari 1 malformation in 3/8. Hair microscopy showed dystrophic features like greenstick fractures and squared-off anagen bulbs. Functional studies are being performed. **Conclusions** Based on the phenotypic findings and PPP1CB's effect on the RAS/MAPK pathway through RAF dephosphorylation, this novel condition can be considered a rasopathy, most similar to NS-LAH.

2590W

De novo mutation in *RBM10* can cause non-lethal TRAP syndrome: Further delineation of an expanding phenotype. *R. Hamid, L. Fairbrother.* Dept Pediatrics, Div Med Gen, Vanderbilt Univ Sch Med, Nashville, TN.

Mutations within *RBM10* (OMIM*300080) cause TARP syndrome (OMIM#311900). TARP is an X-linked recessive disorder associated with **T**alipes equinovarus, **A**trial septal defect (ASD), Pierre **R**obin sequence (PRS) and **P**ersistence of the left superior vena cava (SVC) as well as other congenital anomalies and significant developmental delays. TARP was felt to be universally lethal in the neonatal period until a recent publication of a young boy with long term survival. (Gripp et al) We present a patient with a mild case of TARP syndrome. At the time of evaluation our patient was a 6y old with PRS, sensorineural hearing loss, bicuspid aortic valve, expressive speech delays (felt to speak at a 3y old level), and intellectual disability. Through WES, we identified a pathogenic mutation in *RBM10*, c.170_171delAT (NM_005676), p.Y57X confirming a diagnosis of TARP syndrome. In contrast to other individuals reported, our patient can walk, speak, and progress in the educational system. He did not have some of the cardinal features of TARP including the ASD, Talipes equinovarus, and Persistence of the Left SVC but did have overlap with previously reported individuals (PRS, atypical ears, and SNHL). In addition, compared to the previously reported life expectancy of children with TARP our patient demonstrates long-term survival is possible. Increased ordering of Whole Exome Sequencing (WES) has led to expanded knowledge of previously described Mendelian Disorders. In the past, only patients who had the cardinal features of a disease were typically tested for a particular disease but with advancements in WES, we are now discovering a spectrum of clinical manifestations for previously well described conditions. In this case WES yielded an unexpected result in our patient. Based on our patient's phenotype and that of those previously reported with TARP consideration of the RBM10 was not within our differential diagnosis. Importantly this case highlights the need to think outside of the box when considering differentials and the utility of WES in cases that do not fit the known phenotype of a described Mendelian condition.

2591T

A case of familial transmission of the newly described *DNMT3A*-Overgrowth Syndrome. *G. Lemire, J. Gauthier, J.-F. Soucy, M.-A. Delrue.* Département de pédiatrie, Service de génétique médicale, Centre Hospitalier Universitaire Ste-Justine, Université de Montréal, Montréal, Canada.

DNMT3A-Overgrowth Syndrome (also known as Tatton-Brown Rahman Syndrome) (MIM 615879) has recently been described in 13 individuals with *de novo* heterozygous mutations in *DNMT3A* gene. This autosomal dominant condition is characterized by overgrowth, dysmorphic facial features and moderate intellectual disability. Missense and truncating point mutations, a small in-frame deletion, as well as microdeletion 2p23 have been reported. Moreover, *DNMT3A* is commonly somatically mutated in acute myeloid leukemia. We herein report a family with an affected father and son. The proband is a 12 year-old boy with tall stature, macrocephaly, facial dysmorphism, and intellectual disability with severe language disorder. He also has mild scoliosis. His father is a 49 year-old man with tall stature, macrocephaly, learning difficulties, and minor facial dysmorphism. He had a right occipital osteoma removed at 20 years of age. A heterozygous splice site mutation NM_022552.4 (*DNMT3A*): c.2323-2A>T was found in both individuals by whole exome sequencing analysis. This mutation has not been previously reported and is believed to be pathogenic. Indeed, this substitution involves a highly conserved canonical splice site and is predicted to cause exon skipping. *DNMT3A* has been reported to regulate bone metabolism via osteoclast activity. We therefore suspect that it was involved in the pathogenesis of our patient's bone tumor. This is the first report of a familial transmission of *DNMT3A*-Overgrowth Syndrome, supporting the autosomal dominant inheritance. The proband's phenotype is more severe than his father's, which illustrates variable expressivity in the syndrome.

2592F

Semidominant inheritance in Frank-ter Haar syndrome. *F. Rossignol, P.M. Campeau.* Medical Genetics, CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada.

 Introduction: Frank-ter Haar syndrome (FTHS, MIM 249420) is an autosomal recessive disease characterized by craniofacial, cardiac and skeletal anomalies, including brachycephaly, a prominent forehead, wide fontanels, macrocornea with or without glaucoma, bowing of the long bones and finger deformities. It is caused by homozygous mutations in the *SH3PXD2B* gene, encoding an adaptor protein involved in podosome formation and extracellular matrix degradation. Here, we report a patient with FTHS associated with reduced penetrance in the paternal side of the family.   **Case description:** A 4 month-old child presented with a large anterior fontanel, a prominent forehead, a wide nasal root, mandibular diastasis, supernumerary teeth, a high arched palate, sphenoidal anomalies, a bell-shaped thorax, and hypoplasia of the corpus callosum. Two mutations in *SH3PXD2B* (NM_001017995.2) were found, respectively c.967G>A (p.G323R, rs111492578) and c.2279G>A (p.R760H, not in dbSNP). Both mutations have a prevalence <0.01% in ExAC, and are predicted to be damaging in SIFT and possibly damaging in PolyPhen2.   On family history, the father presented congenital glaucoma without other signs of FTHS. His brother has a patent posterior fontanel in his adult years. Their mother, who was adopted, presented both glaucoma and a patent posterior fontanel. The patient's maternal family history is unremarkable. Segregation analysis in the family is ongoing.   **Discussion:** This is the first case of symptomatic carriers for Frank ter-Haar syndrome. It suggests that further investigations and surveillance in heterozygous carriers of mutations in *SH3PXD2B* might be warranted.

2593W

Optimized exome sequencing to characterize complex syndromic retinal dystrophy phenotypes. *I. Sanchez-Navarro[1,2,3], L. Tian[3], C. Villaverde[1,2], L. Rodrigues Jacy da Silva[1,4], R. Pellegrino[3], F. Mafra[3], O. Zurita[1,2], L. Vazquez[3], Y. Li[3], X. Chang[3], A. Almoguera[3], D. Li[3], J. Snyder[3], M. Corton[1,2], C. Ayuso[1,2], H. Hakonarson[3].* 1) Department of Genetics, Health Research Institute–Jimenez Diaz Foundation University Hospital (IIS-FJD-UAM) Madrid, Spain; 2) Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain; 3) Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia; 4) Universidade de Mogi das Cruzes, São Paulo, Brazil.

 Statement of purpose: To deepen our understating of syndromic retinal dystrophies (SRD), using Clinical Exome Sequencing (CEX) and Whole-Exome Sequencing (WES). **Methods:** 32 SRD cases previously studied in FJD (Madrid) and 7 cases from CAG-CHOP (Philadelphia) biorepositories were selected for the study. These patients presented with systemic phenotypes with eye affectation (mainly retinal dystrophy) and symptoms such as neuronal alteration, renal or cardiac disease, and obesity, among other features. Some of them were clear ciliary syndromes (Bardet-Bield Syndrome, Joubert Syndrome, Alström Syndrome, Jeune Syndrome), other non-ciliary diseases (Stickler) with other cases presenting with less clearly defined systemic diseases.   Commercial clinical exomes (TruSightOne) were used to study the 32 cases from FJD. Libraries were prepared using TruSigthOne kits and sequenced on a NextSeq500 (v2 chemistry) at FJD Genetics Department in Madrid. FastQ files were used to generate VCF using CAG-ARC-CHOP pipeline. Negative cases from this Clinical Exome Study and 7 other SRD not previously characterized cases from The Children's Hospital of Philadelphia (CHOP) were WES-sequenced (Agilent's SureSelect v5 sequenced on HiSeq with v4 chemistry) at the CAG at CHOP. Sequencing data was analyzed in the search of SNVs and CNVs. **Summary of results**: In the first pass analyses, we were able to characterize 29% of the cohort by means of a gene panel approach. At the moment, with the CEX-based approach we have been able to find the cause of the disease in 4 out of 19 cases studied. These findings included the genes *COL2A1*, *IFT140*, *PEX6* and *CRB1*. The previously uncharacterized cases are currently being analyzed at CAG-CHOP. Based on CAG-CHOP previous experience we expect to characterize about 40% of the cases under study.

2594T

Genetic risk factors identified for severe renal disease in Bardet-Biedl syndrome in the largest reported patient cohort. *E. Forsythe[1], K. Sparks[2], S. Best[1], S. Borrows[3], B. Hoskins[2], A. Sabir[4], T.G. Barrett[5], D. Williams[6], S. Mohammed[7], D. Goldsmith[8], D.V. Milford[4], D. Bockenhauer[9], L. Foggensteiner[3], P.L. Beales[1,2].* 1) Genetics and Genomic Medicine Programme, UCL Institute of Child Health, London, United Kingdom; 2) Bardet-Biedl syndrome National Clinical Service, Great Ormond Street Hospital, London, United Kingdom; 3) Nephrology Department, Queen Elizabeth Hospital, Birmingham, United Kingdom; 4) Nephrology Department, Birmingham Children's hospital, Birmingham, United Kingdom; 5) Endocrinology department, Birmingham Children's hospital, Birmingham, United Kingdom; 6) Clinical Genetics Department, Queen Elizabeth Hospital, Birmingham, United Kingdom; 7) Clinical Genetics Department, Guy's Hospital, London, United Kingdom; 8) Nephrology Department, Guy's Hospital, London, United Kingdom; 9) Nephrology Department, Great Ormond Street Hospital, London, United Kingdom.

Introduction: The high risk of renal disease is a cause of great anxiety for patients with Bardet- Biedl syndrome due to the devastating effect on quality of life, morbidity and mortality. Other hallmark features associated with this pleiotropic autosomal recessive ciliopathy include rod-cone dystrophy, obesity, learning difficulties, polydactyly and hypogenitalism. Twenty one disease causing genes (*BBS1-21*) have been identified, and code for proteins that localise to the basal body and cilium. Molecular confirmation of Bardet-Biedl syndrome can be achieved in 80% of patients. The primary renal phenotype is highly variable and no renal genotype-phenotype correlation has previously been identified. Here we present evidence from the largest reported study to date correlating severe renal disease with genotype and mutation type. **Method**: Data from 350 patients attending the United Kingdom adult and paediatric Bardet-Biedl syndrome clinics were collected over a four year period (2010-14). Genotype, urinalysis, sonographic, biochemical and clinical data were analysed. **Results**: One hundred and ninety four adults and 156 paediatric patients attended the national clinics in Birmingham and London. Eight per cent of adults and 6% of children had end stage renal disease. Genotype and mutation type were statistically significant risk factors for severe renal disease (estimated glomerular filtration rate <45 ml/min/1.73m^2). Univariable logistical regression analysis indicated that genotypes *BBS2*, *BBS10* and *BBS12* were significantly more likely to be associated with severe renal disease than mutations in *BBS1* (p values: 0.02, 0.0004, 0.03 respectively). Compound truncating mutations and truncating/missense mutations were also statistically more likely to be associated with severe renal disease than compound missense mutations (p values= 0.0003, 0.02 respectively). Multivariable logistical regression analysis including known clinical renal risk factors (age, hypertension, diabetes, BMI) identified mutations in *BBS10* as an independent risk factor for severe renal disease (p=0.028). Only one patient homozygous for the common M390R mutation in *BBS1* developed end stage renal disease. **Conclusions**: This study describes the largest reported cohort of patients with renal disease in Bardet-Biedl syndrome. Here we identify for the first time valuable genetic risk factors which are directly applicable to prognosis, clinical risk stratification and genetic counselling. .

2595F

SDCCAG8 affects AIMP2 localization and alters its downstream target, p53. *K. Weihbrecht[1,2], C. Searby[1,2], W. Goar[1,2], V.C. Sheffield[1,2], S. Seo[2].* 1) Department of Pediatrics, Univ. of Iowa, Iowa City, IA; 2) Department of Ophthalmology, Univ. of Iowa, Iowa City, IA.

Bardet-Biedl Syndrome (BBS) is an autosomal recessive, pleiotropic disorder, with 21 genes currently identified as causative. One of these genes is serologically defined colon cancer antigen 8 (SDCCAG8 [MIM 613524]), a nephronophthisis-related ciliopathy gene (*NPHP10* or *BBS16*). Patients with mutations in this gene exhibit retinal and renal abnormalities, obesity, and learning disabilities. Currently, little is known about the molecular functions of SDCCAG8 and how loss of SDCCAG8 function leads to the observed phenotypes. Our previous work identified strong interactors of SDCCAG8, consisting mainly of components of the multi-aminoacyl tRNA synthetase complex (MSC). More specifically, 8 out of 9 aminoacyl tRNA synthetases (ARS) as well as aminoacyl-tRNA-synthetase-complex interacting multifunctional protein 2 (AIMP2) interact with SDCCAG8. We also determined that among the MSC components, SDCCAG8 directly interacts with AIMP2. AIMP2 has a known non-canonical role in the apoptotic pathway via p53 activation. In response to DNA damage, a subset of AIMP2 localizes to the nucleus and binds to and prevents the ubiquitination of p53. This leads to an upregulation of apoptosis in response to DNA damage. Recently, we developed two *SDCCAG8* knockout IMCD3 cell lines using CRISPR-mediated technology. Using these cell lines, we examined the effect of *SDCCAG8* loss on AIMP2 localization in these cells. Immunofluorescence microscopy and cell fractionation with immunoblotting show that loss of *SDCCAG8* leads to increased AIMP2 in the nucleus of cells. Additionally, we showed that overall p53 levels increased in knock-out cells. Further work focuses on how these changes affect downstream elements of p53 utilizing a luciferase reporter assay. Based on our early findings, we suspected that SDCCAG8 acts as a negative regulator of AIMP2 to the nucleus under normal cell conditions. However, recent findings show that SDCCAG8 is found at near equal levels in cytoplasmic and nuclear fractions, which suggests a possible role of SDCCAG8 as a chaperone or signal for AIMP2's nuclear localization. Loss of SDCCAG8 may cause a misregulation of normal localization, leading to the increased nuclear AIMP2. To determine the role of SDCCAG8 in this pathway, we will use cell fractionation followed by immunoprecipitation to determine if SDCCAG8 and AIMP2 are bound in the nucleus, which would suggest a role for SDCCAG8 as a chaperone of AIMP2.

2596W

Lipoid proteinosis: A clinical and molecular study in Egyptian patients.
H.H. Afifi[1], K.S. Amr[2], A.M.S. Tosson[3], T.A. Hassan[4], G.Y. El-Kamah[1]. 1) Clinical Genetics Department,Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 2) Medical Molecular Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 3) Pediatrics Department, Faculty of Medicine, Cairo University, Cairo, Egypt; 4) Otolaryngology Head and Neck Surgery Unit, Hearing and Speech Institute, Cairo, Egypt.

　Purpose: Lipoid proteinosis is an autosomal recessive disorder due to the loss of function of *ECM* gene. Clinical features include varying degrees of skin thickening, hoarseness of voice, beaded eyelid papules and less frequently neuropsychiatric abnormalities, and partial alopecia. **Objectives:** Clinical and molecular study of lipoid proteinosis in Egyptians. **Methods:** Clinical examination and molecular analysis of 9 individuals having lipoid proteinosis, and descending from 7 unrelated families, were performed. The families originated from different geographic governorates in Egypt. Extraction of DNA samples of the 9 patients and their parents from peripheral blood by standard methods was performed. Polymerase chain reaction (PCR) amplification of the *ECM1* gene was performed using eight pairs of primers spanning over the 10 exons and splice junctions **Summary and Results:** Patients exhibited a variety of clinical manifestations with skin affection and hoarseness of voice being the consistent feature in all of them. None of them had neuropsychiatric manifestations. We identified four novel homozygous insertion, small deletion, missense, and splice site mutations {c.10_11insC (exon 1), c.690_691delAG (exon 6), c.734G>A (exon 7), and c.1393G>T (intron 9/exon 10)}, as well as one previously published splice site mutation c.71G>C (exon1/intron 1). Parents were heterozygous for the mutation of their offspring. Novel mutations were not detected in 200 Egyptian controls. **Conclusions:** The novel mutation c.1393G>T (IVS10-1G>T) occurred in three unrelated Egyptian patients, which may suggest a founder effect of this allele in Egyptian patients. However, further studies are needed. Identification of pathogenic *EMC1* mutations is important for accurate diagnosis and proper genetic counseling.

2597T

N of 2: Assessing recurrent gene identifications with population and functional data. *S. Akle[1,2], C.A. Cassa[2,3].* 1) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA. USA; 2) Brigham and Women's Hospital, Division of Genetics, Boston, MA, USA; 3) Harvard Medical School, Department of Medicine, Boston, MA, USA.

　The expanding use of clinical genomic sequencing promises not only to improve diagnostics, but also to drive the discovery of new disease genes. When a promising candidate variant is identified in a gene that is not already associated with disease, a second case with a variant in that gene, or an "N of 2", increases the chance of causality. Second case hits may come from collaborating labs or clinics, through the matchmaker exchange, or even from within large clinical sequencing programs (Lee 2014; Yang 2014). Such recurrent observations can help to identify new disease-gene associations.　Due to large clinical sequencing cohorts and the Matchmaker Exchange, these recurrent gene identifications are becoming more common. However, these associations may be due to chance alone. We have previously shown that there is a substantial chance that two individuals in a cohort carry variants in some genes (Akle 2015). Our tool (RD-Match, http://genetics.bwh.harvard.edu/rdmatch/) assesses the likelihood that two individuals simply happen to carry variants in the same gene that is unrelated to the phenotype. This probability can be estimated using data from the Exome Aggregation Consortium, a set of over 60,000 exomes from healthy individuals (Lek 2015).　Here, we extend this work to assess recurrent gene identifications from clinical exome cases. As an example, in a set of 2,000 clinical exome cases, 291 cases involved recurrent genes that were related to a diagnosis. Over 90% of those genes were not available on clinical genetic testing panels (NCBI 2012). In the set of solved exome cases, 91% of these novel genes would not have been conclusively associated with a disease given the recurrence rates, modes of inheritance and disease grouping. However, clinical labs make use of data from other cases, additional phenotypic data, and other resources to make assessments of causality.　Next, we integrate functional validation data into the assessment of N=2 cases. We specifically include three lines of evidence from *in vivo* assays: two assays of cell essentiality (Wang, 2015) and embryonic or early lethal mouse knockout data (JAX/MGI). We calculate a composite probability of association given the probabilities derived from these separately ascertained lines of evidence.　Recurrent observations can expedite the identification of new genes in rare disorders, but investigators must carefully integrate statistical evidence for association with other sources of evidence.

2598F

The identification of novel causative variants and likely non-pathogenic loss-of-function variants in consanguineous families. *P. Makrythanasis[1,2], M. Ansar[1], M.T. Sarwar[3], H. Shaikh[4], S.A. Paracha[3], J. Khan[3], A.A. Khan[4], A. Suhail[3], Z. Iqbal[5], Y.M. Waryah[4], S. Temtamy[6], A. Megarbane[7], A. Masri[8], M.S. Aglan[6], M.S. Zaki[6], S. Kitsiou-Tzeli[9], H. Fryssira[9], N. Al-Allawi[10], A. Sefiani[11], S. Al-Hait[12], L. Al-Gazali[13], N. Jalkh[14], M. Guipponi[1,2], F.A. Santoni[1,2], H. Hamamy[1], E. Frengen[15], A.M. Waryah[4], J. Ahmed[3], S.E. Antonarakis[1,2,16].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 3) Institute of Basic Medical Sciences (IBMS), Khyber Medical University, Peshawar, Pakistan; 4) Molecular Biology and Genetics Department, Medical Research Center, Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan; 5) Department of Neurology, Oslo University Hospital, Oslo, Norway; 6) Department of Clinical Genetics, National Research Centre, Cairo, Egypt; 7) Institut Jerome Lejeune, Paris, France; 8) Pediatric Department, the University of Jordan, Amman, Jordan; 9) Department of Medical Genetics, University of Athens, Athens, Greece; 10) Department of Pathology, College of Medicine, University of Dohuk, Dohuk, Iraq; 11) Département de Génétique Médicale, Institut National d'Hygiène, Rabat, Morocco; 12) Genetics & IVF Department, the Farah Hospital, Amman, Jordan; 13) Department of Paediatrics, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates; 14) Medical Genetics Unit, Saint Joseph University, Beirut, Lebanon; 15) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 16) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.

   Consanguinity and inbreeding increase the level of homozygosity through IBD alleles, thus increasing the frequency of autosomal recessive disorders. We have ascertained 155 consanguineous families with several phenotypes compatible with recessive heredity; mainly intellectual disability/developmental delay or visual impairment, originating from Arab countries and Pakistan. To identify the disease causing variants, we used a combination of homozygosity mapping and exome sequencing. At least five individuals per nuclear family were genotyped and in one individual per family exome sequencing was performed. We have identified pathogenic variants in already known genes in 38% of the families, while 21% of the families presented likely causative variants in candidate genes. In 41% families no causative or candidate variants in the coding regions covered by exome sequencing was identified. We have also looked for (likely) non-pathogenic homozygous loss-of-function variants in order to contribute to the improvement of the diagnostic algorithms and reduce the false positive calls for pathogenicity. We have identified 84 homozygous non-sense variants which did not segregate with the recessive phenotype and therefore are likely tolerated. We conclude that the discovery of novel pathogenic variants for recessive disorders and non-pathogenic homozygous loss of function mutation can be accelerated by analyzing large cohorts of consanguineous families, which in-turn will be beneficial for correct molecular diagnosis, and counselling of the affected families. An international, open-access database of these variants would facilitate the functional analysis of protein-coding genes, and the diagnostic capabilities of (near) Mendelian disorders. .

2599W

Dual genetic diagnosis as a valuable feature of whole exome sequencing. *A.M. Bertoli-Avella[1], O. Brandau[1], K. Kandaswamy[1], N. Nahavandi[1], D. Trujillano[1], P. Bauer[1], R. Abou Jamra[2], A. Rolfs[1,3].* 1) medical reporting, Centogene AG, Rostock, mecklenburg vorpommern, Germany; 2) Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany; 3) Albrecht-Kossel-Institute, University of Rostock, Rostock, Germany.

   Abstract: Introduction: A significant proportion of patients who undergo extensive genetic testing remain undiagnosed. This is due to the fact that many approaches are focused on the identification of a single cause for the complete phenotypic picture of the patient. Such approaches might however miss the diagnosis in cases with multiple genetic disease phenotypes. This has serious consequences for the patients and their families, preventing the access to the right treatment or accurate counselling for pregnancies and prognosis. The advent of next-generation sequencing technologies has provided an opportunity to screen a patient's entire exome in order to establish genetic basis of diseases. This method for cases with unusual and blended clinical presentations is especially useful. Method: Whole exome sequencing (WES) was performed in 1555 index cases according to standardized settings. Data was generated and processed at CAP and CLIA certified laboratory, Centogene. In this cohort with a majority of childhood patients, we analysed the occurrence of clear dual or likely dual diagnoses. Results: We identified 35 families for whom whole exome sequencing provided a dual genetic diagnosis. These cases presented with a broader or atypical phenotype that could not be explained by a single variant. From these 35 families, pathogenic or likely pathogenic variants were reported for 8 families. For the remaining 27 a dual diagnosis seems a likely possibility due to the unexplained partial symptomatology and the identification of a pathogenic or likely pathogenic variant and at least one variant of uncertain significance with high pathogenicity scores in the corresponding gene. Some relevant examples were classic type of Ehlers-Danlos syndrome and X-linked Mental retardation type 102 (*COL5A2* and *DDX3X*), SESAME syndrome and Fanconi anemia, complementation group A (*KCNJ10* and *FANCA*) and Sanfilippo syndrome type B and ichthyosis vulgaris (*NAGLU* and *FLG*). Conclusion: Our result highlights the strength of an untargeted approach for the case of complex disease phenotypes. It clearly underlines the need for a complete standard procedure driven thorough exome analysis independent from an eventual observed diagnosis.   **Keyword**: whole exome sequencing; dual genetic diagnosis; molecular diagnosis.

2600T

Syndrome femur-fibula-ulna (FFU) report on additional findings as atypical presentation. *M. Garcia-Acero[1], J.C. Prieto[1,2].* 1) Instituto de Genetica Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Hospital La Victoria SDS, Bogotá, Colombia.

Introduction: Skeletal dysplasias occur in about 1 case in 4000-5000 live births, with clinical and genetic variability. The classification has been difficult to focus primarily on clinical and radiological criteria associated with histological and genetic criteria, comprising those alterations of growth, number and bone structure. Femur-fibula-ulna (FFU) complex is a non-lethal congenital anomaly of unknown etiology, characterized by a highly variable combination of defects of the femur, fibula, and/or ulna, with striking asymmetry. Axial skeleton, internal organs and intellectual function are usually normal. **Case presentation:** We present a male patient, son of non-consanguineous parents, with clinical evidence of shortening of right forearm, bilateral clubfoot, bilateral dimpled knees, postaxial lower limb polydactyly and syndactyly toes 2-5 right foot, with radiographic findings of agenesis of ulna and fibula rights associated with fusion of sacral vertebrae. **Discussion:** Clinical and radiological findings of the patient are compatible with atypical presentation syndrome Femur-Fibula-Ulna (FFU) group three with vertebral involvement, corresponding to a rare form of skeletal dysplasia limb. The reason to gather in a single syndrome (FFU) is their tendency to occur together in the same patient much more frequently than expected based on opportunities combination of a single defect. They are usually sporadic, affecting males, unilateral, asymmetrical and committing mainly to the upper extremities and the right side. Familial recurrence is exceptional and explaining only by chance, without being observed parent-child transmission or relationship with the age of the same or consanguinity, nor by environmental factors, infectious, iatrogenic or radiation. It has not been explained by chromosome studies, or as deletion or duplication syndromes. **Conclusion:** The FFU syndrome consists of a series of asymmetrical malformations in the limbs characterized by highly variable combinations of congenital anomalies of femur, fibula, and/or ulna, which can occur with abnormal finger/foot on the side of the ulna/fibula. Depending on the involvement of malformed bones, FFU complex is classified into four groups: with an affected limb, with two, three and four. .

2601F

Identification of *MLL2* and *SATB2* in two patients with syndromic craniofacial anomalies. *H. Hain[1], D. Li[1], M. Harr[1,2], E. Zackai[2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 2) Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Whole exome sequencing (WES) offers a powerful diagnostic tool to rapidly and efficiently sequence all coding genes in individuals presenting for consideration of phenotypically and genetically heterogeneous disorders such as syndromic craniofacial anomalies. Here we present two patients with syndromic form of cleft palate who tested negative on targeted gene sequencing. WES analysis followed by Sanger sequencing identified *de novo* and novel mutations in *MLL2* and *SATB2* in two patients, respectively. The first patient has hypertelorism, micrognathia, cleft lip and palate, right ear hearing loss, aplasia cutis and more recent scalp alopecia with pili torti, which fit in Kabuki syndrome. The second patient for the SATB2-associated syndrome has cleft palate, speech delay, intellectual disability, mild pectus excavatum, midfacial flattening, and mild bilateral fifth finger clinodactyly. These case reports highlight the need for careful diagnosis of clinically heterogeneous syndromes, however delayed molecular diagnosis is not likely to have implication for patient health. The report also demonstrates that WES is an unbiased method and free of the constraint of having to develop a hypothesis, and an effective diagnostic tool in providing molecular diagnosis to patients with complex phenotype.

2602W

A novel mutation in the *VIPAS39* gene found in two families with atypical arthrogryposis, renal dysfuntion and cholestasis (ARC) syndrome. *T. Kaname[1], Y. Chinen[2], K. Yanagi[1], M. Iso[1], Y. Matsui[1], K. Hayashi[1], Y. Kuroki[1], T. Tohma[3], Y. Matsubara[4], IRUD-P consortium.* 1) Dept Gen Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan; 2) Dept Pedatr, Univ Ryukyu Grad Schl Med, Okinawa, Japan; 3) Wanpaku Clinic, Okinawa, Japan; 4) Research Institute, National Center for Child Health and Development (Japan), Tokyo, Japan.

Arthrogryposis, renal dysfunction and cholestasis (ARC) syndrome is a rare autosomal recessive multisystem disorder that has neurogenic arthrogryposis, renal tubular dysfunction and intrahepatic cholestasis. ARC syndrome can be caused by homozygous or compound heterozygous mutations in the *VPS33B* gene, whose product acts in intracellular trafficking. In several patients with ARC syndrome, mutations in the *VIPAS39* gene, whose product is a subset of VPS33B-positive vesicles involved in intracellular trafficking, were reported We present six patients in four families (two patients in two families and one patient in two family) with pathologically atypical ARC syndrome. All patients were born to non-consanguineous Japanese parents. The patients had dysmorphic features, bilateral clubfeet, arthrogryposis multiplex, aminoaciduria, ichthyosis and hyperbilirubinemia. The liver biopsies and pathological examination were performed in the patients. Although, multinucleation of hepatocytes and intra hepatocytic bile stasis in porta were observed in the liver of the patients, carcinoembryoinc antigen (CEA) staining revealed atypical spreading pattern compared with the pattern in ARC syndrome. Sanger sequencing of the *VPS33B* gene resulted no pathogenic mutations in all patients. Then, whole exome sequencing analysis was performed in the patients and their families. Of four families, two families (four patients) had the same homozygous mutation around exon 15 of the *VIPAS39* gene (c.913-1G to A.) The splicing mutation affected the *VIPAS39* expression in the patients, which might be a cause of the syndrome. The two families having the same mutation are unrelated but from the same main island of Okinawa, suggesting that there might be a founder effect in this area.

2603T

Chromatin remodelling disorders: The evolving phenotype of EP300 gene syndrome. *M.I. Van Allen[1,2], L. Armstrong[1,2], M. Steinrath[1], K. Meagher[1], T. Oh[1], The CAUSES study.* 1) Dept Med Gen, Univ British Columbia, Vancouver, British Columbia, Canada; 2) Children's and Family Research Centre, B.C. Children's Hospital, Vancouver, B.C.

Identification of new gene disorders via Whole Exome Sequencing (WES) has enhanced our understanding of molecular pathways and the impact of deletions and mutations in critical gene functioning during embryogenesis and postnatally. We present 4 new cases of individuals with EP300 gene alterations: (1) A 4 y/o boy with intellectual disability (ID), short stature (SS), prematurity, gut ischaemia, Tetralogy of Fallot, absent thymus and parathyroid glands, feeding dysfunction (FD), and distinctive facial features with normal thumbs/toes who has a novel *de novo* c.3734_3736delTTG(p.Val1245del) in frame deletion in EP300; (2, 3) A mother with ID and her 4 y/o son with ID, autism (ASD), SS, microcephaly, FD and both with facial dysmorphia and normal thumbs/toes who have the previously reported EP300 c.6915_6918 deletion (p.Asn2305LysFs); (4) A 12 y/o girl with ID, ASD, SS, hydrocephalus, arachnoid cyst requiring a shunt, Chiari 1 malformation, hearing loss, urachal cyst, echogenic kidneys, chronic constipation, and facial dysmorphia (with normal thumbs/toes) who has a *de novo* EP300 c.4505C>T(p.Pro1502Leu) change. The EP300 gene encodes p300 gene, a histone acetyltransferase that regulates transcription via chromatin remodelling and is important in the processes of cell proliferation and differentiation. The clinical features of EP300 gene syndrome are distinctive for multisystem involvement, malformations, and dysmorphic facial features. The clinical presentation overlaps with and is distinct from Rubenstein Taybi, Cornelia de Lange and Wiedemann-Steiner syndromes. All these syndromes have multiple malformations, postnatal growth concerns, neurological dysfunction, and deficits of essential physiological functions, along with microcephaly and mild to more severe ID. Facial features of our patients are phenotypically similar, changing with age, and distinct from other disorders of chromatin modelling, We consider EP300 gene syndrome to be a discreet disorder with a recognizable phenotype, and not a subtype of Rubenstein Taybi syndrome.

2604F

Debunking Occam's razor: Diagnosing multiple genetic diseases by whole exome sequencing. *T.B. Balci[1], T. Hartley[2], Y. Xi[2], X. Yang[3], C.L. Beaulieu[2], A.M. Smith[2], C. Armour[1], J. Richer[1], M. Tetrault[4], J. Schwartzentruber[4], E. Bareke[4], D.A. Dyment[1,2], A.N. Prasad[5], C.A. Rupar[5], C. Prasad[5], R. Mendoza-Londono[6], L. Dupuis[6], B.A. Fernandez[7], G. Horvath[8], J. Majewski[4], A.M. Innes[3], K.M. Boycott[1,2] FORGE and Care4Rare Canada Consortiums.* 1) Medical Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 3) Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 4) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 5) Western University, London Health Sciences Centre, London, Ontario, Canada; 6) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Disciplines of Genetics and Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 8) Division of Biochemical Diseases, Department of Pediatrics, University of British Columbia and BC Children's Hospital, Vancouver, British Columbia, Canada.

   Purpose: Next generation sequencing has changed the landscape of rare disease diagnosis. Recent cohorts of patients who have undergone clinical whole exome sequencing (WES) have shown that while the identification of multiple genetic diagnoses in a single family is unanticipated, it is a reality. Here we outline the diagnostic odyssey of our cases with more than one genetic disease uncovered by WES and review published literature to reveal an overall rate for these events. Methods: We retrospectively analyzed the WES results for 773 families with undiagnosed rare diseases enrolled in the two national research programs in Canada, Finding of Rare Genetic Disease Genes and its successor, Enhanced Care for Rare Genetic Diseases. Reports of WES cohorts that mention cases with multiple diagnoses were reviewed for overall numbers. Results: In our series eight out of 216 solved families (3.6%) had two or more genetic diagnoses explaining their multiple findings. The overall rate of encountering multiple diagnoses among solved WES cases in recent reports, including ours was found to be 4.1% (1.4 - 14.3%). Conclusion: The rate of finding more than one genetic disease in patients with complex phenotypes is significant. These diagnoses of blended phenotypes would be almost impossible to reach without the relatively unbiased nature of WES (or related genome-wide sequencing tests). When initial results do not point to a single unifying cause and only explain part of a patient's findings, the medical geneticist should consider this rare occurrence in addition to, or perhaps even before, other possibilities, such as phenotypic variability.

2605W

Dominant, clustered mutations in desmoplakin (*DSP*) cause erythrokeratodermia-cardiomyopathy (EKC) syndrome, a novel cardio-cutaneous disorder, and impair desmosomal processing. *L.M. Boyden[1], C.Y. Kam[2], A. Hernandez-Martin[3], J. Zhou[1], B.G. Craiglow[1], L.M. Milstone[1], R. Sidbury[4], E.F. Mathes[5], S.M. Maguiness[6], D.A. Crumrine[5], M.L. Williams[5], R. Hu[1], K. Holland[7], A.S. Paller[2], R.P. Lifton[1], P.M. Elias[5], K.J. Green[2], K.A. Choate[1].* 1) Yale University, New Haven, CT; 2) Northwestern University, Chicago, IL; 3) Hospital Infantil del Nino Jesus, Madrid, Spain; 4) University of Washington, Seattle, WA; 5) University of California San Francisco, San Francisco, CA; 6) University of Minnesota, Minneapolis, MN; 7) Medical College of Wisconsin, Milwaukee, WI.

   Disorders of keratinization (DOK) feature marked genotypic and phenotypic heterogeneity. As most disease is cutaneous, the possibility of comorbidity is rarely investigated. In order to better understand the underlying pathobiology, we performed genetic and biochemical analyses of a DOK cohort. Via exome sequencing, we found that three subjects harbored *de novo* missense mutations within the gene encoding desmoplakin (*DSP*), clustered within a seven residue span of a single spectrin repeat. Subjects with these *DSP* mutations all exhibit a novel DOK, which we termed erythrokeratodermia-cardiomyopathy (EKC) syndrome, characterized by erythrokeratodermia and an initially-asymptomatic, progressive, and potentially fatal cardiomyopathy, a morbidity not previously associated with erythrokeratodermia. Desmoplakin is a primary component of desmosomes, the intercellular adhesion junctions most abundant in both epidermis and heart tissue. Affected skin shows aggregation of desmosomes, widening of intercellular spaces, and defects in lipid secretion. We demonstrate that *DSP* mutations in EKC syndrome subjects alter the epidermal localization of desmoplakin, other desmosomal proteins, and the gap junction protein connexin 43. We have since seen two additional probands, each found to have a *DSP* mutation in the same spectrin repeat, who also exhibit erythrokeratodermia and cardiomyopathy. Parental DNA from one of these subjects was not available; we have shown her *DSP* mutation is *de novo* via comparison of her chromosome 6 haplotypes to those of her siblings. Remarkably, her skin phenotype dramatically improved following immunosuppression necessitated by a heart transplant. While mutations in *DSP* are known to cause other disorders, our EKC subjects feature a unique, extreme whole-body erythorkeratodermia, with distinct molecular effects on desmosomal proteins and processing. The cardiac defects in EKC subjects, which were fatal in one, resulted in heart transplant in two, and were undiagnosed prior to our analysis in two, highlight the critical, potentially life-saving importance of this genetic discovery, which necessitates consideration of cardiac disease and genetic diagnosis in all patients presenting with erythrokeratodermia. These findings add a severe, previously undescribed syndrome to phenotypes caused by *DSP* mutation, and identify a specific region of the protein singularly critical to the pathobiology of EKC syndrome and to both epidermal and cardiac DSP function.

2606T

Alterations in the *NAA10* polyadenylation site in two families with X-linked anophthalmia. *J.J. Johnston[1], F.J. Stewart[2], H.M. Chapman[3], C.M. Chou[4], J.C. Sapp[1], A.A. Schäffer[5], S.K. Sen[1], A. Slavotinek[6], T.M. Glaser[3], G.C.M. Black[7,8], L.G. Biesecker[1].* 1) NHGRI, NIH, Bethesda MD; 2) NI Regional Genetics Service, Belfast, UK; 3) Department of Cell Biology and Human Anatomy, University of California, Davis School of Medicine, Davis, CA; 4) Departments of Human Genetics and Internal Medicine, University of Michigan Medical School, Ann Arbor, MI; 5) NCBI, NIH, Bethesda MD; 6) Department of Pediatrics, Division of Genetics, University of California, San Francisco, San Francisco, CA; 7) Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; 8) Centre for Ophthalmology & Vision Sciences, Institute of Human Development, University of Manchester, Manchester, UK.

Mutations in N-alpha-acetyltransferase 10 (*NAA10)* have been identified in several disease phenotypes including Ogden syndrome, Lenz microphthalmia and intellectual disability (ID), both with and without associated features. It is unclear how the alterations in *NAA10* lead to such widely variable phenotypes although substrate specific defects have been suggested. We evaluated a family with X-linked anophthalmia in which coding variants in *NAA10* were not identified. Re-examination of this pedigree based on linkage results and exome sequence analysis identified a non-coding alteration in the polyadenylation signal of *NAA10*. An additional family was identified with a similar phenotype and shown to have a different alteration in the polyadenylation signal. Neither variant was present in the ExAC database. RNA-seq and RT-PCR analysis of the RNA species present in family 1 suggest the use of a downstream polyadenylation signal and reduced overall mRNA levels. These two variants bring to 28 the total known variants in the polyadenylation site in humans associated with disease. The mutational spectrum of *NAA10* is distinct and peculiar. A splice variant has also been associated with syndromic microphthalmia. A missense mutation has been associated with Ogden syndrome. Three other missense variants have been associated with ID with or without LQTS. No loss of function mutations have been described and we presume these would be lethal in males. We propose a model whereby hypomorphic missense mutations with substrate specific effects lead to Ogden syndrome or ID, whereas RNA processing mutations that disrupt protein function cause syndromic microphthalmia. This represents a novel mechanism of disease as prior polyA site mutations, in other genes, have phenotypes similar to loss-of-function ORF mutations.

2607F

Recessive *NCOR2* mutations cause a new syndrome with intellectual disability, white matter anomalies, brachycephaly, and midface hypoplasia. *W. Li[1], B. de Vries[2], G. Mancini[3], A. Ryabets[7], H. Fu[4], T. Gardeitchik[2], A. van den Ouweland[3], M. van Slegtenhorst[3], A. Stegmann[6], X. Yang[5], M. Borchert[8], P. Campeau[4].* 1) Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA; 2) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands; 3) Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) Department of Pediatrics, CHU Sainte-Justine, University of Montreal, Montreal, QC, Canada; 5) Rosalind and Morris Goodman Cancer Research Center, Department of Biochemistry, McGill University and Department Medicine, McGill University of Health Center, Montreal, QC, Canada; 6) Department of Human Genetics, Maastricht University Hospital, Maastricht, The Netherlands; 7) Center for Endocrinology, Diabetes, and Metabolism Children's Hospital Los Angeles, Los Angeles, CA, USA; 8) Vision Center, Children's Hospital Los Angeles, Los Angeles, CA , USA.

Whole-exome sequencing of a pair of twins and two other individuals, from a total of three families, presenting with developmental delay or intellectual disability accompanied by brachycephaly, ventriculomegaly and midface hypoplasia identified bi-allelic mutations in *NCOR2*, a gene encoding the Silencing Mediator of Retinoic acid and Thyroid hormone receptor (SMRT). Features present in a majority of the patients were severe intellectual disability, speech impairment, hypotonia, epilepsy, ventriculomegaly, abnormal myelination, brachycephaly, midface hypoplasia, sparse hair, dysplastic ears, hearing loss, and feeding and respiratory difficulties. Allmutations identified were missense mutations affecting highly conserved residues in the C-terminal half of the protein, and were either not found or had very low minor allele frequencies in variant databases. SMRT is an essential component of the corepressor complex to which histone deacetylase 3 (HDAC3) is recruited for maximal histone deacetylase activity. HDAC3 can also be recruited to another homologous corepressor complex (NCoR), and binding to these complexes lead to a repression of certain nuclear receptors. Previous studies in mice showed *Ncor2* to be essential for forebrain development and neural stem cell maintenance, a finding correlating well with the severe neurological disease identified in the individuals described here. *In vitro* studies are currently underway to assess effects of the identified mutations on transcriptional corepressor functions of SMRT and the deacetylase activity of HDAC3. In conclusion, we define a new recessive condition with multisystem involvement caused by bi-allelic mutations in the SMRT corepressor encoding gene, *NCOR2*. .

2608W

Whole exome sequencing identifies genes converge on PI3K/mTOR and Ras/MAPK pathways as causes of heterogeneous complex lymphatic anomalies. *D. Li[1], T. Wenger[2], M. March[1], C. Seiler[3], L. Tian[1], C. Kao[1], R. Pandey[1], K. Nguyen[1], R. Chiavacci[1], P. Sleiman[1,4], M. Itkin[5], Y. Dori[5], H. Hakonarson[1,4,6].* 1) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Division of Craniofacial Medicine, Seattle Children's Hospital, Seattle, WA 98105, USA; 3) Zebrafish core, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 4) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; 5) Center for Lymphatic Imaging and Interventions, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 6) Divisions of Genetics and Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Complex lymphatic anomalies include several diagnoses with overlapping patterns of clinical symptoms and chylous effusions (pericardial, pleural, or peritoneal) can be present. To identify the underlying genetic basis, exome sequencing (ES) was performed in four families with either generalized lymphatic anomaly or lymphangiectasia diagnosis. In family-1, 6 affected individuals in multigenerational were revealed to have heterozygous *EPHB4* mutation (c.2334+1G>C), a gene previously implicated in a biological pathway related to venous and lymphatic cell fate determination. RNA-seq demonstrated that the *EPHB4* splice-altering mutation creates a cryptic splice donor that causes the retention of the intervening 12 bp of the intron. For family-2, ES revealed a homozygous variant, c.1393-7C>T, in *PIK3R6* in the proband with both parents being heterozygous. PIK3R6 is a regulatory subunit of PI3K complex. Zebrafish knockdown of either *EPHB4* or *PIK3R6* resulted in vessel misbranching and deformities in the lymphatic vessel development, indicative of possibly differentiation defects both in blood and lymphatic vessels and mimicking the presentations of the patients. Western blot analysis using zebrafish lysates, which contained vascular abnormality, confirmed that reduced EPHB4 signalling resulted in downstream mTORC1 overactivation. Strikingly, drugs that inhibit mTOR signalling were able to rescue this misbranching phenotype. In family 3 and 4, a recurrent *ARAF* mutation, c.640T>C:p.S214P, in the conserved phosphorylation site was identified, which putatively result in a ARAF gain of function because the phosphorylation of S214 is responsible for regulation of ARAF. ARAF activation in turn upregulates RFFL, leading to polyubiquitylation and destabilization of PRR5L, a component of mTORC2 and suppressor of PKC phosphorylation, to achieve persistent PKC activation and lead to cell pro-growth and pro-migration. We are currently recruiting ~10 unrelated patients to further validate our discovery and conducting follow-up functional assessments and the zebrafish model to elucidate the role these genes play in lymphatic development. The resulting functional data will be presented together with detailed phenotypic characterization of the additional families examined. In summary we report three novel genes that converge on PI3K/mTOR and Ras/MAPK pathways, presenting a potential avenue for therapeutic intervention for complex lymphatic anomalies.

2609T

Functional characterization of a point mutation in the *MAPK1* gene in a patient with undiagnosed disease. *N. Sosonkina[1], A. Takizawa[2], J.W. Prokop[1], T. Goto[3], M. Grzybowski[2], A. Geurts[2], M. Hirabayashi[3], J. Lazar[1], H.J. Jacob[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Medical College of Wisconsin, Milwaukee, WI; 3) National Institute for Physiological Sciences, Okazaki, Aichi, Japan.

Patients with rare diseases often undergo an exhausting and expensive diagnostic odyssey. Advancements in DNA sequencing, whole exome (WE) and whole genome sequencing (WGS), have offered the revolutionary diagnostic tools to these patients leading to cures and therapies. However, a considerable number of found variants fall into category of variant of uncertain significance (VUS). To make a diagnosis more accurate, the pathophysiological role of a VUS needs to be explained. Here, we report the results of the characterization of a *de novo* variant c.404G>C (p.R135T) in the *MAPK1* gene found using WES in a female proband with the following major symptoms: microcephaly, intellectual disability, speech apraxia, ADHD, craniosynostosis, facial dysmorphic features, and cyclic vomiting syndrome. This *MAPK1* missense mutation was classified as a VUS. We first utilized our deep-Sequence-to-Structure-to-Function analysis tools and suggested that the variant may affect MAPK1 activation. We then sought for creating of what would be the first example of a transgenic rat with knocked-in point mutation to mimic patient variant and potentially patient phenotype. The substitution of C for the wild-type G at nt 398 of the rat *Mapk1* would mimic the patient mutation. We used different combinations of CRISPR/Cas9 system, and manipulated rat fertilized 1-cell embryos. As a result, we obtained one female founder. Another approach was to modify rat ES cells utilizing CRISPR/Cas9 system aiming at knocking-in the *MAPK1* variant. We established two heterozygous ES cell lines that will be used in subsequent experiments. In parallel, we have received skin biopsies from the patient and her parents and performed cell-based assays to investigate the function of the *MAPK1* variant. Interestingly, gene expression studies showed overrepresentation of mutant *MAPK1* allele. In conclusion, we have evaluated different approaches to functional characterization of a VUS found in a patient with rare disease.

2610F

Whole exome sequencing identifies a novel *TDRD7* mutation responsible for the comorbidity of congenital cataract and azoospermia in humans. *Y. Tan[1,2], P. Liang[3], S. Yuan[1], L. Meng[1], C. Sjaarda[3], G. Lu[2], G. Lin[1,2].* 1) Inst. Reproductive and Stem Cell Engineering, Xiangya School of Medicine, Central South University, Changsha, Hunan, China; 2) Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha, China; 3) Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada.

Congenital cataract and azoospermia (both have multiple OMIM entries) in humans are two different diseases occurring in two different organs with very different symptoms and no similarity in known genetic etiology. Here we report the comorbidity of the two diseases in a consanguineous family, in which two brothers and one sister from the same marriage were all affected with congenital cataract, while the two brothers were also affected with infertility as unexplained azoospermia. Via whole exome sequencing of a single patient, we were able to narrow down the candidate causative mutation to a novel frame-shift insertion mutation (NM_014290:c.553_554insA, NP_055105:p.T110N) in the Tudor domain containing 7 (*TDRD7*) gene. The mutation is predicted to generate a severely truncated protein which lacks most of the functional domains in the TDRD7 protein. We verified the presence of this mutation by targeted genotyping and showed it as the only mutation following the expected Mendelian segregation pattern for an autosomal recessive allele in the 4-generation pedigree. Immuno-histological and western blot analyses demonstrated the absence of TDRD7 protein in the seminiferous tubules and skin fibroblasts of the male patients, respectively. Histological and FISH analyses of the testis biopsy samples demonstrated the presence of haploid spermatids but lack of mature sperms in the seminiferous tubules in the male patients. While mutations in *TDRD7* have been previously shown to be associated with congenital cataract in humans and to cause congenital cataract and abnormal spermatogenesis in *Tdrd7-/-* mice, our study represents the first report demonstrating a comorbidity of congenital cataract and azoospermia in humans. The link between these two diseases by TDRD7 suggests that despite involving different organs and cell types with very different cellular morphologies and functions, the developments of lens and sperm share a common biological process in post-transcription regulation, in which TDRD7 plays a critical role as a component of the RNA-granules known to function in RNA-processing. Our results warrant the addition of *TDRD7* to gene panels for genetic screening of cataract and male infertility.

2611W

Utilizing nonsense-mediated decay facilitates candidate disease gene characterization and discovery. *J.J. White[1], Z.C. Akdemir[1], S.N. Jhangiani[2], T. Gambin[3], E. Boerwinkle[2,4], R.A. Gibbs[1,2], C.M.B. Carvalho[1], J.R. Lupski[1,2,5,6], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw 00-665, Poland; 4) Human Genetics Center, University of Texas Health Science Center at Houston, TX 770530, USA; 5) Texas Children's Hospital, Houston, TX 77030, USA; 6) Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA.

Nonsense-mediated decay (NMD) is an RNA surveillance pathway that results in degradation of mRNA transcripts bearing a premature termination codon (PTC). According to the most widely accepted model of NMD, mutant transcripts containing PTCs located in the last ~50-55bp of the penultimate exon and the entire last exon, can escape NMD resulting in translation of a mutant protein product. To bioinformatically select transcripts with variants that may escape from NMD and possibly underlie human disease via a dominant-negative or gain-of-function mechanism, we investigated our in-house cohort of ~5000 exomes consisting of a wide variety of Mendelian phenotypes. We developed an algorithm to computationally predict the location of stop codons in either the -1 or +1 frame for each transcript. We categorized PTCs into those capable of escaping from NMD (NMD-incompetent) and those likely subject to NMD (NMD-competent). By flagging the final NMD-competent PTC we identified the boundary between NMD-competent and NMD-incompetent regions for each transcript allowing us to rapidly predict whether or not frameshift variants can escape from NMD. To identify possible pathogenic NMD-incompetent variants among our disease cohort, we used a binomial test to compare the frequency and distribution of frameshift variants in the predicted NMD-competent and NMD-incompetent transcript regions with the ARIC database of ~11,000 exomes as a control dataset. To further prioritize candidate genes, we designed an NMD-incompetency score metric to rank each gene based on the enrichment of NMD-incompetent vs NMD-competent variants. Among the top ranking genes, we identified examples where protein truncating variants result in disease by a mechanism other than haploinsufficiency. These include *DVL1* (autosomal dominant Robinow syndrome MIM #616331), in which frameshift alleles cluster in the penultimate exon, are hypothesized to escape NMD and mediate pathogenicity via a gain-of-function and/or dominant-negative mechanism (PMID: 25817016). Collectively, we demonstrate that leveraging the knowledge of NMD and an NMD-incompetency score metric may be an effective and efficient tool for discovery of novel disease genes due to production of truncated or altered proteins. Furthermore, we suggest that gain-of-function and dominant-negative mutations are under-recognized in genomic analyses and likely contribute to a wide variety of human disease phenotypes.

2612T

Trio whole exome sequencing identifies *de novo* **pathogenic variants in** *FBN1* **and** *TRPS1* **in a patient with a suspected connective tissue disorder.** *D.B. Zastrow[1,2], L. Fernandez[1,2], J. Kohler[1,2], P. Zornio[1,2], A. Dries[1,2], D. Waggott[1,2], M. Manning[3,4], E. Farrelly[5], A. Hanson-Kahn[5,6], K. Miller[5], M. Walkiewicz[7], C. Eng[7], E.A. Ashley[1,2,6], P. Fisher[1,4,8], J.A. Bernstein[1,4,5], M. Wheeler[1,2], Members of the Undiagnosed Diseases Network.* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 3) Department of Pathology, Stanford School of Medicine, Stanford, CA; 4) Department of Pediatrics, Stanford School of Medicine, Stanford, CA; 5) Lucille Packard Children's Hospital Stanford, Palo Alto, CA; 6) Department of Genetics, Stanford School of Medicine, Stanford, CA; 7) Baylor Miraca Genetics Laboratories, Houston, TX; 8) Department of Neurology, Stanford School of Medicine, Stanford, CA.

   A patient presented with a history of congenital diaphragmatic hernia, inguinal hernia, and recurrent umbilical hernia. She also had joint laxity, hypotonia, and dysmorphic facies. A unifying diagnosis was not identified based on clinical phenotype. As part of her evaluation in the Undiagnosed Diseases Network, trio whole exome sequencing (WES) was performed. After variant analysis and filtering, 89 de novo protein altering variants in 61 genes were found; 2 of these were consistent with the patient's clinical phenotype. We identified pathogenic stop gain variants in Fibrillin 1 (*FBN1* [MIM 134797]) and Zinc Finger Transcription Factor TRPS1 (*TRPS1* [MIM 604386]) resulting in 2 independent autosomal dominant conditions, each with de novo inheritance. A heterozygous known pathogenic variant c.4786C>T (p.R1596X, ClinVar ID 36082) in *FBN1* was detected. *FBN1* mutations are associated with Marfan syndrome (MFS [MIM 154700]) and a spectrum of similar phenotypes. A heterozygous c.1630C>T (p.R544X) pathogenic variant in *TRPS1* was also detected. *TRPS1* mutations are associated with Trichorhinophalangeal syndromes type I (TRPS1 [MIM 190350]) and type III (TRPS3 [MIM 190351]). Features of both MFS (aortic root enlargement (z=3.24), joint laxity and MFS Systemic Score=5) and TRPS1 (sparse hair, bulbous tip of the nose, long flat philtrum, thin upper vermilion border, protruding ears) are seen in the affected individual. MFS commonly presents with long bone overgrowth leading to increased height, which is in contrast to the short stature usually seen in TRPS1 due to premature closure of the growth plates. The height of our patient (97.0 cm: 31rst percentile by CDC; 17th percentile by WHO) may be a result of the differential effects of *FBN1* and *TRPS1* molecular mechanisms on bone formation. This discrepancy and young age of the proband made a clinical diagnosis difficult without genetic testing. Diagnostic criteria for MFS combine family history, medical history, physical exam, and molecular *FBN1* analysis. Our patient did not meet clinical criteria for MFS before WES. It was only after the de novo nonsense *FBN1* variant was found, and genotype-driven echocardiography revealed a mild aortic root enlargement that a diagnosis of MFS could be established. For isolated cases with a suspected genetic etiology, trio WES can allow for a more thorough interrogation of multiple genetic loci than single gene testing, and uncover etiologies responsible for blended phenotypes.

2613F

Expanded phenotypic spectrum of autosomal recessive EGFR-related disease due to a founder mutation in the Roma population: Report of a non-lethal case. *M. Szybowska[1], P. Kannu[2], E. Pope[3], G. Richards[4], I. Lara-Corrales[3], C. Li[1].* 1) Department of Genetics, McMaster Children's Hospital, Hamilton, Ontario, Canada; 2) Department of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Dermatology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) GeneDx Inc, Gaithersburg, MD, USA.

   The epidermal growth factor receptor (EGFR) is known to be involved in modulating the growth and differentiation of multiple tissue types including the heart, brain, epithelium, and hair. It accomplishes this via direct ligand-mediated extracellular to nuclear transmembrane signalling as well as indirect neighbouring pathways. Somatic gain of function mutations affecting the EGF/EGFR pathway are a well-known cause of cancer. Since 2014, a constitutional homozygous loss of function mutation [c.1283G>A; p.G428D] in EGFR has been described in three cases of Roma descent presenting with lethal inflammatory skin and bowel disease [Campbell, et al, 2014, Ganatzsky et al, 2015]. We report a case of a 5 year old girl, ex-27 weeker, of Roma descent who presented at birth with electrolyte imbalance, complex congenital heart defect, cystic kidneys with nephrophthisis, ventriculomegaly with a small subependymal cyst, interstitial lung disease**,** and antenatal polyhydramnios. Physical exam was significant for severe short stature and craniofacial dysmorphism. Dermatological findings are characterized by very thin and translucent skin that broke down easily in the newborn period. The skin lesions evolved to become pustular rashes, prominent on scalp, face and extremities. Loose and wrinkly skin was noted at 4 months assessment, giving rise to progeroid appearance. She also had sparse scalp hair with abnormal texture. Whole exome sequencing revealed the same homozygous EGFR mutations [c.1283G>A; p.G428D] previously reported in the other cases. Despite her very complex medical history, her skin findings have improved overtime and the frequency of infections decreased. She requires regular management for chronic renal disease stage II-III, Fanconi like syndrome, failure to thrive, chronic gastritis with elevated liver enzymes and chronic diarrhea. Presently, she is verbal and ambulatory with mild global developmental delay. It is not clear whether this delay is intrinsic of this syndrome or a consequence of her early neonatal course. In summary, we present a nonlethal case with the same homozygous mutation previous noted to be associated with lethal skin and GI inflammation. This case contributes to our understanding of the natural history and prognosis of a new syndrome. It also highlights the utility of whole exome sequencing in the diagnosis of genodermatoses.

2614W

Non-penetrance in cerebro-costo-mandibular syndrome. *D. Lynch[1], A.M. Innes[1,2], R.E. Lamont[1,2], J.S. Parboosingh[1,2], F.P. Bernier[1,2].* 1) Medical Genetics, University of Calgary, Calgary, Alberta, Canada; 2) Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada.

Cerebro-costo-mandibular syndrome (CCMS) is a rare disorder of skeletal development characterized by micrognathia, posterior rib gaps, and other variable features. We previously identified heterozygous mutations in an alternative exon of the spliceosomal gene *SNRPB* as the cause of CCMS, establishing that CCMS is a dominant disorder. This discovery allows us to offer diagnostic testing and improved genetic counseling to affected individuals and families. In our cohort of seventeen families, we observed both vertical transmission and *de novo* mutations. There were also four instances in three families where an unaffected mother had a heterozygous *SNRPB* mutation, indicating that CCMS is not fully penetrant. The identified mutations lead to deregulation of levels of the three SNRPB transcripts, and patients have a 32-42% reduction in total SNRPB RNA levels compared to controls. Analysis in one non-penetrant family shows that the unaffected mother does not have the reduction in SNRPB transcript levels that is typical in affected patients, which suggests a compensatory mechanism at the RNA level. Identifying the cause of this compensatory mechanism is important from a genetic counseling perspective, but also to elucidate the aetiology of CCMS and the mechanisms by which *SNRPB* is regulated. Sequence capture and massively parallel sequencing of *SNRPB* including coding, intronic, and flanking regions as well as a known *SNRPB* enhancer did not reveal any genetic differences between affected and unaffected family members sharing a *SNRPB* mutation. Preliminary qRT-PCR results in one family show that RNA levels of the *SNRPB* paralog *SNRPN* are increased by 45% in the unaffected mother, but remain unchanged in her affected child. While modifying factors remain unknown and could be genetic, environmental, or both, our preliminary analysis excludes variants in *SNRPB* and its known regulatory regions.

2615T

Genetic and functional comparison of identical germline and somatic *SETBP1* mutations. *R. Acuna-Hidalgo[1], P. Deriziotis[2], G.M. Findlay[3], C. Gilissen[1], M. Steehouwer[1], J.A. Veltman[1,4], S.E. Fisher[2], J. Shendure[3], A. Hoischen[1], B.W. van Bon[1], International Schinzel-Giedion syndrome collaboration.* 1) Department of Human Genetics, Radboud UMC, Nijmegen, Netherlands; 2) Department of Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 3) Department of Genome Sciences, University of Washington, Seattle, WA, United States; 4) Department of Clinical Genetics, University of Maastricht, Maastricht, Netherlands.

Schinzel-Giedion syndrome (SGS) is a rare developmental disorder characterized by multiple malformations, severe neurologic alterations and increased risk of malignancy. SGS is caused by *de novo* germline mutations clustering to a 12 bp hotspot in exon 4 of *SETBP1*. Mutations in this hotspot disrupt a degron (a signal for the regulation of protein degradation) and lead to the accumulation of SETBP1 protein. Identical *SETBP1* hotspot mutations have been observed recurrently as somatic events in leukemia. We collected clinical information of 47 SGS patients (including 26 novel cases) with germline *SETBP1* mutations. Comparing SETBP1 mutations in SGS and in leukemia shows that, despite sharing an overlap, the distribution of mutations within the hotspot differs between both conditions. Overexpression experiments confirm that mutations in the *SETBP1* hotspot destroy a degron and lead to increased SETBP1 protein, with different mutations having varying effects on SETBP1 stability and protein levels. Furthermore, we perform saturation genome editing (Findlay et al. Nature 2014) to analyze the effect of all theoretically possible mutations in the *SETBP1* hotspot and surrounding region. *In silico* modeling and functional studies substantiate that mutations seen more frequently in SGS are functionally weaker than those more prevalent in leukemia. Clinical data supports this finding; SGS patients with germline *SETBP1* mutations shown to be functionally stronger and more prevalent in leukemia show increased cell proliferation *in vitro*, increased incidence of malignancy and decreased survival. Our findings substantiate that, despite their overlap, somatic *SETBP1* mutations driving malignancy are more disruptive to the degron than germline *SETBP1* mutations causing SGS. Drawing on previous studies of somatic *SETBP1* mutations in leukemia, our results support a genotype-phenotype correlation in germline *SETBP1* mutations spanning a molecular, cellular and clinical phenotype.

2616F

Elucidation of *MRAS*-mediated Noonan syndrome with cardiac hypertrophy. *E.M. Higgins[1], J.M. Bos[1], H. Mason-Suares[2], K. Sol-Church[3], K.W. Gripp[3], R.A. Urrutia[1], M.J. Ackerman[1].* 1) Mayo Clinic, Rochester, MN; 2) Partners HealthCare, Boston, MA; 3) Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE.

Noonan syndrome (NS[MIM 163950]) is an autosomal dominant disorder with a characteristic clinical phenotype of facial dysmorphisms, short stature, and congenital heart defects. NS is a member of a family of developmental disorders termed "Rasopathies", caused mainly by gain-of-function mutations in genes encoding RAS-mitogen-activated protein kinase (MAPK) signaling pathway proteins. For our study, whole exome sequencing (WES) and trio-based genomic triangulation was performed on a 15-year old, female with a clinical diagnosis of NS and concomitant cardiac hypertrophy and her unaffected parents. Sequencing data was analyzed using Golden Helix VarSeq software. Filters were applied to identify rare (Sporadic: MAF<.00005 in ExAC, Recessive: MAF<.005 in ExAC), protein altering mutations that were either sporadic, de novo or inherited in a homozygous or compound heterozygous recessive manner. Candidate genes were ranked by ToppGene and Endeavour disease-network analysis candidate gene ranking algorithms based on association to the currently known NS causative genes. A single candidate gene that directly interacts with known NS-associated proteins was identified, and subsequently, a cohort of 110 phenotype-positive, genotype-negative NS patients was sequenced for this gene. Following analysis and filtering, WES identified a total of 176 candidate variants in 67 genes. A cumulative "rank list" based on the ranking software readouts identified *MRAS* (p.G23V, c.68G>T) as the most likely novel NS-susceptibility gene. *MRAS*-encoded RAS-related protein 3 (MRAS[MIM 608435]) interacts directly with the MAPK pathway via SHOC2 and PP1. p.G23V-MRAS localizes to a critical GTP-binding motif and is predicted to result in a gain-of-function by *in silico* prediction tools and by molecular dynamic simulation (MDS). Subsequent analysis of the phenotype-positive, genotype-negative NS cohort identified an additional rare, de novo p.T68I (c.203C>T) *MRAS* missense mutation in a female with NS, pulmonic valve stenosis and cardiac hypertrophy. Herein, with the discovery of two sporadic, de novo variants (p.G23V and p.T68I) identified by WES or direct DNA sequencing, we have established *MRAS* as the newest NS-susceptibility gene. MRAS plays a crucial role in the NS associated Ras/MAPK pathway and interacts directly with several previously identified NS associated proteins.

2617W

***AMMECR1*: A single point mutation causes developmental delay, midface hypoplasia, and elliptocytosis.** *E.G. Seaby[1], G. Andreoletti[1], J.M. Dewing[2], I. O'Kelly[2], K. Lachlan[3], R.D. Gilbert[4], S. Ennis[1].* 1) Human Genetics & Genomic Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Southampton, SO16 6YD, UK; 2) Centre for Human Development, Stem Cells and Regeneration HDH, University of Southampton, IDS Building, Southampton General Hospital, Southampton, SO16 6YD, UK; 3) Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Princess Anne Hospital, Coxford Road, Southampton, UK; 4) Wessex Regional Paediatric Nephro-Urology Service, Southampton Children's Hospital, Southampton, UK and Faculty of Medicine, University of Southampton.

Deletions in the Xq22.3 – Xq23 region, inclusive of *COL4A5*, have been associated with a contiguous gene deletion syndrome characterised by Alport syndrome with intellectual disability (Mental retardation), Midface hypoplasia, and Elliptocytosis (AMME). The extra-renal biological and clinical significance of neighbouring genes to the Alport locus has been largely speculative. Whole exome sequencing was undertaken on maternal half siblings presenting with a phenotype of nephrocalcinosis, early speech and language delay and midface hypoplasia with submucous cleft palate and bifid uvula. In-house genomic analysis included extraction of all shared variants on the X chromosome in keeping with X-linked inheritance. Patient specific mutants were transfected into three cell lines and microscopically visualised to assess the nuclear expression pattern of the mutant protein. In the affected half-brothers, we identified a hemizygous novel nonsynonymous variant of unknown significance in *AMMECR1* (c.G530A; p.G177D), a gene residing in the AMME disease locus. Transfected cell lines with the p.G177D mutation showed aberrant nuclear localisation patterns when compared to the wildtype. Both brothers were then tested for elliptocytosis, which was present in the older sibling. Our study shows that a single missense mutation in *AMMECR1* causes a phenotype of elliptocytosis, midface hypoplasia and mild intellectual disability, previously reported as part of a contiguous gene deletion syndrome. Functional analysis confirms mutant-specific protein dysfunction. We conclude that *AMMECR1* is a critical gene in the pathogenesis of AMME, causing early speech and language delay, infantile hypotonia, midface hypoplasia, elliptocytosis, nephrocalcinosis, sensorineural hearing loss and submucous cleft palate.

2618T

A novel missense mutation in *PUF60* causes syndromic intellectual disability. *D.K. Bourque[1,2], K. Kernohan[3], T. Hartley[3], D. Pohl[1,2,3], K. Boycott[1,3], D. Dyment[1,3], Care4Rare Consortium.* 1) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) University of Ottawa, Ottawa, Ontario, Canada; 3) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.

Intellectual disability (ID) affects 1-3% of the general population and up to 50% of those affected have an underlying genetic cause. Next-generation sequencing offers an efficient means to identify the underlying molecular causes of syndromic ID. Our patient presented to Genetics at 17 years of age with a history of a complex, syndromic form of ID. On past history, she was born to a healthy woman, induced at 36 weeks gestational age after a pregnancy complicated by severe intrauterine growth restriction (IUGR). Her neonatal period was characterized by poor weight gain, hypotonia and congenital ptosis. In childhood, she experienced significant global developmental delay, as well as generalized tonic-clonic seizures, hypothyroidism, delayed puberty, scoliosis, and a fatty infiltration of the filum terminale. At last assessment, she continues to have short stature, microcephaly, and dysmorphic features (including bilateral ptosis, synophrys, flattened nasal bridge, and long philtrum). Comprehensive standard and targeted cytogenetic, molecular, biochemical, and metabolic workups over the course of her life did not reveal a cause for her disorder. She was enrolled in the Care4Rare Consortium and whole exome sequencing was performed on both her and her unaffected parents. She was found to have a de novo missense mutation in exon 6 of *PUF60* (c.G389A, p.Arg130His) on chromosome 8q24.3. This variant had never been seen in our local database, 1000 Genomes, EVS or ExAC. SIFT, PolyPhen-2, and CADD scores were all predicted to be deleterious. *PUF60* is a splicing factor involved in several nuclear processes, including transcriptional regulation. Haploinsufficiency of *PUF60* has been implicated in the phenotypic manifestations of the 8q24.3 deletion syndrome. An SNV in *PUF60* (c.C505T, p.His169Tyr) has been reported in one other individual, who had similar features to our patient (IUGR, postnatal growth retardation, microcephaly, global developmental delay, seizures, scoliosis, fatty filum terminale, and tethered cord); although, it was not associated with any endocrinologic abnormalities (Dauber *et al.* 2013). Functional analysis of that mutation revealed a loss of function effect on the PUF60 protein. Functional validation studies of our newly reported mutation are ongoing.

2619F

De novo **deleterious variants in *SON* are the cause of syndromic intellectual disability.** *M. Tokita[1,2], A. Braxton[1,2], Y. Shao[1,3], A. Lewis[1,3], M. Vincent[4], S. Küry[4], T. Besnard[4], B. Isidor[4,5], X. Latypova[4], S. Bézieau[4], P. Liu[1,2], C. Motter[6], C. Melver[6], N. Robin[7], E. Infante[8], M. McGuire[2,8], A. El-Gharbawy[8], R. Littlejohn[9], S. McLean[9], W. Bi[1,2], C. Bacino[1,3], S. Lalani[1,3], D. Scott[1,3,10], C. Eng[1,2,3], Y. Yang[1,2,11], C. Schaaf[1,3,11], M. Walkiewicz[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Texas Children's Hospital, Houston, TX; 4) CHU de Nantes, Service de Génétique Médicale, Nantes, France; 5) INSERM, UMR-S 957, Nantes, France; 6) Genetic Center, Akron Children's Hospital, Akron, OH; 7) Department of Genetics, University of Alabama, Birmingham, AL; 8) University of Pittsburgh School of Medicine, Children's Hospital of Pittsburgh, Pittsburgh, PA; 9) Children's Hospital of San Antonio, San Antonio, TX; 10) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 11) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX.

Through its role as a mediator of constitutive and alternative splicing, the SON protein influences diverse cellular processes including cell cycle progression, genomic integrity, and maintenance of stem cell pluripotency. In spite of the clear functional importance of SON, the phenotypic effects of deleterious variants in the *SON* gene have not been clearly established. In our WES cohort of over 6,000 individuals, we identified six patients with *de novo* truncating variants in *SON*. The detected variants included one nonsense variant in exon 3, three distinct frameshifting variants in exon 3, and one frameshifting variant in exon 4. We then identified an additional subject with two *de novo* missense variants in exon 3 in *cis* configuration. Further analysis revealed substantial phenotypic overlap among the seven patients with *SON* variants. All seven patients had developmental delay and/or moderate-severe intellectual disability. Three of the seven patients had a history of regression, three had a diagnosis of autism spectrum disorder, and four had documented EEG abnormalities. Of six patients evaluated, five had findings suggestive of cerebral volume loss on imaging. All seven patients had a history of severe feeding difficulties. Variable congenital malformations were also present in the cohort. Mildly dysmorphic features were apparent in all seven patients, the majority of whom had midface retrusion with deep set eyes, down-slanting palpebral fissures, and a smooth or short philtrum. To further explore the potential clinical relevance of *SON*, we reviewed published reports of individuals with copy number variants encompassing *SON*. An analysis of eight phenotypically characterized patients with deletions of *SON* less than 5 Mb in size revealed a severe phenotype that was highly consistent with the features observed in our cohort. Developmental delay, brain anomalies, poor feeding, and congenital malformations were all common among the microdeletion patients. Collectively, this data suggests that deleterious variants in *SON* are the cause of a severe multisystem disorder characterized by intellectual disability, congenital anomalies, and feeding difficulties. Furthermore, the phenotypic overlap between patients with sequence variants and those with microdeletions indicates haploinsufficiency may be the mechanism of disease causation, however, further work will be necessary to confirm this.

2620W

Missense mutations in the *PLK4* gene identified in a patient with autosomal recessive microcephaly and chorioretinopathy. *M. Tsutsumi[1], S. Yokoi[1,2], F. Miya[3,4], M. Miyata[5], M. Kato[6,13], N. Okamoto[7], T. Tsunoda[3,4], M. Yamasaki[8], Y. Kanemura[9,10], K. Kosaki[11], S. Saitoh[12], H. Kurahashi[1].* 1) Division of Molecular Genetics, ICMS, Fujita Health University, Toyoake, Japan; 2) Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 4) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan; 6) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 7) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 8) Department of Pediatric Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 9) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 10) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 11) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 12) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 13) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan.

It has been well documented that mutations in genes encoding centrosomal proteins cause primary autosomal recessive microcephaly, although the association between centrosomal defects and the etiology of microcephaly syndromes is not fully understood. Recently, mutations in the *PLK4 (polo-like kinase 4)* gene, one of the key regulators of centriole duplication, identified in patients with microcephaly, growth failure and retinopathy has been reported. In this study, we performed whole exome sequencing analysis of an autosomal recessive microcephaly and chorioretinopathy (MCCRP2 [MIM616171]) patient and identified novel compound heterozygous mutations in the *PLK4* gene. One of these mutations, c.442A>G (p.(M148V)), resides in the kinase domain, and the other, c.2336G>A (p.(C779Y)), in the polo-box domain. Lymphoblastoid cell line derived from this patient showed that 25% of the mitotic cells had an aberrant spindle morphology, whereas >90% of the control mitotic cells were bipolar. Overexpression experiments of the mutant PLK4 proteins demonstrated that the p.C779Y had no effect on centriole amplification, suggesting a loss of function in centriole duplication. On the other hand, overexpression of the p.M148V proteins could amplify the centrioles and produce excess centrosomes to the same extent as wild-type PLK4. HeLa lysates transfected with the PLK4 expression plasmids were analyzed on a western blot. The altered mobility pattern of both mutant proteins further suggested alterations in post-translation modification, indicating that the p.M148V mutant is not functionally equivalent to wild-type PLK4. Our data lend support to the hypothesis that impaired centriole duplication caused by *PLK4* mutations may reduce mitosis efficiency, thereby inducing cell death or reduced cellular proliferation during early embryogenesis which is involved in the etiology of microcephaly disorder.

2621T

De novo *ACTB* and paternally inherited *COL13A1* missense mutations in a very atypical Baraitser-Winter syndrome case with familial history of joint hypermobility. *D.T. Uehara[1], S. Hayashi[1,2,3], K. Tanimoto[1], Y. Chinen[4], J. Inazawa[1,2,5].* 1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 2) Hard Tissue Genome Center, Tokyo Medical and Dental University, Tokyo, Japan; 3) Department of Neurobiology, Yale University School of Medicine, New Haven, CT, USA; 4) Department of Pediatrics, Faculty of Medicine, University of the Ryukyus, Nishihara, Japan; 5) Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan.

Among 645 Japanese subjects with intellectual disability (ID) and multiple congenital anomalies of unknown etiology screened by multiple microarray platforms, we detected a male patient with a total length of 40 Mb of runs of homozygosity (ROH) regions using SNP array. The proband is a familial case of joint hypermobility, where the father, an aunt and great grandfather from the paternal side were also affected. Additionally, he presented a few features resembling Aarskog-Scott syndrome (OMIM #305400) such as ID, short stature, brachydactyly, shawl scrotum, and hyperactivity. However, an investigation of *FGD1* had been previously performed, with negative results. In order to investigate recessive mutations harbored in the ROH regions, whole-exome sequencing (WES) was performed in the parent-child trio. Although no candidate gene lying in the ROH regions could be identified, two genes elsewhere with missense mutations emerged as the most plausible candidates: a *de novo ACTB* mutation (c.616C>T, R206W) and a paternally inherited *COL13A1* mutation (c.478G>A, G160S), both predicted as probably damaging according to *in-silico* tools. Gain-of-function missense mutations in *ACTB* have been associated with Baraitser-Winter syndrome (OMIM #243310), characterized by ID, hypertelorism, broad nose with large tip and prominent root, arched eyebrows, ptosis, iris or retinal coloboma, and pachygyria. The last three features are the most prominent and characteristic of Baraitser-Winter syndrome; however, none of them are present in our patient. *COL13A1* encodes the alpha chain of one of the nonfibrillar collagens and might be related to the joint hypermobility shown by the affected individuals from the paternal side, although it is unclear whether the mutation is segregating in the family. This case illustrates the various shifts in the diagnostic hypothesis before reaching a more accurate diagnosis after WES.

2622F

LRP4 mutations in bilateral renal agenesis with severe limb deficiencies. *A. Jacquinet[1,4], A. Lee[2], A. Swenerton[3], M.S. Patel[1], S. Langlois[1].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 2) Department of Pathology and Laboratory Medicine,University of British Columbia, Vancouver, Canada; 3) Provincial Medical Genetics Program, BC Children's & BC Women's Hospitals, Vancouver, Canada; 4) Department of Medical Genetics, University of Liège, Liège, Belgium.

   Introduction : Severe limb deficiencies have been reported in association with bilateral renal agenesis in several syndromes. Recently, truncating mutations in *LRP4* have been associated with a lethal autosomal recessive form of Cenani-Lenz syndrome (CLS). The objectives of our study was to delineate the phenotype of the third fetus, and through a search of syndromes, identify a group of disorders with overlapping phenotypes that may be allelic with CLS or due to mutations in the same pathway.   **Results :** A male fetus at 22 weeks' gestation was found to have bilateral renal agenesis by ultrasound. The pregnancy was interrupted and autopsy and skeletal survey revealed shortening of forearms and legs below the knees, radioulnar synostosis, foreshortened triangular shaped tibia and fibula, and oligosyndactyly with disorganized metacarpals, metatarsals and phalanges. Bilateral renal agenesis, absent ureters and small bladder were confirmed. Additional features included downslanting palpebral fissures, high-arched palate, small mouth, low-set and posteriorly rotated ears, nuchal oedema, pulmonary hypoplasia, hypospadias and absent olfactory nerves. An autosomal recessive disorder was suspected, given a prior obstetric history of termination at 18 weeks for unilateral limb anomalies and cleft lip. Based on the aforementioned renal and limb anomalies, the lethal form of CLS was considered. Sequencing of *LRP4* identified a paternally inherited variant (c.919+3G del) and a maternally inherited variant (c.4808T>A ;p.Ile1603Val). Both variants were rare in the general population and bioinformatically predicted to be damaging. We compared the phenotype with the two previously reported fetuses as well as with syndromes and disorders that may present prenatally with genitourinary anomalies plus acromesomelia of upper and lower limbs. We found significant overlap with a subset of cases of acro-renal-uterine-mandibular syndrome presenting with high-arched palate plus or minus micrognathia and microglossia, severe limb anomalies affecting the middle and distal segments, oligosyndactyly, renal agenesis, uterine anomalies and absence of olfactory nerves.   **Conclusion :** Based on the clinical features seen in three lethal cases of CLS, we propose that acro-renal-uterine-mandibular syndrome may be allelic with CLS or due to mutations in the molecular pathway involving *LRP4*.

2623W

Whole exome sequencing identifies moderate *GLMN* variants from sporadic Bean's syndrome. *L. Hu[1], J. Yin[1], K. Wu[1], Z.P. Qin[2], X.Y. Kong[1].* 1) Institute of Health Sciences, Shanghai, Shanghai, China; 2) Linyi Cancer Hospital, Linyi, Shandong.

   Bean's Syndrome or Blue Rubber Bleb Nevus Syndrome (BRBNS) mostly affects skin and gastrointestinal tract in early childhood. BRBNS is a rare venous malformation associated disorder, which has similar histological features to venous malformation (VM). Therefore, the two disorders might have similar underlying pathogenesis mechanism. Both familial and sporadic VM can be explained by mutations in *TEK*, the etiology of BRBNS is still undetermined so far. In addition, somatic mutations in *PIK3CA* have also been identified as causative factor in *TEK* mutation negative sporadic VMs. We applied whole exome sequencing to one sporadic VM and two sporadic BRBNSs with blood-derived samples as normal control. One of the BRBNSs was performed as parent-offspring-trio. After variant calling, we filtered the variant list against dbSNP138, 1000 Genome, ESP6500 for novel or rare variations and we only focused on protein-altering variations, such as missense, nonsense, frameshift and splicing site mutations. After that we identified one somatic *TEK* (c.2740C>T, p.L914F) from the sporadic VM affect, which has been proven to be causative. In addition, we identified one rare *GLMN* missense and one novel *GLMN* nonsense mutation from the unrelated BRBNS affects, respectively. However, incomplete penetrance was shown in one *GLMN* mutation carrier. Based on the whole exome sequencing, we found that the estimated cellularity of lesion tissues was about 0.3 and both VM and BRBNSs did not suffer from genomic instability. Because both *TEK* and *PIK3CA* positive VM affects have abnormal mTOR signaling, we supposed that abnormal mTOR might also invloved in *GLMN* mutation. In comparison to GLMN wide type, we found mTOR signaling activated much more in GLMN mutants transfected HUVECs and were sensitive to Rapamycin treatment. In addition, the symptoms of the other two independent BRBNSs were relieved, especially the gastrointestinal bleeding after Rapamycin treatment. Therefore, abnormal mTOR signaling could also be the underlying causative mechanism of BRBNSs.

2624T

SMAD2 mutations associated with DORV, other congenital heart anomalies, seizures, developmental delay and microform holoprosencephaly.
M. Shinawi[1], W.K. Chung[2], D. Pineda-Alvarez[3,6], G. Douglas[3], J.J. Murphy[4], J. Shimony[5], T.L. Toler[1]. 1) Dept Pediatric Gen, Box 8116, Washington Univ, St Louis, MO; 2) Dept of Pediatric & Medicine, Columbia University Medical Center, New York; 3) GeneDx, Gaithersburg, Maryland, USA; 4) Rush University Medical Center, Chicago, Illinois; 5) Dept Radiol, Washington Univ, St Louis, MO; 6) Now at Courtagen, Woburn, MA.

　　BACKGROUND: TGF-β signaling pathway is important for pattern formation and differentiation and is essential for the specification of primary body axes. Mutations in *SMAD2* have been reported in 2 patients from large cohorts of subjects with complex congenital heart disease (CHD) (DORV, dextrocardia with unbalanced complete atrioventricular canal and pulmonary stenosis), 2 patients with holoprosencephaly and more recently were associated with arterial aneurysms and dissections. **METHODS:** Whole exome sequencing (WES) was performed using Agilent Clinical Research Exome kit and Illumina HiSeq 2000 100 bp paired-end reads. Sanger sequencing was used to confirm positive exome findings. **RESULTS:** Patient 1 is a 2-year-old female who presented with a complex CHD (double outlet right ventricle (DORV), D-malposed great vessels, confluent branch pulmonary arteries, ASD, VSD, PDA, mitral and pulmonary atresia) and dysmorphic features. She later exhibited subglottal stenosis, laryngomalacia, staring spells with abnormal EEG, poor weight gain, short stature, and developmental delay. Brain MRI revealed a single central incisor and focused holoprosencephaly testing revealed a variant in *TGIF1* (c.320A>T, p.Gln107Leu) believed to be benign based on frequency in controls and previously reported functional analysis. WES identified a *SMAD2* mutation (c.475G>T, p.Glu159Ter) which was not found in the mother with the father not available for testing. Patient 2 is a 10-year-old female who also presented in first week of life with complex CHD (DORV with overriding aorta, VSD, ASD, PDA). She later exhibited developmental delay, learning disability, short stature and poor weight gain, seizures, some dysmorphic features, and mild-moderate conductive hearing loss. WES identified a *de novo* missense *SMAD2* mutation (c.935G>C, p.Cys312Ser). The mutations in our patients were not observed in 6500 individuals of European and African American ancestry in the NHLBI Exome Sequencing Project. **CONCLUSIONS:** Our data add to the accumulating evidence of the important role of SMAD2 in heart and brain malformations. We also describe the first patient, to our knowledge, with both complex CHD and holoprosencephaly associated with a *SMAD2* mutation, although the possible contribution of the *TGIF1* variant to holoprosencephaly cannot be excluded. Additional patients with *SMAD2* mutations are needed to better characterize the phenotypic and genotypic spectra of SMAD2-related disorders. .

2625F

Whole genome sequencing identifies compound heterozygous mutations in *RNU4ATAC* in a large family characterized by developmental, infectious and inflammatory disease.
Z. Deng[1], A. Almeida de Jesus[1], J. McElwee[2], R. Siegel[1], E. Hanson[1]. 1) NIAMS, National Institute of Health, Bethesda, MD; 2) Merck Research Laboratories, Boston, MA.

　　Introduction: Genome sequencing (WGS and WES) can identify monogenic disease mutations in patients with early onset symptoms suggestive of immune dysregulation. Here we report genetic findings from individuals with cognitive and skeletal defects, recurrent infection, and multisystem autoimmune disease arising in a single family. **Cases**: The three affected males were all born with intrauterine growth retardation to non-consanguineous parents. All have short stature, musculoskeletal abnormalities, developmental delays, dysmorphic features, polyendocrinopathy, combined immunodeficiency, and autoimmunity. Extensive clinical and cellular phenotyping coupled with targeted sequencing failed to yield a diagnosis. **Methods**: WGS was performed on the parents, 3 affected sons, 3 unaffected sons and 2 unaffected sisters. Variants were jointly called with the standard BWA-PICARD-GATK pipeline and the GenomeSTRiP CNV caller. A search for the likely causative mutation was performed using a combination of scripts that we developed and expert review. **Results**: High genome coverage was obtained for each sample (30X to 60X). Initial mutation searches were focused on rare protein coding changes consistent with X-linked, autosomal recessive or de novo inheritance models. No probable candidate mutations were identified. Analysis of CNV data and haplotype analysis of the X chromosome also was negative. Linkage analysis identified 4 regions in the genome with maximum LOD score achievable for the given pedigree. One of the candidate regions on chr2 harbors a non-coding RNA gene, *RNU4ATAC*, that plays an essential role in splicing minor introns found in about 800 human genes recently linked to Roifman Syndrome. The remarkable clinical similarities of our patients to Roifman Syndrome suggested *RNU4ATAC* as a likely candidate gene. We found two rare mutations in this gene that segregated with the disease: c.[13C>T] and c.[116A>C]. The first mutation has been reported as pathogenic in Roifman Syndrome. The second mutation is of unknown significance but it resides right next to a gene region critical to Sm protein-binding. It is also extremely rare (absent in public databases and more than 3700 internal WGS controls). **Conclusions**: Using WGS and mutation analysis, we have found the likely disease-causing mutations in patients with clinical features consistent with Roifman syndrome. Functional studies are currently underway to understand the effect of these mutations on immune cell function. .

2626W

Second-site mutations in a patient with KID syndrome cause reversion of the dominant mutation p.D50N in *GJB2* by inhibition of gap junction channels formation. *S. Gudmundsson[1], M. Wilbe[1], S. Ekvall[1], A. Ameur[1], N. Cahill[1], L.B. Alexandrov[2], M. Virtanen[3], M. Hellström Pigg[3], A. Vahlquist[3], H. Törmä[3], M-L. Bondeson[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Science for Life Laboratory, Uppsala, Sweden; 2) Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM; 3) Department of Medical Sciences, Dermatology, Uppsala University, Sweden.

Revertant mosaicism (RM) is a naturally occurring phenomenon where the pathogenic effect of a germline mutation is corrected by a second somatic event, giving rise to a mosaic appearance. RM has been observed in several inherited conditions, including inherited skin disorders. Here, we report on RM in a Keratitis-ichthyosis-deafness (KID) syndrome patient and further investigate the underlying molecular mechanisms. The patient was diagnosed with KID syndrome at the age of 3 based on characteristic skin lesions, hearing deficiency and keratitis. Investigation of *GJB2*, encoding the gap junction channel (GJC) protein connexin 26 (Cx26), revealed heterozygosity for the recurrent *de novo* germline mutation, c.148G>A, p.D50N.   Intrudingly, at age 20, the patient started to develop spots of healthy-looking skin that grew in size and number on the inside of her thighs, within her widespread erythrokeratodermic lesions. Like 15% of reported KID syndrome patients, she developed squamous cell carcinoma (SCC). Interestingly the SCC occurred at the same location as the RM. To further investigate the mechanism, Single Molecule Real-Time (SMRT) deep sequencing was performed on cDNA and gDNA from two biopsies from healthy-looking spots, generating over 10'000 sequence reads over the *GJB2* locus in all sample.   We identified five somatic nonsynonymous mutations in frequencies between 2.4-12.5%. Three mutations have previously been reported as pathogenic, causing hearing loss, and two are novel variants predicted as disease causing. All five mutations were independently present *in cis* with the p.D50N mutation, but absent in biopsies of lesional skin. No secondary mutations could be detected on the wild-type (wt) Cx26 allele. Functional studies of Cx26 in HeLa cells displayed co-expression of Cx26-D50N and wt Cx26 in GJC plaque. However, Cx26-D50N with second-site mutations displayed no formation of GJC plaque or co-expression with wt Cx26.   Conclusively, we show that the second-site mutations inhibit Cx26-D50N expression in GJCs, which results in expression of wt Cx26 alone and the reversion of the disease-mechanism, explaining reverted skin phenotype in our patient. To our knowledge, this is the first time RM is reported in a KID syndrome patient.

2627T

Re-evaluation of a "negative" clinical exome reveals a novel candidate for cleft palate and hypertelorism. *A. Gutierrez-Uzquiza[1], D. Li[1], E. Bhoj[1,2], M. Harr[2], E. Zackai[2,3], H. Hakonarson[1,2,3,4].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; 4) Divisions of Genetics and Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Whole exome sequencing (WES) has proven to be a powerful tool for diagnosis in Mendelian disorders. A few large studies on the application of WES have reported a diagnostic yield of ≈25-28%, however the difference in yield in research setting has not been fully investigated. Here we present a family with an apparently autosomal dominant form of cleft palate who tested negative on clinical WES. However, when WES data was obtained from clinical lab and re-analyzed to exclude synonymous variants, variants with minor allele frequency (MAF) greater than 0.5%, and variants previously identified in controls by our in-house exome variant database, we identified a single nucleotide substitution, c.2708C>T:p.(S903F), in *PDGFRA* gene as the most likely candidate. The patient is a 4-month-old male with developmental delay, multiple congenital anomalies (cleft palate, posterior ear pits, tethered chord/meningocele with caudal appendage and left clubbed foot and hypertelorism) and family history is significant for mother born with cleft palate and posterior ear pits and a brother with congenital hyperinsulinism. *PDGFRA* (OMIM 173490), Platelet-derived Growth Factor Receptor alpha, plays a major role in palatogenesis, craniofacial and neural crest development. *Pdgfr[-/-]* mice have complete cleft palate (CP) that results from an extracellular matrix defect within the cranial neural crest-derived palatal mesenchyme. *PDGFRA* has been associated with isolated CP, however, our family clearly expands the phenotype to include syndromic CP. This report highlights the challenges in recognizing disease variants and that "negative" exome should be worth a second look.

2628F

De novo pathogenic variants in genes associated with autosomal recessive diseases. *J. Pappas[1,4], D. McKnight[2], M.R. Hegde[3], E. Moran[4], J. Borsuk[1], K. Daley[1], E. Reich[1].* 1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, NY; 2) GeneDx, Gaithersburg, MD; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 4) Center for Children, NYU Hospital for Joint Diseases, New York, NY.

Whole exome and whole genome sequencing in families has demonstrated significant genetic disease burden due to de novo pathogenic variants (Veltman JA et al 2012; Ku CS et al 2012). De novo single nucleotide variants have been associated with sporadic occurrence of autosomal dominant and X-linked diseases (Ku CS et al 2013). There are very few cases of autosomal recessive disease due to de novo pathogenic variants in the medical literature (Kaur J et al 2016). We present two unrelated non-consanguineous families with a child affected by autosomal recessive disease due to an inherited maternal pathogenic variant and a de novo pathogenic variant likely on the paternal allele. Case 1 is a 3 year old female with congenital cataracts, developmental delays and progressive spasticity. Whole exome sequencing of case 1 and her parents revealed compound heterozygosity in the *RAB3GAP2* gene associated with Martsolf syndrome. The p.R385C pathogenic variant was maternally inherited and the p.Q429P was de novo likely on the paternal allele. The mother was of European non-Jewish ancestry and the father of Ashkenazi Jewish and Irish ancestry. The mother's age was 31 and the father's 34 at the delivery of case 1. Clinical evaluations concurred with the diagnosis of Martsolf syndrome. Case 2 is a 15 year old male, known to us since birth, with cataracts, glaucoma, retinal detachment, osteoporosis, multiple fractures, platyspondyly, chest deformity and developmental delays. Whole exome sequencing of case 2 and his parents revealed compound heterozygosity in the *XYLT2* gene associated with SpondyloOcular syndrome. The c.1584delC pathogenic variant was maternally inherited and the c.628+2T>C was de novo likely on the paternal allele. The mother was of northern European (non-Jewish) ancestry and the father of Ashkenazi Jewish ancestry. The mother's age was 30 and the father's 34 at the delivery of case 2. The clinical diagnosis of case 2 was SpondyloOcular syndrome. Both Martsolf syndrome and SpondyloOcular syndrome are rare conditions reported in a handful of cases due to homozygous pathogenic variants/consanguinity (Handley MT et al 2013, Taylan F et al 2016, Munns CF et al 2016). Our cases demonstrate that de novo pathogenic variants can be the mechanism of rare autosomal recessive disease in the offspring of non-consanguineous couples. The de novo pathogenic variant was likely on the paternal allele in both cases. Advanced paternal age was not observed in our cases.

2629W

Mandibulofacial dysostosis with alopecia: Report of an adult case with a novel EDNRA mutation. *P. Ahimaz, A. Iglesias.* Clinical Genetics, Columbia University Medical Center, New York, NY.

Endothelin receptor type A (EDNRA) is involved in the endothelin signaling pathway. Animal models have shown that EDNRA signaling plays a prominent role in mandibular neural crest cell specification. In 2015, Gordon et. al, reported four patients with *de novo* mutations in the EDNRA gene and consistent features of mandibular dysplasia, cleft palate, eyelid anomalies, hearing loss and alopecia. The term 'Mandibulofacial Dysostosis with Alopecia' (MFDA) was coined to describe this new syndrome. Here we report a 21 year old female of Bangladeshi descent with MFDA features and developmental delay. Upon examination, she was found to have tall stature, a receding hairline, sparse eyebrows, absent lower eyelashes, mild malar hypoplasia, square nasal tip, mandibular dysplasia, cleft palate, prominent ears and tapering fingers. Further testing showed that she also has bilateral hearing loss, hypoplastic kidney and acro-osteolysis. A SNP chromosomal microarray was obtained, which was normal. Subsequent whole exome sequencing analysis revealed a novel heterozygous missense p.S168L (c.503C>T) variant in coding exon 2 of the EDNRA gene. This nucleotide substitution is in a highly conserved region and is predicted to be pathogenic. Based on the clinical and molecular findings, our patient represents the fifth reported case of MFDA and the S168L variant is the third EDNRA alteration documented to cause this rare condition. In addition to the central MFDA features of mandibular dysplasia, general hypotrichosis, hearing loss and cleft palate, our patient also had atypical findings of renal malformation and acro-osteolysis, which have not been seen with this condition previously. The novelty of the S168L variant along with the clinical features seen in our patient demonstrate possible genotype-phenotype variability associated with this newly described disorder, which has not been well explored owing to its rare incidence. Further cases will help to elucidate the genotype-phenotype correlation and expand the phenotype.

2630T

Isolated optic nerve hypoplasia in 5 family trios: A clinical and exome study. *P. Bitoun[1,2], A. Boland-Auge[3], D. Bacq-Daian[3], E. Pipiras[4], B. Benzacken[4], A. Alexander[5], S. Kuzbari[6], V. Renault[3], J.F. Deleuze[3].* 1) Groupe Medical Jarente, Génétique Médicale, Paris, France; 2) SIDVA91, Juvisy s/orge 91; 3) Centre National de Génotypage CEA, Evry 91; 4) Laboratoire d'Embryo Cyto Génétique , APHP CHU Jean Verdier, Bondy 93; 5) GeneUS, Pearlgen, Austin TX, USA; 6) Banque d'ADN, Hôpital Robert Debré, Paris.

   Purpose: Optic nerve hypoplasia (ONH) is a rare congenital malformation with severe visual incapacity.Children are often only temporarily blind at birth and progress to severe visual impairment.ONH is often syndromic (septo-optic dysplasia / de Morsier syndrome) with MRI pituitary anomalies. Our purpose is to explore currently unknown genetic causes of isolated ONH. **Material and Methods:** 5 Probands (2 females/3 males) with normal MRI and parents were enrolled in genetic study after informed consent. Pediatric, dysmorphologic and ophthalmologic examinations and endocrine work up was performed. Patients had SNP array CGH and Exome Sequencing.Male patients were genotyped for HESX. DNA was extracted from lymphoblastoid cell lines or whole blood and with stringent quality control.DNA was sequenced at a mean of 30X in > 92% of samples using an Illumina platform; Bam file coherence and duplicate elimination used Sam Tools and Picard Tools, QC used house tools and Bedtools, variant call using GATK, annotation using sfnpEff and snpSift selecting variants with severe or moderate impact, variant Filters with de novo dominant hypothesis, absence in 1k, 65k genomes and Exac databases using Gene US from Pearlgen. **Results:** Patients had normal intelligence and distance VA around 20/200 , 3 had learning difficulties, one severe dyspraxia, and one had multiple endocrine deficiencies despite normal MRI and was treated micropenis and cortisol deficiency. 3 patients had sleeping disorders treated with melatonin. No HESX mutations were identified. A single 410,4 kb deletion of the 2p26.1 was identified on SNP array CGH. All parents were healthy and 3 fathers were unavailable. Candidate genes identified by NGS exome analysis needing further Sanger validation will be presented including analysis of 46 variants in 6 genes present in at least 2 families . **Discussion:** A 2p16.1 deletion was identified on array CGH. Whole Exome NGS analysis results in 5 family trios yielded 6 novel gene candidates for optic nerve hypoplasia.

2631F

Werner´s syndrome: First case report in Colombia. *A. Rincon Bolivar[1], J.C Prieto[1,2], J.A Rojas[1].* 1) Instituto de Genetica Humana, Universidad Javeriana, Bogota, Colombia, Bogota, Bogota, Colombia; 2) Hospital la Victoria SDS, Bogota, Colombia.

   Introduction: Werner´s syndrome is an autosomal recessive genetic disease, which is mainly characterized by scleroderma-like skin alterations, juvenile cataracts and signs of premature aging. Additionally these patients have other premature complications associated with age such as atherosclerosis, diabetes mellitus, hyperlipidemia, osteoporosis and high incidence of some types of cancer. **Case presentation:** We report a case of a 48-year-old male patient, who attends our medical genetics consult because he had noticed his voice weakening with a high pitched since about the age of 35 years, associated with premature graying and skin lesions. Patient has personal history of parental consanguinity and diabetes mellitus type 2 in treatment and family history of leukemia and melanoma. Physical examination on admission revealed the main findings of Werner´s syndrome, he appeared much older than his stated age, presents sclerotic skin lesions predominantly in the feet, "bird-like facial appearance", beak-shaped nose, high-pitched voice, gray hair, bilateral cataracts, short stature and hypogonadism. He has a brother also with characteristic clinical findings of the disease but with a more severe phenotype. WRN gene sequencing identifies the homozygous variant NM_00553.4: c.2581C> T (NP_000544.2: pGln861Ter), this generates a stop codon at position 861 and has been classified as pathogenic. This truncated protein affects the helicase domain making it unable to enter the cell nucleus. Werner protein plays a critical role in maintain the structure, integrity and repairing DNA. Patient monitoring continues without reporting other associated complications such as osteoporosis and cancer. **Discussion:** In conclusion this is the first reported case of Werner´s syndrome in the Colombian population, in whom clinical phenotype is similar to previously reported in other populations. We report this case because of the low frequency of this disease to avoid misdiagnosis and allow taking preventive measures in these patients to promptly identify complications associated with age, especially malignancies.

2632W

National Institute on Aging Cell Repository. *F. Rahimov, A. Green.* Coriell Institute for Medical Research, Camden, NJ.

The NIA Aging Cell Repository is a resource facilitating cellular and molecular research studies on the mechanisms of aging and the associated degenerative processes. The Repository was established in 1974 and the cells in this resource have been collected using strict diagnostic criteria and banked under the highest quality standards of cell culture. The collections of the Repository include highly-characterized, viable, and contaminant-free cell cultures for research on such diseases as Alzheimer disease, Progeria, Parkinsonism, Werner syndrome, and Cockayne syndrome as well as Gerontology Research Center samples from the Baltimore Longitudinal Study of Aging, the Longevity Collection including the Aged Sib Pair Collection, the Adolescent Study, Apparently Healthy Controls, Animal Models of Aging and both human and animal differentiated cell types. Cell lines and DNA may be used by academic and commercial entities. Biospecimens requested by academic/government/non-profit investigators are provided at no cost to the user. In addition, the Repository utilizes secure laboratory information management systems to monitor inventory and distribution of samples. The full catalog can be accessed at https://catalog.coriell.org/1/NIA.

2633T

Understanding the genetic basis of very early onset inflammatory bowel disease (vIBD). *J. You[1,2], N. Sobreira[2,3], D. Valle[2,3], A. Guerrerio[4].* 1) Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 3) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 4) Pediatric Gastroenterology and Nutrition, Johns Hopkins Children's Center, Baltimore, MD 21205, USA.

vIBD has an age of onset of < 6 years of age and a severe course. These features suggest a strong genetic contribution to the etiology of vIBD. As part of the Baylor Hopkins Mendelian Genomics Center, we are performing whole exome sequencing (WES) on 58 vIBD probands (38 probands are from multiplex families [4 likely autosomal dominant and 32 likely autosomal recessive] and 20 are isolated cases). Currently, we have completed WES on 27 probands and analyzed them by selecting rare (MAF <1%) functional variants (missense, nonsense, frameshift and splicing). We identified 5 heterozygous missense variants in genes associated with disorders already known to be associated with vIBD such as *NOD2* (3 probands), *PLCG2* (1 proband), *GUCY2C* (1 proband). One variant, *NOD2*-Q574X, is not present in ExAC, HGMD and ClinVar; another, *NOD*-R311W, has a MAF of 0.05% in ExAC, is classified as variant of uncertain significant in HGMD and not present in ClinVar; and a third, *NOD2*-E778K, has a MAF of 0.02% in ExAC and is not present in HGMD and ClinVar. The *PLCG2*- Y482H variant has a MAF of 0.3% in ExAC and is not present in HGMD and ClinVar. The *GUCY2C*-N689K variant has a MAF of 0.4% and is not present in HGMD and ClinVar. Segregation analyses of these variants in available family members showed imperfect segregation suggesting incomplete penetrance. Our analysis also identified a novel candidate gene, *ZNF275* (p.V12I), in one family. *ZNF275* is an X-linked gene previously associated with immunodeficiency by a GWAS. This variant was also present in the affected brother and unaffected mother and was not present in the unaffected father and sister. We also investigated genes that were mutated in two or more probands and identified *TDRKH* as a novel candidate gene. Two unrelated probands are compound heterozygous for, p.A359V and p.A359S in *TDRKH*. Both have a MAF of 0.5% in ExAC and neither are present in HGMD and ClinVar. We plan to perform additional analyses including a polygenic analysis, in which we investigate the same sets of genes mutated in two or more families; and a pathway analysis, searching for involvement of the same pathway in multiple affected families.

2634F

Whole genome sequencing reveals a novel *FOXP3* frame-shift mutation responsible for fetal demise in a large multigenerational pedigree of recurrent male intrauterine fetal death. *D.J. Tester[1], N.C. Ackerman[1], F.S. Cowchock[2], M.J. Ackerman[1].* 1) Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 2) Center for Aging and Human Development, Duke University Medical School, Durham, NC.

Intrauterine fetal death (IUFD) or stillbirth occurs in approximately one out of every 160 pregnancies and accounts for 50% of all perinatal deaths. Post-mortem evaluation often fails to elucidate an underlying cause in many IUFD cases. Approximately 25% of IUFD have been attributed to genetic etiologies. Here, whole genome sequencing (WGS) was used to identify the underlying genetic cause for recurrent second trimester IUFD of males in a multigenerational pedigree. There were a total of 12 male cases of IUFD. Most deaths occurred at ≤ 20 weeks gestation. Hydrops fetalis was diagnosed at death in the most recent case. A fetal whole genome chromosome microarray analysis was normal. WGS was performed on DNA from 5 healthy obligatory carrier females and an unaffected male offspring of a carrier. Following variant filtering based on a recessive X-linked inheritance pattern, a rare FOXP3 frame-shift mutation (p.D303fs) that results in a premature truncation of the protein was discovered. Sanger sequencing confirmed the mutation in the affected fetus. The *FOXP3* gene encodes for a transcriptional regulator critical to the function of T-regulatory cells. FOXP3 mutations are associated with immune dysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome which predominantly affects males and may present with a potentially life-threating complex autoimmune disorder in early childhood. Recent case reports have described FOXP3 mutations in three additional families with recurrent IUFD in males. Hydrops fetalis has been diagnosed in affected fetuses due to severe autoimmune hemolytic anemias. Here we demonstrate the utility of WGS-based pedigree analysis to identify the genetic cause for recurrent IUFD when chromosome studies, including microarray analysis, are normal. FOXP3 mutational analysis should be considered in cases of male IUFD. Further studies are needed to determine the prevalence of FOXP3 mutations in IUFD.

2635W

Multidisciplinary Mendelian gene discovery and clinical translation. *N.Y. Frank[1,2], N. Stitziel[1,3], S. Chopra[1], N. Carmichael[1], E. Hoffman[1], E. Krieg[1], J. Krier[1,2], S.R Sunyaev[1,2], R.L. Maas[1], D. Vuzman[1,2], Brigham Genomic Medicine.* 1) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Broad Institute of MIT in Harvard, Cambridge, MA; 3) Cardiovascular Division, Washington University School of Medicine, St. Louis, MO.

Despite major progress in defining the genetic basis of diverse Mendelian diseases, the molecular etiology of many cases remains unknown. Brigham Genomic Medicine (BGM) is an ongoing collaboration between clinical geneticists, computational biologists, and basic researchers aimed at identification of novel genes responsible for distinct familial disease phenotypes encountered in clinical practice. BGM has developed a unique genomic analysis pipeline which won the international 2012 CLARITY Challenge and has now been successfully used to solve undiagnosed cases, even when prior WES/WGS analyses have been unrevealing. Here we present three novel clinical cases where WGS followed by experimental validation led to successful genetic diagnosis with important implications for patient care. In one case, a family was referred to the BGM for a multigenerational history of aortic aneurysms and dissections. WGS in two first cousins with aortic dissection revealed a missense mutation in the lysyl oxidase (LOX) gene. In a second case, WGS revealed mutation in the WISP3 gene in a 34 year old male with precocious polyarthritis previously mis-attributed to rheumatoid arthritis. Lastly, in a third illustrative case, a missense mutation in the gene encoding complement component 3 (C3) was identified in a family with a novel hereditary syndrome of paroxysmal digital ischemia. In all three cases, these disease gene discoveries not only informed family members about genetic risk, but also led to the development of patient-specific guidelines for clinical management. These cases illustrate how genetic resolution even in long-standing individual cases of previously undiagnosed disease can have unexpected beneficial consequences.

2636T

Case vignettes of clinical and genetic disease discovery at the Stanford Center for Undiagnosed Diseases. *P. Zornio[1,2], J. Kohler[1,2], A. Dries[1,2], D. Waggott[1,2], L. Fernandez[1], E. Ashley[1,2,5], P. Fisher[1,3], J. Bernstein[1,4], M. Wheeler[1,2], Members of the UDN.* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, California; 3) Department of Neurology, Stanford University, Stanford, California; 4) Department of Pediatrics, Stanford University, Stanford, California; 5) Department of Genetics, Stanford University, Stanford, California.

The Stanford Center for Undiagnosed Diseases (SCUD) is designed for patients with rare and undiagnosed diseases when standard clinical options have been exhausted. In collaboration with the National Institutes of Health, the Undiagnosed Diseases Network is comprised of 7 Clinical Sites (including the Stanford Center for Undiagnosed Diseases), 2 Sequencing Cores, one Model Organism Core, one Metabolomics Core, one Central Biorepository, and one network-wide Coordinating Center. The objectives are to 1) Provide answers to patients with mysterious conditions that have long eluded diagnosis, and 2) Advance medical knowledge about rare and common diseases. Since our launch in September, 2015, the SCUD has over two dozen active cases, with approximately 1/3 of patients receiving a clinical and/or molecular diagnosis. Here we present several vignettes of our first cases to demonstrate the novel approaches to diagnosis, in particular relating to novel genetic disease discovery.

2637F

CRISPR/Cas9 as a therapeutic tool for the germline correction of the BBS1 M390R mutation. *M. Cring, C. Searby, C. Carter, V. Sheffield.* University of Iowa, Iowa City, IA.

Bardet-Biedl Syndrome (BBS [MIM 209900]) is a pleiotropic disorder that results in a variety of phenotypes in humans and animal models including blindness, male infertility, polydactyly, intellectual disability, hydrocephalus, and obesity. BBS is a model ciliopathy, which affects approximately 1 in 1000 people worldwide. No current efficient treatments for ciliopathies exist, illustrating a critical need for the development of new therapeutic strategies. The most common cause of BBS is the M390R mutation in the *BBS1* gene (BBS1 [MIM 209901.0001]), which is implicated in approximately 25% of all molecularly diagnosed BBS cases. BBS1 is an essential component of the BBsome that is required for basal body and ciliary function. The BBS1 M390R mutation leads to the classic aforementioned phenotypes of BBS, including the lack of flagellar formation in spermatogenesis. Immunocytochemical analyses have shown that sperm cells in a mouse model of M390R are completely devoid of flagella. Our objective is to implement gene editing techniques using CRISPR/Cas9 to rescue flagellar formation in spermatocytes *in vitro* and *in vivo* in the *Bbs1*$^{M390R/M390R}$ mouse model. We have produced several CRISPR/Cas9 constructs in saCas9 and spCas9 plasmids targeting near the M390R locus in the mouse *Bbs1* gene, including an allele specific gRNA. Using cell free cleavage assays, we have identified several constructs that efficiently cleave near the M390R codon. On going work is aimed at rescuing flagellar formation *in vitro* by culturing primary testicular cells from *Bbs*$^{M390R/M390R}$ mice. These cells are electroporated with CRISPR/Cas9 constructs and single-stranded oligonucleotide homologous donor templates (ssODNs). The cells will be tested for functional rescue of flagellar formation by immunocytochemistry. On going work is also aimed at functional rescue of *Bbs1*$^{M390R/M390R}$ *in vivo* by injection and electroporation of *Bbs*$^{M390R/M390R}$ mouse testes with the CRISPR/Cas9 constructs and ssODNs. Functional *in vivo* rescue will be determined by successful impregnation of female mice, as well as by histology. By correcting the germline *in vivo*, all downstream progeny should be corrected. This study will provide data on the efficacy of *in vivo* germline correction.

2638W

Penetrance in the EHR record of 76 DiscovEHR Cohort participants with two recurrent pathogenic variants. *K. Manickam[1], D. Hartzel[1], N. Abdul-Hsun[2], F. Dewey[2], D. Lindbuchler[1], M. Barr[1], A. Lazzeri[1], M. Hallquist[1], A. Buchanan[1], W.A. Faucett[1], M. Murray[1]. 1) Precision Health Center, Geisinger Medical System, Forty Fort, PA; 2) Regeneron Genetics Center, Tarrytown, NY.*

Participants in the Geisinger Regeneron DiscovEHR cohort will only receive incidental or secondary genomic findings that are classified by stringent criteria as pathogenic/likely pathogenic (P/LP). Early return of results efforts have focused on two disorders: Familial Hypercholesterolemia (FH; *LDLR, APOB, PCSK9*) and Hereditary Breast and Ovarian Cancer (HBOC; *BRCA1, BRCA2)*. The identification of a P/LP variant is not sufficient to establish a condition diagnosis. We examined de-identified EHR records for evidence of penetrance in carriers of two specific P/LP variants in 76 unrelated participants with either: *BRCA2* 5722_5723delCT and *APOB* c.10580G>A. In total, there were 24 carriers of the *BRCA2* variant and 52 carriers of the *APOB* variant. We used established criteria for diagnosis of these disorders, which are based on signs, symptoms, and family history. In FH, three established diagnostic criteria are used to determine pre-test risk. All three criteria use cholesterol level and family history to assess risk (some use additional phenotypic information). For HBOC risk, criteria use personal history of cancer and family history to determine prior probability of a variant. Among the *APOB* variant carriers, 79% had hypercholesterolemia and this together with the sequence variant met minimal criteria set for FH diagnosis. This was a significantly higher percentage meeting FH diagnosis compared to other *APOB* variant carriers (43%, n=53) in the same cohort. Among the *BRCA2* variant carriers, 13% of females had a personal history of cancer consistent with the diagnosis. Overall the current age of BRCA2 is 59 and 37.5% were male. In analysis of available family history, 10 were seen for a clinical visit and a full pedigree was taken as the gold standard. In this group, only 40% had a family history that would indicate high risk for HBOC. All 76 patients with these variants will be re-contacted and offered in person visits for targeted phenotyping. We anticipate that the penetrance ascertainment will be increased through this clinical engagement effort. EHR based phenotypes in the DiscovEHR cohort for these recurrent P/LP variants in 2 genes on the ACMG 56 list are noted to have reduced penetrance. The return of results efforts in this cohort will allow for a calculation of penetrance in the pre-return EHR record as compared to penetrance in the post-return record.

2639T

***In silico* models for genotype/phenotype correlations of *F9* gene mutations causing severe Hemophilia B.** *L. Meléndez-Aranda[1,2], A.R. Jaloma-Cruz[2], M.M. Romero-Prado[1]. 1) Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Jalisco, México; 2) Centro de Investigación Biomédica de Occidente, Guadalajara, Jalisco, México.*

Introduction and Objective: *In silico* modeling of the amino acid substitutions found in Hemophilia B may identify the structural and functional effects on the mature FIX protein. In this work we apply informatics tools to describe the genotype/phenotype possible correlations of selected *F9* gene mutations at exons 2 and 3 causing severe hemophilia B and their effect on the structure and function of the propeptide and Gla domains of FIX protein. **Materials and methods**: Severe phenotype-associated mutations at exons 2 and 3 of *F9* gene were selected from the current Haemophilia B Mutation Database (http://www.factorix.org/). The PolyPhen program was used to perform multiple sequence alignments using the structural information of FIX protein (Polymorphism Phenotyping, http://genetics.bwh.harvard.edu/pph2/). The theoretical three-dimensional (3D) structure analysis of *F9* missense mutations was performed with the Swiss Model Workspace platform (http: //swissmodel.expasy.org/workspace/). **Results:** The PolyPhen program showed that amino acids in positions V30 and R46 (propeptide) as well as G58 and Y84 (Gla domain) are all highly conserved and have a probability of deleterious effect higher than 0.998 if replaced. The Swiss program revealed that the selected mutations in exon 3 affect the hydrogen bonding network of the Gla domain; in turn, this impairs the orientation of side chains of amino acids and their normal conformational structure (omega loop between amino acids at 3 to 11 positions) for binding to calcium, which is critical for interaction with phospholipids in platelet membranes. The limitations of the 3D structural analysis by crystallography (Swiss Model Workspace) prevented us to assess the effect of amino acid substitutions at the propeptide, which nevertheless has a strong impact on the structure and function of the mature FIX. **Conclusions**: Selected mutations in the propeptide and Gla domains affect highly conserved amino acids and have a significant functional effect on the carboxylation of glutamic acid residues via disruption of the interaction of mature FIX with platelets. The inclusion of the FIX propeptide in the established informatics models would have a strong impact on the structure and function of the FIX protein analysis.

2640F

The Israeli national genetic carrier screening program: Experience and outcome, CF and SMA as examples. *A. Singer[1], I. Grotto[2], P. Stafler[3], C. Vinkler[4].* 1) Department of Community Genetics,, Public Health Services, Ministry of Health, Tel Aviv, Israel; 2) Public Health Services, Ministry of Health, Jerusalem, Israel; 3) Cystic Fibrosis Center, Schneider Children's Medical Center of Israel, Petach Tikva, Israel; 4) Inst Medical Genetic Wolfson Medical Center, Holon, Israel.

The Israeli population is combined of various ethnic groups which are quite conserved. These groups are Jewish (Ashkenazi and non-Ashkenazi) and arabs (Muslem, Christian and Druze). High frequency of genetic diseases is unique to the Israeli population as a result of its religious isolated small communities, and high rate of consanguinity. Founder mutations are known to be present in these communities and were identified over the years (e.g. Tay Sachs). With the possibility to perform a molecular diagnosis of many genetic diseases, many disorders became candidates for screening. The Israeli Society of Medical Geneticists (ISMG) made recommendations to expand the carrier screening program and include all of the severe genetic diseases in which the carrier frequency has 1:60. These recommendations were accepted by the Israeli MOH and today 18 tests are offered free of charge to the general population. The program offers carrier screening for cystic fibrosis, fragile X, and spinal muscular atrophy for the general population and additional tests for severe diseases with high carrier frequency to specific ethnic groups according to the frequency of founder mutation in each group. Methods: Tests are offered by the primary physician as part of preconception recommendations. Pretest information are given by trained professionals. Carriers are referred to genetic counseling for discussion of pre-gestational and prenatal tests options. Fetal diagnostic tests (PGD, CVS/Amnio.) are included in the Israeli national health basket. Results: More than 400,000 tests were done during 2013-2014 for the above 18 traits. Of these 117,291 were CF carrier screening tests. The CF registry proves the impact of this program showing a marked reduction in CF birth rates from 14.5:100,000 live birth to 6:100,000 live birth during the last 20y with a shift towards milder mutations. It should also be mentioned that prenatal diagnosis allows later early intervention improving survival and quality of life of CF patients. Epidemiological data regarding the impact of this program on SMA will be presented.

2641W

Clinical exome reduces the gap between diagnostic and research. *B. Cogné, S. Schmitt, P. Boisseau, S. Küry, X. Latypova, W. Deb, T. Besnard, E. Séché, M. Nizon, M. Vincent, S. Mercier, B. Isidor, S. Bézieau.* Medical genetics unit, CHU Nantes, France.

Background: Whole exome sequencing (WES) has been proved to be an efficient tool in the diagnosis of monogenic diseases. The high diagnostic yield of WES justifies its clinical implementation in defined indications, thereby shortening the diagnostic odyssey. Furthermore, during the diagnostic test, the access to non-OMIM genes offers the possibility to enhance the identification of new disease-causing genes, especially for nonspecific phenotypes. **Method**: We selected 55 patients affected with heterogeneous genetic pathologies without any obvious disease-causing gene or previous unsuccessful targeted genes analyses; WES was performed in the proband only. Data were analyzed following the GATK's best practices and variants were prioritized using an in-house phenotype-guided analysis, using Human Phenotype Ontology terms. A first analysis was focused on rare indels, nonsynonymous SNV and canonical splice sites (+/-2bp) identified in clinically relevant genes (with HPO match); if negative, the analysis was extended to rare synonymous and intronic variants in OMIM and non-OMIM genes. CNVs were analyzed using CANOES and custom scripts were developed to annotate and filter CNVs, linking them to rare indels and SNVs identified in WES, in order to easily detect recessive conditions involving one CNV. **Results**: A diagnosis could be reached in 38% of patients (21/55), including nine autosomal recessive, seven autosomal dominant, three X-linked recessive and two X-linked dominant disorders. The CNV pipeline allowed the detection of a deletion of two exons in COQ2, which, together with a frameshift variant, was responsible for coenzyme Q10 deficiency. For "negative" exomes, we focused on variants absent from databases and predicted as deleterious (CADD phred score > 20), which highlighted three candidates variants in non-OMIM genes: one appeared *de novo* after segregation analyses and the two others had matches in GeneMatcher with patients presenting with a similar phenotype. **Discussion**: Our diagnostic yield is in the range of those already published in WES studies. For negative WES, we expect some of them to be resolved with the identification of new disease-causing genes. It is thus important to implement a local database that can be easily interrogated. Furthermore, the identification of new disease-causing genes could greatly benefit from the data sharing of candidate variants identified during the diagnostic procedure, reducing the gap between diagnosis and research.

2642T

A splicing mutation in *VPS4B* causes Dentin Dysplasia I. *F. Xiong[1], Q. Yang[1], D. Cheng[2], Da. Cheng[1], C. Liu[1], Y. Liu[1,3], Q. Yu[1], J. Xiong[1], J. Liu[2], K. Li[2], L. Zhao[4], Y. Ye[1], L. Hu[1], Z. Tian[5], X. Shang[1], L. Zhang[5], X. Wei[1], W. Zhou[1], D. Li[6], W. Zhang[4], X. Xu[1,7].* 1) Department of Medical Genetics, Southern Medical University, Guangzhou, Guangdong, China; 2) School of Stomatology, Zhengzhou University, Zhengzhou, China; 3) Department of Prenatal Diagnosis Center, Maternal and Child Health Hospital, Dongguan, China; 4) Department of Cell Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; 5) Department of Stomatology, Nanfang Hospital, Southern Medical University, Guangzhou, China; 6) Department of Forensic Science, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; 7) Correspondence author.

Dentin dysplasia I (DDI) is a genetically heterogeneous autosomal-dominant disorder characterized by rootless teeth with abnormal pulpal morphology, the etiology of which presents as genetically heterogeneous.In this study,using a cohort of a large Chinese family with 10 DDI patients, we mapped to a 9.63 Mb candidate region for DDI on chromosome 18q21.2-q21.33. We then identified a mutation IVS7+46C>G which resulted in a novel donor splice site in intron 7 of the *VPS4B* gene with co-segregation of all 10 affected individuals in this family. The aberrant transcripts encompassing a new insert of 45-bp in size were detected in gingival cells from affected individuals. Protein structure prediction showed that a 15-amino acid insertion altered the ATP-binding cassette of VPS4B. The mutation resulted in significantly reduced expression of mRNA and protein and altered subcellular localization of VPS4B, indicating a loss of function of VPS4B. Using human gingival fibroblasts, the *VPS4B* gene was found to act as an upstream transducer linked to Wnt/beta-catenin signaling and regulating odontogenesis. Furthermore, knock-down of *vps4b* in zebrafish recapitulated the reduction of tooth size and absence of teeth similar to the tooth phenotype exhibited in DDI index cases, and the zebrafish mutant phenotype could be partially rescued by wild-type human *VPS4B* mRNA. We also observed that *vps4b* depletion in the zebrafish negatively regulates the expression of some major genes involved in odontogenesis. The study identifies *VPS4B* as a disease-causing gene for DDI, which is one of the important contributors to tooth formation through the Wnt/beta-catenin signaling pathway.

2643F

***GNPTAB* gene mutations in the Lithuanian mucolipidosis II and II/III patients.** *L. Ambrozaityte[1,2], B. Burnyte[1,2], L. Cimbalistiene[1,2], B. Tumiene[1,2], A. Utkus[1,2], I. Kavaliauskiene[1,2], T. Rancelis[1,2], V. Kucinskas[1,2].* 1) Dept. of Human & Medical Genetics, Faculty of Medicine, Vilnius University, Santariskiu 2, LT-08661 Vilnius, Lithuania; 2) Center for Medical Genetics, Vilnius University Hospital Santariskiu Klinikos, Santariskiu 2, LT-08661 Vilnius, Lithuania.

Mucolipidosis (ML) II and ML IIIα/β are allelic autosomal recessive metabolic disorders due to mutations in the *GNPTAB*, leading to a deficiency of the enzyme N-acetylglucosamine-1-phosphotransferase, which causes the lysosomal enzymes to accumulate in plasma. Lysosomal storage results in a global impairment of many lysosome dependent pathways (e.g. autophagy and endocytosis), leading to cellular dysfunction and death. ML II and IIIα/β patients have a complex phenotype, characterized by failure to thrive, developmental delays and manifests as thickened skin, coarse facial features, hypertrophic gingiva, thoracic deformity, kyphosis, clubfeet, and restricted joint movement. The sequence of the *GNPTAB* gene (NM_024312.4) has been identified by Sanger sequencing for six unrelated Lithuanian patients. Sequence variant c.1123C>T (p.Arg375Ter) (CM096593) has been detected to all of them; in a homozygous state in one case (ML II) or compound heterozygous state with the other detected mutations as c.1196C>T (p.Ser399Phe) (CM062765) or 2715+2T>G (p.?) (CS100356) (ML II/III). Data from the population analysis of 98 self-reported healthy unrelated individuals (equal male and female) from the Lithuanian population was used to identify mutations carrier frequencies. Whole exome sequencing has been performed on Applied Biosystems 5500 SOLiD™ system according to the optimized manufacturer's protocols, obtaining average coverage of 20x. In total 16 different variants have been detected in the Lithuanian population samples of the *GNPTAB*, five – exonic. The identified mutations of the Lithuanian ML II and II/III families have not been detected in the population samples. The genotype phenotype analysis of our patients correlate with the published data on ML II and II/III. Early knowledge of the *GNPTAB* genotype should favour gathering clinical information more prospectively and may influence the development of successful therapies.

2644W

Airway response to CFTR-directed therapeutics for cystic fibrosis is modulated by *SLC26A9*. G. He[1], T. Gonska[2,3,4], K. Keenan[4], W. Ip[4], P.Y. Böelle[5,6], F. Lin[1], N. Panjwani[1], J. Gong[1], W. Li[1,7], D. Soave[1,8], B. Xiao[1], E. Tullis[9], H. Rabin[10,11], M. Parkins[10,11], A. Price[12], P. Zuberbuhler[13], H. Corvol[5,14], F. Ratjen[4,15], L. Sun[8,16], C.E. Bear[17,18,19], L.J. Strug[1,7,8], J.M. Rommens[1,20]. 1) Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 2) Department of Pediatrics, University of Toronto, Toronto, Ontario, M5G1E2, Canada; 3) Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 4) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 5) Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Paris 06, Paris 75005, France; 6) Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital St. Antoine, Biostatistics Department; Inserm U1136, Paris 75012, France; 7) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 8) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Ontario, M5T3M7, Canada; 9) Department of Respiratory Medicine and Li Ka Shing Knowledge Institute, St. Michael's Hospital, University of Toronto, Toronto, Ontario, M5B1W8, Canada; 10) Department of Medicine, The University of Calgary, Calgary, Alberta, T2N4N1, Canada; 11) The Department of Microbiology, Immunology and Infectious Disease, The University of Calgary, Calgary, Alberta, T2N4N1, Canada; 12) Division of Paediatric Respirology, Department of Paediatrics, Children's Hospital at London Health Sciences Centre, London, Ontario, N6A5W9, Canada; 13) Department of Pediatrics, University of Alberta, Edmonton, Alberta, T6G1C9, Canada; 14) AP-HP, Hôpital Trousseau, Pediatric Pulmonary Department; Institut National de la Santé et al Recherche Medicale (INSERM) U938, Paris 75012, France; 15) Division of Respiratory Medicine, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 16) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, M5S2Z9,Canada; 17) Program in Molecular Structure and Function, The Hospital for Sick Children, Toronto, Ontario, M5G1X8, Canada; 18) Department of Biochemistry, The University of Toronto, Toronto, Ontario, M5S1A8,Canada; 19) Department of Physiology, The University of Toronto, Toronto, Ontario, M5S1A8, Canada; 20) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S1A8, Canada.

Recent developments in cystic fibrosis (CF) therapeutics has included the strategic targeting of different *CFTR* mutation types with potentiator and corrector drugs. While such targeted approaches hold tremendous promise, there is variability in patient response. *SLC26A9* has been shown to be a modifier of early intestinal and pancreatic CF disease phenotypes. Recognizing that SLC26A9 is itself a solute transporter and is expressed and interacts with CFTR at the cell surface in lung-derived culture models, we sought to better understand its relationship to lung disease severity and to airway therapeutic response in CF. SLC26A9 association with lung disease severity was assessed in individuals from the Canadian and French CF Gene Modifier Consortia. Rs7512462 was found to influence lung function measures when at least one gating deficient CF-causing allele, G551D, was present, but not in individuals homozygous for the common F508del CF-causing allele, p=0.04. In addition, Canadian patients with at least one G551D allele showed differential lung function response to potentiator therapy dependent on their SLC26A9 rs7512462 genotype, p=0.0006. A relationship between SLC26A9 and 'gating deficient' CF-causing alleles in lung function severity and in potentiator therapeutic response, is consistent with SLC26A9 – CFTR interaction at the cell surface. As corrector therapeutics to improve CFTR delivery for 'processing-deficient' alleles are just becoming available, we investigated corrector drug-treated primary bronchiole cells of individuals possessing F508del alleles. Differences in forskolin-activated current responses (CFTR function with corrector treatment versus no treatment) were dependent on SLC26A9 rs7512462 genotype, p=0.02, implying knowledge of rs7512462 genotype may predict patient responses and facilitate improved management of patient care. Our findings suggest that variation in SLC26A9 is associated with airway disease severity provided residual CFTR is present, or can be prompted, to the surface of airway cells. Further, the full realization of personalized medicine in CF should incorporate genetic information beyond CFTR disease gene alleles, including *SLC26A9* variation.

2645T

Gene identification in a family with skin fragility and woolly hair syndrome. A.J.M.H. Verkerk[1], A.J.M. Hoogeboom[2], P.A. Arp[1], P. Jhamai[1], R. Kraaij[1], J.G.J. van Rooij[1], J. Verlouw[1], J. Saris[2], A.M.W. van den Ouweland[2], S. Pasmans[3], S. Koljenovic[4], V. Noordhoek Hegt[4], C. Dean[5], J.J. Hublin[6], J.A. Stoop[4], R. Willemsen[2], A.G. Uitterlinden[1]. 1) Internal Medicine, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 2) Clinical Genetics, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 3) Dermatology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 4) Pathology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 5) Department of Cell and Developmental Biology, University College London, London, United Kingdom; 6) Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Woolly hair syndrome, as defined in the literature, encompasses phenotypes presenting with an abnormal variant of tightly curled and twisted scalp hair and diffuse to progressive hair loss. Woolly hair grows slowly, is brittle and stops growing after a few inches. The woolly hair occurs with or without mild to severe skin anomalies and in some cases cardiomyopathy. Autosomal dominant as well as autosomal recessive forms exist. We present a small Dutch family with two children affected with woolly hair syndrome and mild skin anomalies showing a recessive pattern of inheritance. Homozygosity mapping using Affymetrix SNP 6.0 arrays identified a number of homozygous areas in the DNA of the affected children, with one large 7.6 Mb region on chromosome 1. Subsequent whole exome sequencing (NimbleGen SeqCap EZ V2) was performed on an Illumina HiSeq2000 sequencer. Data analysis identified a novel causative gene for the phenotype in these patients. Sanger sequencing verified carrier status of a splice mutation in the parents and a homozygous mutation in the patients. RT-PCR experiments showed aberrant splicing. Immuno-histochemical studies on skin tissue are ongoing to learn more about the functionality of the involved protein.

2646F

Dominant and recessive disease trait inheritance associated with variation at a single locus: Lessons learned from next-generation sequencing. *J. Posey[1], T. Harel[1], J.A. Rosenfeld[1], Z.H.C. Akdemir[1], M.K. Eldomery[1], P. Liu[1,2], M. Walkiewicz[1,2], F. Xia[1,2], D. Muzny[2,3], C.M. Eng[1,2], C.A. Shaw[1,2], E. Boerwinkle[1,3,4], R.A. Gibbs[1,2,3], S.E. Plon[1,3,5,6,7,8], Y. Yang[1,2], J.R. Lupski[1,3,5,6].* 1) Human & Molec Genetics, Baylor College Med, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Human Genetics Center, The University of Texas Health Science Center, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Department of Pediatrics, Texas Children's Hospital, Houston, TX; 7) Texas Children's Cancer Center, Houston, TX; 8) Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX.

The approach to Mendelian genetics often assumes a consistent relationship between mode of inheritance for a given disease and its associated gene. Despite this, several genes have been identified for which both monoallelic and biallelic variants lead to disease. The resulting phenotypes can be distinct (e.g. *BRCA2*) or share substantial overlap (e.g. *COL7A1*) despite differences in severity or phenotypic spectrum. Independent of disease mechanism, genotype-phenotype correlations may be associated with localization to specific protein domains (e.g. *NALCN, EGR2*), type or severity of variants (e.g. *APC*), and escape or trigger of nonsense-mediated decay (e.g. *HBB, RHO*). We hypothesize that molecular diagnoses may be overlooked when identified variants do not observe the expected mode of inheritance. We have identified several cases of dominant and recessive inheritance at a single locus, involving known and novel genes (e.g. *EMC1*). To investigate this systematically, we analyzed the first 345 cases referred to a diagnostic laboratory for trio whole exome sequencing (WES) to identify cases for which a heterozygous *de novo* variant was identified in a known disease gene described as causing a disease trait with autosomal recessive inheritance. This analysis revealed 7 cases (2.0%) with *de novo* variants in known autosomal disease genes associated with disease phenotypes that overlap with the reported phenotype; no second variant or deletion was identified in the exome data. Six cases involved missense variants (*PNPLA6, PIP5K1C, AGRN, SEC24D, ZBTB24, RBBP8*) and 1 involved a splice site variant at the -4 position (*UQCRB*). For 2 of these cases with *de novo* variants in genes associated with Boucher-Neuhauser syndrome (*PNPLA6*, MIM #215470) and Lethal congenital contractural syndrome 3 (*PIP5K1C*, MIM #611369), the variant was a novel missense variant that occurred in a conserved region of the protein; *in silico* bioinformatic analyses suggested a potentially damaging effect on protein function, possibly resulting in either an antimorphic or neomorphic allele. Although at present one cannot rule out a second regulatory variant undetected by WES, a single variant may be sufficient to cause a phenotype at some loci associated with recessive traits. Further investigation of both dominant and recessive disease alleles at a single locus will provide insights into disease mechanisms and genotype-phenotype correlations; moreover it may enhance molecular diagnostic rates.

2647W

Persistence of positive renal and cardiac effects of migalastat in patients with Fabry disease with amenable mutations following 30 months of treatment in the ATTRACT study. *D.G. Bichet[1], D.P. Germain[2], R. Giugliani[3], D. Hughes[4], R. Schiffmann[5], W. Wilcox[6], J. Castelli[7], E.S. Cantor[7], J. Kirk[7], N. Skuban[7], J. Barth[7].* 1) Hôpital du Sacré Coeur, University of Montreal, Montreal, Quebec, Canada; 2) Division of Medical Genetics, University of Versailles–St. Quentin en Yvelines (UVSQ), Montigny, France; 3) HCPA/UFRGS Porto Alegre, Porto Alegre, Brazil; 4) Royal Free Hospital, University College London, London, UK; 5) Baylor Research Institute, Dallas, TX, USA; 6) Department of Human Genetics, Emory University, Atlanta, GA, USA; 7) Amicus Therapeutics, Cranbury, NJ, USA.

Fabry disease is an X-linked disorder of lysosomal α-galactosidase A (α-Gal A) deficiency that causes substrate accumulation and clinical sequelae. Migalastat is an oral pharmacological chaperone that stabilizes amenable mutant forms of α-Gal A in patients with Fabry disease, increasing lysosomal activity. The ATTRACT study (NCT00925301) comprised an 18-month open-label randomized comparison of enzyme replacement therapy (ERT)-experienced patients switching to migalastat HCl 150 mg every other day (n=36) or remaining on ERT (n=21), and a 12-month open-label-extension (OLE) with migalastat. Thirty-one male and female patients with amenable mutations in the migalastat group completed the 18-month randomized period and entered the 12-month OLE. Forty-nine patients with amenable mutations received ≥1 dose of migalastat during the combined 30-month study periods. In patients receiving migalastat for 30 months, the mean (95% confidence interval [CI]) annualized rates of change (mL/min/1.73 m^2) from baseline in eGFR$_{CKD-EPI}$ and mGFR$_{iohexol}$ were −1.7 (−2.7, −0.8) and −2.7 (−4.8, −0.6), respectively. These annualized rates of change in eGFR$_{CKD-EPI}$ and mGFR$_{iohexol}$ are comparable to those previously reported in patients receiving ERT for 18 months: −1.0 (−3.6, 1.6) and −3.2 (−7.8, 1.3), respectively. In patients receiving migalastat for 30 months, the mean (95% CI) change from baseline in left ventricular mass index (LVMi, g/m^2) for all patients and for those with baseline left ventricular hypertrophy (LVH) were −3.8 (−8.9, 1.3) and −10.0 (−16.6, −3.3), respectively. For patients receiving ERT, previously reported 18-month changes in LVMi were −2.0 (−11.0, 7.0) for all patients and 4.5 (−20.9, 29.9) for patients with baseline LVH. LVMi reduction from baseline to month 30 in patients treated with migalastat with baseline LVH was statistically significant. The results indicate that in patients switched from ERT, the renal and cardiac effects of migalastat observed after 18 months persist over 30 months, making migalastat a promising first-in-class oral chaperone treatment for male and female patients with amenable mutations.

2648T

Efficacy of migalastat in a cohort of male patients with the classical form of Fabry disease in the phase 3 FACETS study. *D.P. Germain[1], R. Giugliani[2], D.G. Bichet[3], W. Wilcox[4], D. Hughes[5], H.M. Amartino[6], R. Schiffmann[7], C. Viereck[8], M. Yao[8], N. Skuban[8], J. Castelli[8], J. Barth[8].* 1) Division of Medical Genetics, University of Versailles–St. Quentin en Yvelines (UVSQ), Montigny, France; 2) HCPA/UFRGS Porto Alegre, Porto Alegre, Brazil; 3) Hôpital du Sacré-Coeur, University of Montreal, Montreal, Quebec, Canada; 4) Department of Human Genetics, Emory University, Atlanta, GA, USA; 5) Royal Free Hospital, University College London, London, UK; 6) Hospital Universitario Austral, Buenos Aires, Argentina; 7) Baylor Research Institute, Dallas, TX, USA; 8) Amicus Therapeutics, Cranbury, NJ, USA.

Fabry disease is an X-linked lysosomal storage disorder caused by *GLA* mutations, resulting in deficient lysosomal α-galactosidase A (α-Gal A) activity. Males with the "classical" Fabry disease phenotype generally have an earlier onset of disease and more severe symptoms. Migalastat is an oral pharmacological chaperone that stabilizes mutant α-Gal A to facilitate normal trafficking to lysosomes. Subgroup analyses were undertaken to explore the benefit of migalastat in this cohort of male patients from a 24-month phase 3 study (enzyme replacement therapy-naïve; FACETS). Male patients with classical phenotype were defined as having multiorgan system involvement (based on medical history and baseline assessments) and peripheral blood mononuclear α-Gal A activity <3% (n=14; mean ± standard deviation 42.4 ± 14 years). Changes from baseline (CFB) were calculated for estimated glomerular filtration rate using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula (eGFR$_{CKD-EPI}$), left ventricular mass index (LVMi), plasma globotriaosylsphingosine (lyso-Gb$_3$), and diarrhea symptoms based on the Gastrointestinal Symptoms Rating Scale (GSRS-D; scored on a 7-point Likert-type scale [1=absence of burden to 7=severe discomfort]). For male patients with classical phenotype, the annualized CFB (mean [95% confidence interval]; baseline [n=14]: 87.8 ± 33.6 mL/min/1.73 m^2) to month 24 (n=11) for eGFR$_{CKD-EPI}$ was −0.3 (−2.8, 2.3). For LVMi, the CFB (baseline [n=13]: 114.2 ± 28.5 g/m^2; baseline left ventricular hypertrophy was present in 7/13 patients) to month 24 (n=8) was −12.6 (−25.0, −0.2). For plasma lyso-Gb$_3$, the CFB (baseline [n=7]: 99.8 ± 35.3 nmol/L) to month 12 (n=7) was −36.8 (−69.9, −3.7). For GSRS-D, the CFB (baseline [n=14]: 2.5 ± 1.66) to month 24 (n=10) was −0.9 (−2.1, 0.4). These results demonstrate the clinical benefit of migalastat in male patients with amenable mutations and the classical phenotype of Fabry disease in terms of renal function, LVMi, plasma lyso-Gb$_3$, and diarrhea symptoms. These results are consistent with those seen in the broader study population.

2649F

Podocyte globotriaosylceramide (GL-3) content in male adult patients with Fabry disease is reduced following 6 months of treatment with migalastat. *B. Najafian[1], A. Sokolovskiy[1], J. Barth[2], J. Castelli[2], H. Williams[2], M. Mauer[3].* 1) University of Washington, Seattle, WA, USA; 2) Amicus Therapeutics, Cranbury, NJ, USA; 3) University of Minnesota, Minneapolis, MN, USA.

Fabry disease is an X-linked lysosomal storage disorder caused by *GLA* mutations, resulting in deficient α-galactosidase A activity and accumulation of glycosphingolipids, including globotriaosylceramide (GL-3) and plasma globotriaosylsphingosine (lyso-Gb$_3$). Kidney podocyte injury in Fabry nephropathy is associated with intracellular GL-3 burden, as demonstrated by correlations between proteinuria and volume fraction of podocyte cytoplasm filled with GL-3 and podocyte foot process width, both of which increase with age in young Fabry disease patients. Treatment with the oral pharmacological chaperone migalastat (150 mg QOD) reduces peritubular capillary GL-3 in patients with Fabry disease. This study aimed to establish whether migalastat also reduces podocyte GL-3 in these patients, using a novel approach to estimate the total volume of GL-3 inclusions/podocyte (V[Inc/PC]). Kidney biopsy samples from ERT-naïve male patients with Fabry disease with *GLA* mutations amenable to migalastat (N=8), taken at baseline and again after 6 months of migalastat treatment, were studied by masked unbiased electron microscopy stereology. The mean ± SD V(Inc/PC) of all patients decreased from 2568 ± 1408 μm^3 at baseline to 1282 ± 792 μm^3 after 6 months of migalastat (*P*=0.0182). There was a correlated reduction in mean podocyte volume from 6680 ± 2835 μm^3 at baseline to 3525 ± 2084 μm^3 (*P*=0.004) after 6 months of migalastat (*r*=0.98, *P*=0.00003). These findings indicate that the podocyte cytoplasmic shrinkage was proportional to GL-3 loss; thus, the volume fraction of podocyte cytoplasm attributable to GL-3 did not change significantly. The magnitude of podocyte GL-3 volume reduction following migalastat correlated with improvement of foot process width (*r*=0.82, *P*=0.02), indicative of reduced podocyte injury. Mean plasma lyso-Gb$_3$ also decreased from 118 ± 48 nM at baseline to 75 ± 42 nM after 6 months of migalastat (*P*=0.0004). This decrease correlated with % reduction in GL-3 volume/podocyte (*r*=0.79, *P*=0.02). There was a trend between decrease in GL-3 volume/podocyte and proteinuria (*r*=0.69, *P*=0.06) following treatment with migalastat for 6 months, but no association was found with glomerular filtration rate. In this study, migalastat treatment was associated with a loss of GL-3 inclusions in podocytes in patients with Fabry disease. The sensitive quantitative methods used herein can assess treatment efficacy for this important cell type over a relatively short period of time.

2650W

Migalastat improves gastrointestinal symptoms in patients with Fabry disease: Results from a double-blind, placebo-controlled phase 3 trial (FACETS). *R. Schiffmann[1], D.G. Bichet[2], D. Hughes[3], R. Giugliani[4], W. Wilcox[5], S.P. Shankar[5], D.P. Germain[6], C. Viereck[7], J. Castelli[7], M. Yao[7], N. Skuban[7], J. Barth[7].* 1) Baylor Research Institute, Dallas, TX, USA; 2) Hôpital du Sacré-Coeur, University of Montreal, Montreal, Quebec, Canada; 3) Royal Free Hospital, University College London, London, UK; 4) HCPA/UFRGS Porto Alegre, Porto Alegre, Brazil; 5) Department of Human Genetics, Emory University, Atlanta, GA, USA; 6) Division of Medical Genetics, University of Versailles–St. Quentin en Yvelines (UVSQ), Montigny, France; 7) Amicus Therapeutics, Cranbury, NJ, USA.

Gastrointestinal (GI) symptoms are a prominent manifestation of Fabry disease, with profound and often debilitating negative effects on patients. Migalastat is a first-in-class, oral pharmacological chaperone for patients with Fabry disease with amenable *GLA* mutations. FACETS was a 24-month trial with a 6-month double-blind, placebo-controlled (DBPC) stage, conducted in male and female patients with Fabry disease who were naïve to enzyme replacement therapy. The Gastrointestinal Symptoms Rating Scale (GSRS) was used to assess patient-reported GI symptoms in patients with amenable mutations (n=50). Patients received migalastat HCl 150 mg every other day or placebo. GI symptoms improved with migalastat in 3 of 5 domains (diarrhea, reflux, indigestion) assessed by GSRS. During the DBPC stage, a statistically significant decrease in diarrhea symptoms from baseline to 6 months was seen with migalastat compared with placebo (P=0.03). Based on the minimally clinically important difference for diarrhea, an improvement from baseline of ≥0.4 units was noted: 69% of patients on migalastat had a clinically relevant change vs. 11% of patients on placebo (P=0.012). In patients with symptoms at baseline, a statistically significant improvement in reflux was seen at 6 months compared with baseline (P=0.047). During the open-label extension (months 6-24), a statistically significant improvement from baseline was also observed for indigestion (P<0.05), and there was a trend toward improvement in constipation. Treatment with migalastat led to statistically significant, clinically relevant improvements in GI symptoms assessed by GSRS in patients with Fabry disease with amenable *GLA* mutations.

2651T

***ASAH1* gene mutations cause acid ceramidase deficiency, which has a broad phenotypic spectrum including two distinct diseases: Farber disease and spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME).** *E.H. Schuchman[1], A. Solyom[2], X. He[1], C. Simonaro[1].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mt. Sinai, New York, USA; 2) Plexcera Therapeutics LLC, New York, USA.

Mutations in the *ASAH1* gene lead to acid ceramidase deficiency, accumulation of the lipid ceramide, and a broad phenotypic spectrum, culminating in two recognized diseases: Farber disease and Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME). The prevalence of both diseases is currently unknown and awareness of them is limited due to their rarity; both are likely underdiagnosed. All known phenotypes are associated with autosomal recessive inheritance. Farber disease represents a broad clinical spectrum presenting from infancy through late childhood, associated with the pro-inflammatory and pro-apoptotic characteristics of ceramide. Information from a growing cohort of over 30 living Farber patients reinforces the validity of the characteristic symptoms: early-onset polyarticular arthritis, subcutaneous nodules and dysphonia. However, it also reveals that there are patients who present with only one or two of these symptoms, and that the spectrum of disease includes remarkably attenuated forms with relatively little associated disability. Slowly progressive disease may only be diagnosed in adulthood. Cases of Farber patients diagnosed at over 40 years of age have recently been published, and indicate that peripheral osteolysis can also be caused by *ASAH1* mutations. Less in currently known about the natural history of spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME), although it is also proving to be a disease with a heterogeneous presentation. The broad spectrum of phenotypes associated with *ASAH1* mutations indicate that it should be considered for inclusion in targeted diagnostic panels for motor neuron disease, therapy resistant epilepsy, therapy resistant arthritis or contractures, and peripheral osteolysis. There is no specific therapy currently available for Farber disease or SMA-PME. Haematopoietic stem cell transplantation in Farber has shown variable results and carries a severe burden for the patients. Acid ceramidase enzyme therapy is currently under development.

2652F

Nine-year follow-up in a patient receiving migalastat pharmacological chaperone therapy for Fabry disease. *C. Swift[1], R. Schiffmann[1], J. Barth[2], N. Skuban[2], J. Castelli[2].* 1) Baylor Research Institute, Dallas, TX, USA; 2) Amicus Therapeutics, Cranbury, NJ, USA.

Fabry disease is a rare X-linked lysosomal storage disorder associated with altered autonomic function and neuropathic pain. This report will describe the long-term effects of oral migalastat on Fabry disease-associated pain in a patient with an amenable *GLA* mutation. In 2005, a 37-year-old male patient presented with an 8-year history of severe acroparesthesia in his hands and feet and a sensation of stiffness, predominantly in the lower extremities, that markedly affected his quality of life. He was treated with multiple pain medications, including daily oxycodone, tizanidine, and baclofen, and over-the-counter analgesics. The patient was diagnosed with Fabry disease following a kidney biopsy showing typical lysosomal inclusions in podocytes. His white blood cell (WBC) α-galactosidase A (α-Gal A) activity was markedly decreased to 13.4-25% of normal due to an A143T mutation. The patient began receiving migalastat (various doses) shortly after diagnosis and continued for 4.5 years (phase 2 study), followed by migalastat 150 mg every other day for an additional 4.5 years (phase 3 study). No drug-related adverse events were observed. α-Gal A activity increased 5-fold in WBCs and 30-fold in skin. Cardiac and renal function, as well as urine globotriaosylceramide levels, remained stable on migalastat treatment. Paresthesias and sensation of stiffness progressively improved, and the patient was able to decrease and then discontinue all of his pain medications by the 8th year of treatment with migalastat. Long-term migalastat treatment in this patient with Fabry disease was associated with increased α-Gal A activity in WBCs and skin and reversed symptoms of small-fiber neuropathy over the course of treatment. In this patient case, long-term treatment with migalastat ameliorated Fabry disease-associated pain, resulting in discontinuation of daily pain medications and enabling the patient to resume normal activities of daily living that were not possible prior to treatment.

2653W

A phenotype data sharing format for diagnosis and discovery. *C. Mungall[1], J. Jacobsen[2], J. Balhoff[3], J. Nguyen-Xuan[1], K. Shefchek[4], D. Keith[4], H. Hochheiser[5], S. Lewis[1], S. Köhler[6], P. Robinson[6], J. McMurry[4], T. Groza[7], D. Smedley[2], Monarch Initiative.* 1) Lawrence Berkeley National Laboratory, Berkeley, CA, USA; 2) Sanger Institute, Hinxton, UK; 3) RTI International, Durham, NC, USA; 4) Oregon Health and Sciences University, Portland, OR, USA; 5) University of Pittsburgh, Pittsburgh, PA, USA; 6) Charité – Universitätsmedizin Berlin, Germany; 7) Garvan Institute of Medical Research, Sydney, Australia.

While great strides have been made in exchange formats for genomic sequence and variation data (e.g. Variant Call Format; VCF), the same is not true for phenotypic features. Similarly, bioinformatics tools have rich models for genomic or phylogenetic data types, but lack a uniform phenotype representation. This is due in part to the diversity of phenotypic descriptions, from clinical observations, through QTLs, and newer high-throughput phenotype measurements. Further, phenotypic features are represented in different formats in a diversity of places: publications, databases, health records, registries, clinical trials, and even social media. In these contexts, biocuration has been pivotal to obtaining a computable representation, but is still deeply challenged by the lack of standardization, accessibility, persistence, and computability. As a result, it is extremely difficult to exchange, aggregate, and operate over phenotypic data in the same way that we do sequence data. To address this problem, we have designed an exchange format standard for flexible, extensible, and expressive representation of a broad range of phenotypes in humans or any other species. The Phenotype eXchange Format (PXF) (www.phenopackets.org) works hand-in-hand with phenotype ontologies such as the Human Phenotype Ontology, but allows for representation of other fields such as quantitative measurements, environments and exposures, together with a means of capturing the complete evidence that goes into an association. The goal is to enable PHI-free open exchange of phenotypic data suitable for computational creation by registries and clinics and outside paywalls for journals during publication of genotype-phenotype studies. This will allow consumption of these phenotypic data for computational use by clinical labs for defining gene panels, for diagnostic pipelines, for rare disease patient matchmaking, and for deposition and aggregation in public knowledge bases such as the Monarch Initiative. Open exchange of phenotypic data will allow innovation of algorithms and sharing of our greater understanding of the correlations between genotype, environmental factors, and phenotypic outcomes in more holistic and translational manner.

2654T

A high throughput TaqMan CFTR mutation genotyping workflow. *T. Harts-horne, J. Lang, E. Swartzman, P. Brzoska, L. Tracy.* Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA.

Cystic fibrosis (CF) is an autosomal recessive disease and is the most common genetic disease of Caucasians, affecting 1 in 3000 newborns and having a carrier frequency of 1 in 25. It a progressive incurable disease that is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which impacts mucin production in lungs and other organs. To aid identifying individuals that carry CFTR mutations, we have developed of a high throughput, low cost CFTR genotyping workflow. TaqMan™ SNP Genotyping Assays were developed to nearly 200 CF-causing mutations that were selected from the CFTR1 and CFTR2 databases based on population frequency and predicted functional activity. The mutations occur primarily in coding and regulatory regions and include single nucleotide base substitutions, triallelic mutations, clustered mutations, small insertion deletions, large deletions, and repetitive sequences. The varied nature of the mutations presented challenges to assay development using a single chemistry and platform; these were overcome using nonstandard assay design strategies and analysis methods. CFTR assays were tested on both 384-well and OpenArray™ (3072-well nanofluidic plate) plates run on a real-time PCR system, using Coriell cell line gDNA samples carrying CFTR mutations for a subset of the mutations, and using synthetic DNA controls for all three genotypes per variant to demonstrate robust allelic discrimination. CFTR assays were also tested with DNA isolated from blood and buccal cell samples, and assay accuracy and concordance studies were performed. We will present our complete sample-to-data analysis workflow for high throughput CFTR mutation detection, and cover the development of 200 qualified TaqMan SNP genotyping assays to key CFTR mutations. Data will be presented for nonstandard assays and data analysis methods used to genotype difficult targets including homopolymer mutations and large deletions. Availability of these CFTR assays will enable tailored studies wherein: (1) targets are selected based on CFTR mutation occurrence in populations of interest; (2) small or large numbers of samples can be run on customizable OpenArray panels of 60, 120 or 180 assays, or on TaqMan Array 384-well plates; and (3) highly accurate and reproducible CFTR genotypes can be acquired from blood and buccal cell samples.

2655F

Embodying the ACMG/AMP standards and guidelines for the interpretation of sequence variants in process and code; considerations and the path forward. *E.A. Worthey[1,2], J. Anderson[1], G. Beard[1,2], C. Birch[1,2], D. Brown[1,2], L. Handley[1], J. Harris[1], J. Kelly[1], S. Newberry[1], M. Schroeder[1,2], F. Shater-Ferdosian[1], A. Uno-Antonison[1], A. Weborg[1], C. White[1], B. Wilk[1], D.P. Bick[1,2], Other members of the HudsonAlpha Clinical Sequencing Lab.* 1) Software Development and Informatics, HudsonAlpha Institute, Huntsville, AL; 2) Clinical Services Laboratory, HudsonAlpha Institute, Huntsville, AL.

The advent of genomic sequencing has altered how molecular changes are identified and is transforming medicine. It has uncovered causal molecular changes that would not otherwise have been found and has increased diagnostic success rates in patients with rare genetic disease. Our ability to accurately predict the impact of molecular changes has dramatically improved over the last decade, but much work still remains to interpret the impact on structure and function of molecular changes even within coding regions. Over the last two years there has been increased focus on the topic of consistency in variant classification methods and practices both amongst individual directors within and between Molecular Diagnostic labs. Studies seeking to evaluate consistency and discrepancy in interpretation have identified significant discord; exceeding 30% in some cases. Such findings have prompted formation of cross institutional collaborations aimed at producing definitive and follow-able guidelines for variant interpretation. One goal of such collaborations is identification of the reasons underpinning disparate calls so that as a community we can minimize their occurrence. In 2015 working groups from ACMG and AMP jointly produced final guidelines for interpretation of variants. One intent was to identify factors that should be prioritized in order to maximize classification accuracy. Another goal was to provide a comprehensive evidence-based approach to variant classifications. These guidelines have subsequently been adopted by many labs and embodied in various clinical analysis processes and interpretation software platforms. In many cases, such as ours, these guidelines have superseded or led to alteration or refinement of existing prioritization and classification schemas. The analysis presented here takes an in depth look at concordance and differences between the prior in house schema and the classifications rendered through encoding of these guidelines. Specifics examples will be shown that highlight areas where the updated guidelines need further refinement. We have learnt a lot from *in silico* embodiment of these guidelines; examples of datasets that can be ingested in order to support translation of ACMG evidence codes will also be highlighted. Finally we examine how close we are to being able to embody these guidelines in code to produce accurate suggested clinical interpretations without the need for significant director alteration during interpretation.

2656W

Using mouse genomic and phenotypic data to predict human genetic disease genes. *J.T. Eppig, S.M. Bello, J.A. Kadin, J.E. Richardson, C.L. Smith, MGI Staff.* The Jackson Laboratory, Bar Harbor, ME.

   Our ability to sequence exomes and genomes provides extraordinary opportunity to identify the genetic causes of hereditary human diseases. However, the large number of genetic variants uncovered from a patient provides serious challenges in identifying the causal gene or genes. The laboratory mouse, as the most studied of mammalian model systems, provides unique insights into the dissection of genetic mechanisms of human disease. Comparative phenotyping and directed gene mutation can aid in identification of candidate gene mutations and development of mouse models that recapitulate specific human genetic mutations.   The Human-Mouse Disease Connection is a translational tool (www.diseasemodel.org) designed for exploring and comparing human and mouse phenotypes and their associations with known human diseases, and to provide rapid access to mouse model resources and supporting references. Searches can be initiated based on human or mouse data using one or more parameters, including genes, genomic locations, or phenotypes or disease terms. Results are displayed initially as a visual color-coded grid comparing phenotypes and diseases associated with human and mouse orthologs; or users can choose to view data in tabular format based on genes or on diseases that correspond to the search parameters. New features provide for flexibility for targeted searching (e.g. UBE3A AND Angelman Syndrome) or broad searching (e.g. Hermansky-Pudlak Syndrome OR Storage Pool Platelet Disease). The ability to visually compare the breadth of human/mouse phenotype similarity provides new evidence for underlying genetic models, highlights potential human candidate genes for disease based on mouse phenotypes, and identifies mouse genes that can be engineered to develop new human disease models.   Supported by NIH grant HG000330.

2657T

Assessing the clinical utility of whole blood transcriptome analysis for rare, undiagnosed disorders. *Z. Zappala[1,2], L. Fresard[2], K.D. Kernohan[4], K.S. Smith[2], E.A. Ashley[1,3], J.D. Merker[2], K.M. Boycott[4], S.B. Montgomery[1,2].* 1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA; 2) Department of Pathology, Stanford University, School of Medicine, Stanford, CA; 3) Department of Cardiovascular Medicine, School of Medicine, Stanford University, Stanford, CA; 4) Children's Hospital of Eastern Ontario, Ontario, Canada.

   Clinical exome and genome sequencing is increasingly being used to identify genetic variation underlying disease in patients suffering from rare, undiagnosed genetic disorders. Exome sequencing provides a cost effective way to identify pathogenic variation in protein-coding regions of the genome, while genome sequencing enables the discovery of variation that is either poorly covered or not well captured by exome sequencing (e.g. non-coding variation and structural variation). Both approaches rely on our ability to recognize pathogenic variation – either as rare, predicted loss-of-function/gain-of-function mutations or as large structural/copy number variation. As a result, the molecular diagnostic yield of clinical sequencing strategies is generally reported to be between 25-40%, leaving the majority of patients undiagnosed. We hypothesized that transcriptome analysis of undiagnosed patients would provide functional evidence for pathogenic variation and improve the molecular diagnostic success rate for a diverse set of rare disorders. To test this hypothesis, we performed whole blood transcriptome analysis on patients suffering from heritable cardiovascular disease, heritable cancer predisposition, and other unexplained pediatric disorders with a suspected genetic etiology. By comparing these patients' transcriptomes with a large set of healthy controls from the Depression, Genes, and Networks cohort (n = 909), we identify rare regulatory phenotypes including extreme gene expression, allele-specific expression, and novel splicing patterns. We then highlight aberrant regulatory phenotypes that affect known disease genes or are involved in processes associated with each patient's individual disease etiology and validate the phenotype with quantitative PCR. By integrating transcriptome analysis with genome and exome sequencing data, we identify rare variation associated with these regulatory events that are compelling for clinical diagnosis. Finally, we assess the utility of whole blood transcriptome analysis and compare it to other tissues that can be easily collected in order to understand the limitations of transcriptome analysis in clinical settings where disease presentation is diverse and biopsy of the most relevant disease tissue can be prohibitively invasive. Ultimately, we demonstrate how these integrated genomics analyses can improve the diagnostic yield for the majority of patients with unexplained genetic disorders.

2658F

Initiative on Rare and Undiagnosed Diseases in Pediatrics (IRUD-P) in Japan. *Y. Matsubara[1], K. Hata[1], T. Kaname[1], K. Kosaki[2], IRUD-P Consortium[3].* 1) Research Institute, National Center for Child Health and Development (Japan), Tokyo, Japan; 2) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 3) IRUD-P, Japan.

The Initiative on Rare and Undiagnosed Diseases (IRUD) is a national consortium designed to help patients and their families suffering from rare and undiagnosed disease conditions in Japan. The project started in July 2015. It is funded by the Japan Agency for Medical Research and Development (AMED). The aims of the project are to make diagnosis on patients with rare and undiagnosed diseases, to construct their genome database with clinical information, and to make banking system of precious specimens. The pediatric version of IRUD (IRUD-P) is coordinated by dual centers, the National Center for Child Health and Development (NCCHD) and Keio University, while the adult version of IRUD is led by the National Center for Neurology and Psychiatrics (NCNP). The IRUD-P assigned 17 regional core clinical centers which evaluate clinical symptoms of patients who are referred from local hospitals and clinics and perform first-line laboratory examinations. The obtained data are carefully examined by experts specialized in rare diseases and selected patients are enrolled into IRUD-P. The IRUD-P mainly accepts patients with undiagnosed diseases, but also provides genomic or genetic analysis for rare diseases in which genetic tests are not readily available in Japan. The consortium has four analytical centers for NGS: Tohoku University, NCCHD, Keio University and Yokohama City University. From July 2015 to April 2016, approximately 700 patients, who passed the first screening in the core clinical centers, were consulted to the IRUD-P centers. Specimens accompanied with medical information (n=2,300) were collected from patients and their families, mainly in trios, and sent to the centers. So far we have analyzed 350 patients and genetically confirmed diagnosis in 102 patients. Four patients were found to have new diseases or mutations in novel genes. For the remaining undiagnosed patients, the IRUD-P consortium is developing data sharing network system based on a phenomizer program. The DNA and blood cell specimens have been stored at the National Center Banking Network in Japan. The IRUD-P aims to promote genomic medicine in Japan and to build international collaboration and data-sharing on rare and undiagnosed diseases.

2659W

The Finnish national collection of balanced translocations and inversions facilitates gene mapping. *T. Varilo[1,3], T. Luukkonen[2,3], M. Pöyhönen[4], R. Salonen[5], K.O.J. Simola[6], K. Aittomäki[4], A. Palotie[1,2,7], J.D. Terwilliger[8,9], J. Ignatius[10].* 1) Dept Molecular Medicine, NPHI, Helsinki, Finland; 2) FIMM Institute for Molecular Medicine Finland, Helsinki, Finland; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Dept. of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 5) Dept. of Medical Genetics, Norio-centre, Rinnekoti Foundation, Helsinki, Finland; 6) Dept. of Pediatrics, Tampere University Hospital, Tampere, Finland; 7) Broad Institute of Harvard and MIT, Cambridge, MA, United States; 8) Columbia Genome Center, Columbia University, New York, NY, United States; 9) New York State Psychiatric Institute, New York, NY, United States; 10) Dept. of Clinical Genetics, Turku University Central Hospital, Turku, Finland.

We have gathered a nationwide collection of all carriers of reciprocal balanced translocations and inversions from every Medical Genetics Department and Clinical Genetics Laboratory in Finland. Our database contains medical records of carriers and their relatives carrying the same rearrangement identical by descent as well as other relatives for comparison. To date, we have systematically assembled 3016 carriers and their relatives, and gathered samples (n=124; DNA, RNA and cells from each). We are in process of collecting more biological samples, and linking our data to the comprehensive national disease registers. Additionally, we collaborate with the International Breakpoint Mapping Consortium (Tommerup N et al. 2015 Cancer Genetics). We utilize this valuable collection as a gene mapping tool. We have searched distinct diseases or traits segregating with a given chromosomal break in the families, and identified their precise molecular location on DNA level. We focus on families, which appear to have a striking correlation of a balanced translocation or an inversion and an abnormal phenotype. In individuals from each of these families, we have performed standard cytogenetic analyses, copy number analyses (genotyping and aCGH), genome-wide paired-end sequencing, and capillary sequencing with an objective to identify specific breakpoints for each translocation and rule out other causative genetic factors. In our first gene-mapping pilot, we identified a potential positional candidate gene for intracranial and aortic aneurysm (Luukkonen T et al. 2012 JMG). We have two other manuscripts in preparation. The first is a familial translocation t(1;12)(q43;q21.1), which truncates a *CHRM3* GENCODE isoform in a family suffering from strokes. The other one is a familial translocation t(2;18), which is associated with dyspraxia. Our results demonstrate the feasibility of genome-wide paired-end sequencing for the characterization of balanced rearrangements and identification of candidate genes in patients with potentially disease-associated chromosome rearrangements. This unique registry will be of benefit to both researchers and clinicians by facilitating diagnostic purposes, genetic counseling, and subsequent follow-up.

2660T

MyGene2: Radically open data sharing to accelerate the pace of gene discovery. *J.X. Chong[1], J.H. Yu[1], Q. Yi[2], E. Zhang[3], D.A. Nickerson[2], T. Groza[3], M.J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) 1.Phenomics Group, Garvan Institute of Medical Research, Sydney, Australia; 4) 1.Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA, USA.

Most rare diseases are Mendelian conditions (MC), which means that mutation(s) in a single gene can cause disease. Despite the increasing pace of gene discovery and utilization of exome/whole genome sequencing (ES/WGS) in clinical settings, ~70% of families who undergo clinical ES/WGS remain without a diagnosis—and therefore are unable to benefit from advances in precision medicine. In some of these families, one or more candidate genes are identified but it is unclear which, if any, are "causal," while in others, there is no efficient way to meaningfully prioritize among the tens to hundreds of genes with segregating variants. Some families in this situation have turned to social media in an attempt to "match" with other families with the same condition or test results. However, this process is highly inefficient, and most families do not have the means or technical expertise to do matching effectively.   To lower barriers to public sharing of health and genetic data with the goal of facilitating case-matching, gene discovery, and genotype-phenotype relationship studies, we developed a website, MyGene2 (http://www.mygene2.org). MyGene2 enables families with MC to publicly share health and genetic information (e.g. VCFs and BAMs) and allows clinicians and researchers to publicly share de-identified gene data on behalf of their patients and/or research participants. Users can search MyGene2 for genes of interest and browse health and candidate gene data contributed by all families who have created profiles. Users who contribute data themselves can network with other users and are notified automatically of matches. Tools to enable phenotype matching, real-time public reporting and crowd-sourced evaluation of putative novel gene discoveries, and direct queries by families of their own sequence data are in development. MyGene2 provides a unique opportunity for radically open data sharing with the goal of accelerating the pace of gene discovery for rare diseases. Its use can empower families to directly engage with clinicians and researchers and explore their own data for "answers," and widespread adoption could transform gene discovery efforts by reducing the timeline to publication from months or years post-discovery to days. We are confident that open sharing of health information linked to candidate genes will not only accelerate the pace of gene discovery and translation into diagnostics and clinical care, it will democratize the process of gene discovery.

2661F

Association between the angiotensinogen (*AGT*) gene M235T polymorphism, the methylenetetrahydrofolate reductase (*MTHFR*) gene C677T polymorphism and preeclampsia in Vietnam. *T. Minh[1], T. Cao Ngoc[2], N. Nguyen Viet[1], H. Nguyen Vu Quoc[2], T. Le Mai Hoang[1], U. Le Thanh Nha[1].* 1) Department of Medical Genetics, Hue University of Medicine and Pharmacy; 2) Department of Obstetrics and Gynecology, Hue University of Medicine and Pharmacy.

Background: The AGT M235T and MTHFR C677T polymorphisms has been considered to associated with preeclampsia (PE), but the results from previous studies were conflicting. The aim of this study was to: (1) determine the frequencies of M235T genotypes of AGT gene and the frequencies of C677T genotypes of MTHFR gene in normotensive and preeclamptic pregnant women, and (2) survey the association between these polymorphisms and PE. **Patients and methods:** 68 preeclamptic pregnant women and 272 normotensive pregnant women were determined M235T genotypes of AGT gene by allele – specific PCR technique, were determined C677T genotypes of MTHFR gene by PCR-RFLP with DNA samples extracted from whole blood. **Results:** The frequencies of 235MM, 235MT and 235TT genotypes in normotensive pregnant women were 1.8%; 15.8% and 82.4%, respectively. These frequencies in preeclamptic pregnant women were 1.5%; 14.7% and 83.8%, respectively. The frequencies of 677CC, 677CT and 677TT genotypes in normotensive pregnant women were 83.1%; 15.1% and 1.8%, respectively. These frequencies in preeclamptic pregnant women were 67.6%; 30.9% and 1.5%, respectively. There was no association between the M235T polymorphism and PE. There was the association between the MTHFR C677T polymorphism and PE (p = 0.0074). **Conclusion:** There was no association between M235T polymorphism and PE, but there was the association between the MTHFR C677T polymorphism and PE.

2662W

Evaluation of antigenotoxic and antioxidant effect of curcumin and alpha lipoic acid in radiated lymphocytes of patients with Ataxia Telangiectasia. *C.E. Monterrubio Ledezma[1,2,3,4], L. Bobadilla-Morales[1,4,6], R. Silva-Cruz[1,4], H.J. Pimentel-Gutiérrez[6], C. Ortega-De la Torre[6], M.A. Ramírez-Herrera[5], J.R. Corona-Rivera[1], A. Corona-Rivera[4,6].* 1) Biología Molecular y Genómica, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Área de Genética y Laboratorio de Citogenética de Clínica Down en la Clínica de Atención Especial, DIF Jalisco, México; 3) Secretaria de Salud Jalisco, México; 4) Laboratorio de Citogenética Genotoxicidad y Biomonitoreo del Instituto Genética Humana Dr. Enrique Corona Rivera, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jalisco, México; 5) Laboratorio de Neurofisiología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jalisco, México; 6) Unidad de Citogenética, Servicio Hematología y Oncología Pediátrica, OPD Nuevo Hospital Civil Dr. Juan I. Menchaca, Guadalajara, Jalisco.

Introduction. Ataxia telangiectasia (AT) is a chromosomal instability syndrome caused by mutations in the *ATM* gene. Is a low incidence disease 1/40,000; characterized clinically by ataxia associated with a progressive neurological damage, immunodeficiency, telangiectasias, predisposition to cancer, and susceptibility to damage by ionizing radiation (IR) mediated by overproduction of oxidant radicals. Oxidative stress is the imbalance to favor a pro-oxidant state with consequent damage to biomolecules in the body. The antioxidant and antigenotoxic effect of curcumin and alpha lipoic acid (ALA) have not been evaluated in AT cells patients, therefore it is interesting to evaluate its effect since there is no treatment for these patients. Previously AT lymphoblastoid lines treated with ALA finding significant levels reduction of proteins involved in cell cycle that are elevated due to oxidative stress but without making techniques such as lipid peroxidation, besides unevaluated DNA damage. **Objective.** To evaluate the effect of ALA and curcumin to reduce oxidative damage caused by IR in lymphocytes from patients with AT measuring the concentration of lipid peroxidation and DNA fragmentation by comet assay. **Material and methods**. We used peripheral blood lymphocytes from 4 patients with AT as a study model. Lymphocytes microcultures during 72 h were exposed to curcumin 25 µmol and / or ALA 80 µmol and subsequently exposed to 1 Gy IR. After harvesting, lipid peroxidation test and comet assay were performed. **Results.** Irradiated lymphocytes showed no significant differences compared with control cells, curcumin showed a higher damage value compared to control and did not decrease the damage caused by radiation, even it associated with increased damage, while ALA had a smaller value and decreased the damage caused by radiation, as well as the ability to decrease the concentration of lipids peroxidation. **Conclusions**. Curcumin had a genotoxic effect, and pro-oxidant, whereas ALA had an antioxidant and antigenotoxic effect on lymphocytes from patients with AT *in vitro*.

2663T

Exome sequencing identifies *ATAD5* as a novel predisposition gene for familial uveal melanoma. *M.H. Abdel-Rahman[1,2], R. Pilarski[2], D. Hedges[2], G. Boru[1], K.M. Sample[1], K. Sadek[1], J.B. Massengill[1], D. Kinnamon[2], P. White[3], B. Kelly[3], S. Yoon[4,5], K. Myung[4,5], F.H. Davidorf[1], C.M. Cebulla[1].* 1) Ophthalmology, The Ohio State University, Columbus, OH; 2) Division of Human Genetics, Department of Internal Medicine and Comprehensive Cancer Center, Columbus, Ohio; 3) Biomedical Genomics Core, Nationwide Children Hospital, Columbus, Ohio; 4) Genome Instability Section, Genetics and Molecular Biology Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland; 5) Center for Genomic Integrity, Institute for Basic Science, Korea.

Objectives: About 12% of uveal melanoma (UM) patients have features suggestive of hereditary cancer predisposition. Germline mutation in *BAP1* explains only a small subset of these patients, suggesting the existence of other candidate genes. Methods: Whole exome sequencing (WES) was carried out on a proband and his mother, both with UM. Tumor tissue on the proband was also sequenced. Validation of the mutation was carried out by direct sequencing. Functional studies of candidate genes and downstream signals were assessed in lymphoblastoid cell lines generated from the proband's peripheral blood leukocytes and from his tumor tissue. Direct sequencing of the coding region of *ATAD5* was carried out on an additional 26 unrelated UM patients. Results: Using WES, germline private deleterious variants were identified in 13 genes in common between the proband and his mother. Four of these (*ATAD5*, *DIDO1*, *PIK3C2B* and *MMS19*) have been previously associated with cancer. No evidence of biallelic inactivation of any of the four genes was observed in the UM tumor tissue from the proband. In view of association of haploinsufficiency in mammalian *Atad5* with genomic instability and tumorigenesis we further focused on this gene. The mutation in *ATAD5* (c. 3517G>A, p. Glu1173Lys) was in the functional ATPase domain of the protein. A decrease in the protein expression of the *ATAD5*, though not significant, was observed in the peripheral blood leukocytes of the proband compared to the controls. Accumulation of PCNA, an established downstream target of ATAD5, was observed in the proband's leukocytes suggesting failure of its unloading. Low level microsatellite instability was observed in the proband's tumor. A high frequency of somatic mutations was observed in the proband's UM compared to the average of the reported mutations in UM. Variants of uncertain significance in *ATAD5* were observed in several UM patients by direct sequencing. Conclusions: Our results suggest that the germline mutation in *ATAD5* is the main driver of UM, and possibly other cancers, in this family. Together with a recent study suggesting a role for germline *ATAD5* mutations in development of ovarian cancer, our results support *ATAD5* as a rare hereditary cancer predisposition gene for UM and other tumors in humans. .

2664F

Personal and family history in patients with high penetrance germline findings through paired tumor/normal sequencing. *J. Everett[1,3], K. Hanson[1], V.M. Raymond[1], M. Jacobs[1], E.M. Stoffel[1], J. Innis[1,2], D. Robinson[3], Y. Wu[3], P. Vats[3], R. Lonigro[3], C. Kumar[3], R. Mody[4], A. Chinnaiyan[3,5].* 1) Dept of Internal Medicine, University of Michigan, Ann Arbor, MI; 2) Dept of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Center for Translational Pathology, University of Michigan, Ann Arbor, MI; 4) Dept of Pediatrics, University of Michigan, Ann Arbor, MI; 5) Dept of Pathology, University of Michigan, Ann Arbor, MI.

   Introduction: The Michigan Oncology Sequencing Project (MI-Oncoseq) uses exome sequencing of tumor-normal pairs to identify therapeutic targets. Pathogenic and likely pathogenic (P/LP) germline variants in high or moderate penetrance cancer genes are disclosed. Increased use of tumor sequencing has raised concerns about unanticipated germline findings. We examined correlation between personal history (PHx), family history (FHx) of cancer, and P/LP variants in high penetrance genes. **Methods:** Adults with refractory tumors and pediatric patients with any cancer are study eligible. Genetic counselor-collected, cancer focused pedigrees were reviewed for patients with P/LP variants in high penetrance cancer genes to determine if genetic testing or clinical criteria were met for the gene/syndrome identified. **Results:** Between August 2011 and March 2016, 691 patients completed sequencing. 27 patients (3.9%) had a P/LP germline variant associated with an autosomal dominant high penetrance syndrome (*AIP, APC, BRCA1, BRCA2, DICER1, FH, MLH1, MSH2, PDGFRB, PTPN11, RB1, RET, SBDS, SMARCA4, SOS1, TP53*). 10/27 (37%) had FHx meeting testing criteria, and in 7 of these 10 cases the variant was known from prior testing. An additional 11/27 (40.7%) had PHx of cancer or other findings associated with the gene identified. 6/27 (22%) did not meet clinical or testing criteria. Reasons for not meeting criteria included syndromes with non-cancer findings not queried at pedigree collection (3) and mosaicism (1). Two patients had true unanticipated P/LP variants: *APC* in a patient with hepatocellular carcinoma, and *RET* in a patient with no PHx/FHx of related cancers. **Conclusion:** Among patients having tumor-normal sequencing, 3.9% had a P/LP germline variant in a gene associated with a high penetrance cancer risk syndrome. 77% of these findings were clinically consistent with PHx/FHx. True unanticipated findings in high penetrance cancer genes were rare in our cohort (2/691; 0.2%) suggesting that concerns about unanticipated findings are unwarranted when PHx/FHx is collected and considered.

2665W

***BRCA* testing in Ovarian tumors initiated by a PAthologist (OPA): A pre-screen for germline testing and therapy choice.** *I.E. Fakkert[1], A.R. Mensenkamp[1], E.M. Leter[2], J.A. de Hullu[3], R.W. Willems[4], G.W. Woldringh[1], M. Simons[4], M. Jongmans[1,5], H. Bulten[4], M.J.L. Ligtenberg[1,4], N. Hoogerbrugge[1].* 1) Department of Human Genetics, Radboud university medical center, Nijmegen, Netherlands; 2) Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; 3) Department of Obstetrics and Gynecology, Radboud university medical center, Nijmegen, The Netherlands; 4) Department of Pathology, Radboud university medical center, Nijmegen, The Netherlands; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands.

   Background: Dutch guidelines advise germline DNA *BRCA1* and *BRCA2* testing to all women with ovarian cancer (OC). Approximately 20% of all OC patients have a *BRCA* mutation in their tumor DNA, of which 75% are germline and 25% somatic mutations. Both may benefit from PARP inhibitor therapy. A tumor DNA *BRCA* test for all patients with newly diagnosed OC initiated by a PAthologist (OPA) may serve as a pre-screen for germline DNA *BRCA* testing and as a predictive test for response to PARP inhibitor therapy. **Methods:** Pathologists from multiple laboratories were invited to submit formalin fixed, paraffin embedded (FFPE) samples of all newly diagnosed epithelial OCs for tumor DNA *BRCA* testing. *BRCA* testing was performed using a combined approach of single molecule molecular inversion probe (smMIP)-based targeted next generation sequencing (NGS) and *BRCA1* Multiplex Ligation-dependent Probe Amplification (MLPA) for the detection of gross deletions and duplications in DNA from these samples. Turnaround time for tumor DNA *BRCA* testing was calculated from dates of request to test result. Test results were shared with patients by the gynecologic oncologist, who referred patients with a positive tumor DNA *BRCA* test for germline testing. **Results:** From October 2015 to April 2016 54 tumor DNA *BRCA* tests were requested for 51 women. *BRCA* testing was not feasible in five of the provided tumor samples, because of low tumor cell percentage (N = 2) or low DNA quantity or quality (N = 3). For three of these *BRCA* testing was performed on a different sample. In women with complete tumor DNA *BRCA* testing, median age at OC diagnosis was 63 (range 29-86). *BRCA* mutations were detected in 11 tumors (22%; 6 *BRCA1* and 5 *BRCA2*), of which eight were high-grade serous cancers, one low-grade serous, one endometrioid and one undifferentiated. Median age of women with tumor DNA *BRCA* mutations was 58 (47-85). Median turnaround time was 14 days (7-29). Germline testing was performed in four women and revealed two germline *BRCA1* mutations. **Conclusion:** Tumor DNA *BRCA* testing initiated by the pathologist was feasible in most tumors and turnaround times were short. Knowledge on tumor *BRCA* mutation status can be used to select patients for germline testing and to aid therapy choices shortly after OC diagnosis. **Future plans:** Uptake of tumor DNA *BRCA* testing will be evaluated with data from a national pathology registration database. Patients and physicians experiences will be evaluated by questionnaires.

2666T

Expanded phenotype in multiple endocrine neoplasia type 1 in *MEN1* **mutation carriers.** *J.M. Gass, K.J. Harris, P.S. Atwal.* Center for Individualized Medicine, Department of Clinical Genomics, Mayo Clinic , Jacksonville , FL.

   Introduction: Multiple endocrine neoplasia type 1 (MEN1) is a hereditary cancer associated with many different endocrine as well as non-endocrine tumors, caused by mutations in the *MEN1* gene. Clinical diagnosis is established in an individual with at least 2 MEN1-associated tumors or an individual with one MEN1-associated tumor and a first degree relative with a clinical diagnosis of MEN1. Primary hyperparathyroidism affects the majority of MEN1 individuals by age 50 years. Additionally, *MEN1* mutations trigger familial isolated hyperparathyroidism. We describe a 76-year-old female who presented to our Genetics Clinic with a family history of primary hyperparathyroidism and the identification of a pathogenic *MEN1* variant. **Methods:** Molecular testing for the known *MEN1* mutation c.1A>G as well as biochemical testing, MRI of the brain and abdomen were all performed using standard methods. **Results:** Molecular testing revealed our patient possessed the MEN1 pathogenic variant previously identified in her two offspring. This particular variant was also reported in a family with isolated hyperparathyroidism. Physical exam revealed red facial papules and onset in her seventies, involving her cheeks, nose and upper lip. Formerly, she was diagnosed with rosacea by a Dermatologist and noted no improvement with treatment. Clinically, these lesions appear to be facial angiofibromas. Glucagon was mildly elevated and pancreatic polypeptide was nearly seven times the upper limit of the normal range. Brain MRI was normal. However, an MRI of her abdomen revealed a 1.5 cm lesion at the tail of the pancreas with normal adrenal glands. The patient underwent spleen sparing distal pancreatectomy and subsequent pathology was consistent with a well differentiated pancreatic neuroendocrine tumor. **Conclusion:** Age-related penetrance and variable expressivity are well documented in families with MEN1. It is thought that nearly all individuals with MEN1 manifest disease by age 40. We present a case of late-onset MEN1 in the absence of the most common feature, primary hyperparathyroidism, but with the presence of a PNET and cutaneous findings. This family expands the phenotype associated with the c.1A>G variant and highlights the importance of providing comprehensive assessment of *MEN1* mutation carriers in families that at first blush may appear to have isolated hyperparathyroidism. .

2667F

Whole exome sequencing identifies *TSC1/TSC2* **biallelic loss as the primary and sufficient driver event for renal angiomyolipoma development.** *K. Giannikou[1], I.A. Malinowska[1], T.J. Pugh[2], R. Yan[1], Y.Y Tseng[2], C. Oh[2], J. Kim[2], M.E. Tyburczy[1], Y. Chekaluk[1], Y. Liu[1], N. Alesi[1], G. Finlay[3], C.L. Wu[4], S. Signoretti[1], A. Sivachenko[2], M. Meyerson[2], G. Getz[2], J. Boehm[2], E. Henske[1], D.J. Kwiatkowski[1], * equal contribution.* 1) Division of Pulmonary and Critical Care Medicine and of Genetics, Brigham and Women's Hospital, Harvard Medical School, 20 Shattuck Street, Boston MA 02115, USA; 2) Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 3) Tufts New England Medical Center, Boston, MA 02111, USA; 4) Massachusetts General Hospital, Boston, MA 02114, USA.

   Renal angiomyolipoma is a kidney tumor of mesenchymal cells in the perivascular epithelioid (PEComa) family that is common in both Tuberous Sclerosis Complex (TSC) and Lymphangioleiomyomatosis (LAM). Multiple, bilateral angiomyolipomas in TSC can contribute to renal dysfunction and large angiomyolipomas can lead to life-threatening hemorrhage. Angiomyolipomas are known to have bi-allelic inactivating mutations in *TSC2* and less commonly in *TSC1*. The frequency and contribution of other somatic genetic events is unknown so far. We performed whole exome analysis in 32 resected tumors (n=30 angiomyolipoma, n=2 LAM) from 15 subjects, including three with TSC. Two germline and 23 somatic inactivating mutations in *TSC2* were identified in 30 of 32 samples (94%), as well as one germline *TSC1* mutation, most of them novel. Twenty of 32 (62%) samples showed copy neutral LOH (CN-LOH) mostly in *TSC2* and less commonly in *TSC1. A*t least 8 different LOH regions were defined by SNP analysis. Nine of 32 (28%) samples had small second hit mutations in *TSC2*. Whole exome sequencing in comparison to normal tissue identified 78 somatic non-synonymous coding region variants in 23 tumors, mainly missense and non-recurrent, all confirmed by other methods. Overall, there was a median of 4 somatic mutations per sample (range 0-12). Singleton missense mutations were identified in three known cancer associated genes *(BAP1, ARHGAP35* and *SPEN)*; their functional significance in angiomyolipoma development is uncertain. Comparison of 16 angiomyolipoma specimens from a TSC subject showed distinct genetic aberrations, indicating that most of the tumors arose independently. However, three of tumors had identical somatic mutations and CN-LOH suggesting that they derived from a common precursor cell followed by dispersion with acquisition of additional unique somatic mutations. Finally, analysis of an abdominal LAM tumor and LAM cell clusters isolated from pleural chylous fluid of a sporadic LAM subject revealed four identical somatic mutations and CN-LOH, indicative of a common clonal origin. In conclusion, our results indicate that *TSC2* and less commonly *TSC1* mutations are the primary and essential genetic driver events in renal angiomyolipoma/LAM development, and suggest that other somatic mutations are rare passenger events and likely do not contribute to tumor development. Angiomyolipomas appear to have one of the lowest somatic mutation rates ever reported for human tumors.

2668W

Unique features of GREM1 induced hereditary mixed polyposis syndrome in four Ashkenazi families. *Y. Goldberg[1], M. Schechter[2], T. Adar[2], E. Goldin[2], R. Beeri[3], E. Levy-lahad[3], N. Sharon[4], H. Baris[4], E. Half[5], I. Lerer[6], T. Peretz[1], S. Liebrerman[3].* 1) Sharett institute of Oncology, Hebrew University-Hadassah Medical Center,, Jerusalem , Israel; 2) Gastoenterology institute, Shaare Zedek Medical Center, Jerusalem, Israel; 3) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; 4) Gastoenterology institute, Rambam Medical Center, Jerusalem, Israel; 5) Genetic institute, Rambam Medical Center, Jerusalem, Israel; 6) Department of Human Genetics and Metabolic Diseases, Hebrew University-Hadassah Medical Center, Jerusalem, Israel.

A 40 kb duplication upstream of the *GREM1* gene has been identified as causing Hereditary Mixed Polyposis Syndrome (HPMS) in isolated Ashkenazi Jewish families, conferring a greatly increased risk of colon polyps and ultimately colon cancer. The paucity of confirmed carrier families leaves many clinical questions in diagnosing and managing HMPS unresolved. We present a collection of four non related families with HPMS who were referred to high risk clinics. The probands have undergone extensive work-up, found to carry the 40kb duplication and did not carry any other pathogenic variant in genes causing polyposis syndromes. The phenotype was extremely variable and unexpected; One of the carriers had Desmoid tumor at age of 48 and metastatic duodenal adenocarcinoma at age 67; A small tubular-adenoma was detected in a 13 year old girl; Another family fulfilled the Amsterdam criteria for Lynch syndrome. The number, pathology and age of onset of the colonic polyps in other carriers were also very variable. To estimate the incidence of HMPS, we looked for other carriers in a wider cohort of 184 Ashkenazi Jews who had colorectal cancer (142) or colonic polyps (42) and identified only one (1/184 = 0.5%) carrier. These data illustrate the phenotypic variability, and the overlap between HMPS and other colorectal and polyposis syndromes. It should help shape the diagnostic criteria, clinical guidelines and recommendations for these families who are at higher risk for gastro-intestinal cancer.

2669T

Whole exome and genome sequencing to identify familial multiple myeloma risk genes. *S. Lipkin[1], V. Joseph[2], C. Snyder[3], C. Vachon[8], N. Camp[5], J. McKay[9], R. Niesvizsky[1], S. Chen-Kiang[1], J. Garber[7], M. Daly[4], H. Yu[1], X. Wei[1], H. Lynch[3], R. Klein[6], C. Dumontet[9], K. Offit[2].* 1) Dept Medicine, Weill Cornell, New York, NY; 2) Memorial Sloan-Kettering Cancer Center, New York, New York; 3) Hereditary Cancer Institute, Creighton University; 4) Broad Institute, Cambridge, MA; 5) Huntsman Cancer Institute, University of Utah; 6) Department of Genetics, Mount Sinai School of Medicine; 7) Dana Farber Cancer Institute, Boston, Massachusetts; 8) Mayo Clinic, Rochester, Minnesota; 9) IARC, Lyon, France.

Important strides in our understanding of the genetic basis of cancer susceptibility over the past three decades have impacted the management of both adult and pediatric malignancies. The incorporation of cancer genetic testing into oncology signifies one of the first applications of precision medicine, since it allowed tailored cancer screening, prevention, and in some cases, therapeutic measures. More recent discoveries of mutations of *TP53, RUNX1, CEBPA, GATA2, KLHDC8B,* and *NPAT* have characterized novel syndromes of hereditary predisposition to hematological malignancies. As part of a program to understand the constitutional risk factors associated with inherited risk for many types of hematological malignancies, we have performed whole exome and genome sequencing on probands with familial and early onset Multiple Myeloma. Here we will describe mutation burden analyses, specific candidates and mechanistic validation of novel risk loci previously not associated with familial predisposition to Multiple Myeloma.

2670F

Genomic analysis of inherited breast cancer among Palestinian women.
S. Lolas-Hamameh[1,2,3], L. Kamal[1], D. Dweik[1], M. Salahat[1], P. Renbaum[3], T. Walsh[4], M.K. Lee[4], S. Gulsuner[4], S. Casadei[4], M.C. King[4], E. Levy-Lahad[2,3], M. Kanaan[1]. 1) Hereditary Research Lab, Bethlehem University, Bethlehem, Palestinian Territory; 2) Faculty of Medicine, Department of Human Genetics, Hebrew University of Jerusalem, Jerusalem, Israel; 3) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem 91031, Israel; 4) Departments of Medicine and Genome Sciences, University of Washington, Seattle, Washington, 98195, USA.

Breast cancer incidence in Palestinian women has historically been low, but is rapidly increasing with increased education and later initiation of child bearing. Perhaps because of its historically low incidence, breast cancer in Palestinian women is strikingly familial. We studied inherited breast cancer predisposition in Palestinian women in a hospital-based cohort of 854 breast cancer patients. The cohort was divided into a mutation discovery series (N=391) of 224 women with positive family history and 167 women diagnosed by age 40 years, and general patient series (N=463). The discovery series was tested by the BROCA panel for 28 known and candidate breast and ovarian cancer genes. Considering only unambiguously damaging mutations (i.e. truncations, complete deletions, splice mutations leading to a mutant message, and experimentally validated loss-of-function missenses), in the discovery cohort we identified 38 different mutations, 17 not previously reported, in 47/391 cases (12.0%). Of the 47 patients with mutations, 47% (22) carried mutations in *BRCA1* or *BRCA2*, 12.8 % (6) in *TP53*, 10.6% (5) in *ATM*, 8.5% (4) in *CHEK2*, and the remaining 22.3% (10) in *BARD1*, *BRIP1*, *CDH1*, *ATR, MRE11A*, *PALB2* or *XRCC2*. One patient was a *BARD1/CHEK2* double heterozygote. The general patient series and 100 Palestinian controls were tested using allele-specific assays for 36 of these mutations, revealing 13/463 (2.8%) carriers among the patients and none among the controls. Overall, 14 mutations, in 6 genes, were recurrent in 37 unrelated patients. All recurrent mutations examined (BRCA2, CHEK2, TP53) occurred on a shared haplotype. Family history was the strongest predictor of harboring a mutation. Carrier frequency was 26% (18/70) among patients with family history and young diagnosis, 11% (17/154) among patients with family history only, and 7.2% (12/167) among patients with young onset only (P = 0.0003). *BRCA1* and *BRCA2* carriers were significantly more likely than carriers of mutations in other genes to have triple negative tumors (P = 0.04). This is the first large study to evaluate the genetics of inherited breast cancer in the Palestinian population. The mutational spectrum is highly heterogeneous, including 12 different genes with 38 different mutations, 45% of them novel. Furthermore, most familial cases did not harbor mutations in any BROCA gene, so these families can serve as a future resource for gene discovery. (Supported by the Breast Cancer Research Foundation).

2671W

Germline mutation burden in hereditary cancer predisposing genes: Pan-cancer analysis of the 51,289 individuals from the DiscovEHR study.
R.P. Metpally, S. Krishnamurthy, D. Smelser, S. Kip, D. Carey, H. Williams, Geisinger-Regeneron DiscovEHR Collaboration. Geisinger Clinic, Danville , PA.

Cancer is a complex disease with familial and environmental influences. Alterations in certain genes have been attributed to cancer risk and are providing molecular insights into tumor biology as may serve as putative drug targets. In this study, we describe the prevalence and spectrum of germline variants among cancer predisposing genes and highlight correlations between germline genotypes with tumor types. **Methods:** Using Geisinger-Regeneron DiscovEHR cohort, we sequenced the whole exomes of 51,289 study participants. Individuals were divided into those with (9,911) and without a cancer diagnosis. We analyzed the DNA sequences of 126 hereditary cancer genes of each individual. Pathogenicity of rare missense variants was predicted through weighted ensemble scoring using various damage prediction tools. **Results:** The top 5 malignancies in our population included skin (3069 individuals), prostate (1561), breast (1396), lung (639) and colon (626) cancer. Cancer susceptibility variants (26%) including Clinvar pathogenic (7.8% of individuals), novel missense pathogenic (9.5%) and nonsense (6.5%) were found in our cancer cohort. Genes with susceptibility variants in at least 40 individuals from our cohort are shown in Table 1. Top variants were distributed among several tumor types including skin, prostate, breast, colon, bladder, lung, thyroid, uterine, lymphoma and leukemia. **Conclusions:** Cancer predisposing germline mutations were found in a substantial number of DiscovEHR participants with cancers. Knowledge of these mutations and integration with somatic information will help to predict cancer susceptibility, and thus be beneficial in counseling patients, cascade screening of their families and directing more personalized and disease-specific care, all of which are important steps towards precision medicine.

Gene	Nonsense	Clinvar Pathogenic	Novel missense pathogenic	Total	Percent	Breast	Colon	Prostate	Skin	Thyroid
TERT	27	95	9	131	1.32	23	9	18	24	6
HNF1A	98	22	6	126	1.27	10	10	17	31	8
TET2	114	0	0	114	1.15	23	3	16	34	10
MSH3	22	0	86	108	1.09	18	7	20	27	7
FANCL	1	99	1	101	1.02	13	5	15	18	4
CDKN1A	3	0	91	94	0.95	10	4	15	21	8
FANCA	10	3	81	94	0.95	19	7	20	15	9
CHEK2	9	62	6	77	0.78	16	10	14	12	2
DNMT3A	44	0	33	77	0.78	20	6	14	13	3
MUTYH	3	60	1	64	0.65	12	2	10	16	4
ERCC2	25	17	15	57	0.58	7	1	11	15	2
FANCM	38	0	19	57	0.58	13	1	10	13	4
ATM	11	34	11	56	0.57	14	1	9	10	1
BLM	7	41	6	54	0.54	8	3	7	15	5
MITF	2	46	3	51	0.51	6	2	2	14	1
BRCA2	7	40	1	48	0.48	12	2	7	9	2
BRCA1	22	20	2	44	0.44	11	1	7	8	1
POLE	3	0	41	44	0.44	8	1	5	11	3
ERCC6	11	7	23	41	0.41	4	2	7	13	1

2672T

PTEN hamartoma tumour syndrome: Two Colombian cases with Cowden Syndrome and Bannayan-Riley-Ruvalcaba Syndrome. *A. Nova[1], J.C Prieto[1,2], J.A Rojas[1].* 1) Bogota D.C, Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, Colombia, Bogota, Bogota, Colombia; 2) Hospital La Victoria SDS, Bogota, Colombia.

PTEN hamartoma tumour syndrome (PHTS) is the term adopted to describe Cowden syndrome (CS), Lhermitte–Duclos disease (LD), Bannayan-Riley-Ruvalcaba syndrome (BRRS), Proteus syndrome (PS), and Proteus-like syndrome. Recently, it also has been identified in a fraction of patients with autism spectrum disorder (ASD) who also display macrocephaly with and without additional developmental phenotypes characteristic of PHTS. For this reason currently this term describes a spectrum of disorders that have been linked to germline mutations in *PTEN* (phosphatase, tensin homologue, deleted on chromosome 10) that encodes a tumor suppressor protein, which mediates cell-cycle arrest and apoptosis. The prevalence of detectable PTEN germline mutations in these syndromes is 80% for CS patients, 83% for LD patients, 60% for BRRS patients, 50% for PLS patients, and 20% for PS patients. CS is a multiple hamartoma syndrome with a high risk of benign and malignant tumors of the thyroid, breast, endometrium and other organs. BRRS is a congenital disorder characterized by macrocephaly, intestinal polyposis, lipomas, and genital lentiginosis. While most cases are inherited in a family for generations, following an autosomal dominant pattern, at least 10% and perhaps as many as 44% of cases are due to a new (*de novo*) mutation. PHTS confers increased risks for specific malignancies, most notably breast, thyroid, renal, and endometrial cancers. Benign tumors that affect a variety of tissues are common and can range from subtle skin papules, requiring no treatment, to devastating vascular anomalies. Because the diagnosis of PHTS is difficult to establish by the variable and often subtle external manifestations, many individuals remain undiagnosed, so the true prevalence is unknown. Here we present two Colombians cases, affected by CS and BRRS syndrome respectively, both are carriers of a mutation in PTEN protein, aimed at avoid misdiagnosis and allow taking preventive measures in these patients to promptly identify malignancies and complications.

2673F

Physical activity is inversely associated with colorectal cancer risk in Lynch syndrome. *M. Pande[1], A.K. Win[2], K. Basen-Engquist[1], S.S. Hardikar[3,4], Y.N. You[1], M.A. Jenkins[2], S.K. Peterson[1], P.M. Lynch[1].* 1) The University of Texas MD Anderson Cancer Center, Houston TX, USA; 2) Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia; 3) Department of Epidemiology, University of Washington, Seattle WA, USA; 4) Division of Public Health Sciences, Fred Hutchison Cancer Research Center, Seattle WA, USA.

Background: The beneficial effect of physical activity on colorectal cancer (CRC) risk is well established, however little is known about the influence of physical activity on risk of CRC for those with a genetic predisposition to CRC such as carriers of germline DNA mismatch repair (MMR) gene mutations (Lynch syndrome). **Methods**: We analyzed data from 1865 MMR mutation carriers recruited into the Colon Cancer Family Registry from the USA, Canada and Australia between 1998 and 2012. Self-reported physical activity was summarized over the person's adult life starting at age 20 years, in metabolic equivalent hours per week (MET-h/week). Exposure was assessed prior to the outcome event (CRC) or censoring (diagnosis of a polyp or non-CRC cancer or questionnaire completion) whichever came earliest. Physical activity was categorized as <3.5, ≥3.5-<8.75, ≥8.75-<17.5, ≥17.5-<35, ≥35 MET-h/week (3.5 = 1 hour; 8.75 = 2.5 hours; 17.5 = 5 hours; 35 = 10 hours of brisk walking). A weighted Cox proportional hazards regression was used to estimate the association between physical activity and CRC, while adjusting for relevant confounders. **Results**: There were 691 CRC events over an observation of 80,625 person-years. Compared with the lowest level of physical activity (<3.5 MET h-week), higher levels of physical activity were associated with a decreased risk for CRC (Table 1), however, the association was not linear. The inverse association was maintained after adjusting for potential confounders, although estimates were attenuated. **Conclusion**: Physical activity was inversely associated with the risk of CRC in Lynch syndrome and may warrant counseling of these high-risk individuals to follow national physical activity guidelines.

No	Clinical indication	Karyotype	Microarray karyotype	Results	P
1	Abnormal ultrasound findings A=18	46,XY,del(5)(p13p15.1).ish 5p15.2(D5S23, D5S721)x2	arr[hg19] 5p14.3p13.2(21,057,088-37,749,526)x1	Cri-du Chat not involved, but *NIPBL* deleted, known to cause Cornelia De Lange syndrome 1	
2	Fetus with cardiac echogenic foci and echogenic bowels. A=22+6	47,XY,+mar.ish(15q11.2q13)(D15S11-,D15Z1-)	arr[hg19](1-22)x2,(XY)x1	Marker had no significant genomic content	-
3	Missed abortion	46,XX,add(6)(p24)	arr[hg19] 3p26.3p24.2(61,891-24,670,223)x3,6p25.p25.1(222,200-5,277,366)x1	Microarray refined the abnormality as dup(3p) and del(6p)	0.026
4	Dandy Walker A=23+1	46,XX	arr[hg19] 4p16.3p16.1(2,931,120-6,494,394)x1	Del 4p results in Weyers acrofacial dysostosis or Ellis-van-Crevald syndrome	0.018
5	Fetus showed symmetrical I UGR A=21+3	46,XX,del(13)(q12q14)	arr[hg19] 13q12.3q14.13(29,116,876-47,290,217)x1	Does not involve *RB1*, but results in development delay and other significant developmental or morphological phenotypes. *BRCA2* also deleted	0.117
6	Twins. Lower fetus with IUGR, CVS, VSD A=19+3	46,XX,der(21)ins(21;?)(q11.2;?)	arr[hg19](1-22,X)x2	Surprisingly normal.	0.068

*adjusted for sex, CFR center, body mass index at age 20, cigarette smoking, alcohol consumption, diabetes.

2674W

Mutational profiles of cancer-relevant genes in Lynch syndrome tumor-igenesis. *N. Porkka[1], S. Valo[1,2], T.T. Nieminen[1], S. Eldfors[3], P. Peltomäki[1].* 1) Department of Medical and Clinical genetics, University of Helsinki, Helsinki, Finland; 2) Department of Biosciences, University of Helsinki, Helsinki, Finland; 3) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

Predisposition to Lynch syndrome (LS) is caused by germline mutations in mismatch repair (MMR) genes, notably *MLH1*, *MSH2*, *MSH6*, and *PMS2*. It is the most common form of hereditary colorectal cancer (CRC) accounting 3 – 5% of all CRCs. Defects in MMR gene function cause accumulation of replication errors in the genome. This study was designed to profile different tumors from MMR gene mutation carriers for somatic alterations, starting from colorectal adenomas and carcinomas. The tumors originated from a nation-wide registry of LS families. Frequency of the CpG island methylator phenotype (CIMP) was analyzed using the methylation-specific multiplex ligation-dependent probe amplification test, which detects the methylation status of eight CIMP marker genes simultaneously. Next-generation sequencing of tumor and matching normal samples was conducted using the Nimblegen Comprehensive Cancer Panel, which covers the coding regions of 578 known cancer related genes. Somatic mutations (missense, nonsense, frameshift, in-frame coding indels and splice site mutations) with VarScan somatic p-value < 0.01 were recorded and categorized based on the allele frequency (< 25% vs. ≥ 25%) and type of mutation (truncating vs. non-truncating). CIMP occurred in 18% of (high-dysplasia) adenomas and 50% of CRCs. All adenomas and CRCs were MMR-deficient by MMR protein expression and microsatellite instability analysis. A great majority of the 578 cancer-associated genes showed somatic mutations in all tumor groups. The average frequency of somatic mutations per tumor was higher in adenomas compared to carcinomas, 1514 (range 73 – 4123) vs. 1103 (range 116 – 2379). Focusing on four known CRC driver genes *APC, BRAF, KRAS,* and *TP53*, the somatic mutation frequencies of each were higher in adenomas than carcinomas (although not statistically significant probably due to limited sample series). With adenomas and carcinomas combined, the average frequency of somatic mutations per tumor was 1245, which was significantly (P = 0.007) higher than that in Lynch syndrome breast carcinomas (671). Analysis of additional types of LS tumors and sporadic counterparts is in progress. The genes examined for somatic mutations represent multiple pathways relevant for cancer initiation and progression. The patterns of somatic alterations we found suggest that several molecular mechanisms, genetic and epigenetic, contribute to the high mutation frequencies characteristic of LS tumors.

2675T

Exome sequencing and analyses in a familial breast cancer cohort to identify novel cancer predisposition variation. *K. Praveen[1], W. Chang[2], M.B. Terry[3], J.H. Chung[1], S. Mukherjee[1], J. Staples[1], V. Arunachalam[1], S. Balasubramanian[1], Q. Huang[4], J. Penn[1], J. Reid[1], J.D. Overton[1], Y. Shen[4], F.E. Dewey[1], C. Gonzaga-Jauregui[1], W.K. Chung[2].* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Departments of Pediatrics and Medicine, Columbia University Medical College, New York, New York; 3) Department of Epidemiology, School of Public Health, Columbia University, New York, New York; 4) 4Departments of Systems Biology and Biomedical Informatics, Columbia University Medical College, New York, New York.

   Introduction: Familial breast cancer (FBC) accounts for ~30% of all breast cancer in the population. Germline variants in the predisposition genes that have been identified to date, such as *BRCA1, BRCA2, CHEK2* etc., collectively account for ~35% of FBC. Hence, a majority of familial cases have an unknown genetic cause and may have variants in novel cancer predisposition genes. **Methods:** We performed whole exome sequencing (WES) on a cohort of 1,331 individuals with a family history of breast cancer. The cohort was composed of 737 Ashkenazi Jewish, 550 Dominican individuals and 44 individuals of unknown ancestry. A total of 562 FBC patients and 669 unaffected relatives belonging to 491 families were sequenced. We analyzed all probands for pathogenic and loss-of-function variants in known breast cancer predisposition genes, and conducted family based analyses when multiple family members (≥3 affected individuals) were available for segregation of variants. We also performed case-control single variant association analysis on the cohort using mixed linear models. These results were then used for pathway-based analyses to look for enrichment of variants in pathways in breast cancer patients. **Results:** A screen for variants in known breast cancer predisposition genes identified known and likely pathogenic variants in 12 of 29 genes tested (31 variants). These variants likely increase breast cancer predisposition in 35 families. Amongst the prioritized candidates from the family based analysis, we identified genes involved in DNA repair, apoptosis and cell division, and innate immunity. There were no significant single variants from association analysis after multiple testing correction. However, pathway analysis revealed an enrichment of variants in metabolic and signaling pathways, such as Notch signaling and PI3K/AKT/mTOR pathways. **Conclusions:** Our family based analysis of FBC revealed novel candidate genes for FBC predisposition that function in pathways important in cancer susceptibility. We did not observe a meaningful overlap of candidate genes between families suggesting that there is considerable genetic heterogeneity in predisposition to FBC in patients who do not harbor causative variants in one of the major known predisposition genes. However, we noted a significant enrichment of variants in pathway based analysis suggesting that while FBC patients in our cohort carry variants in different genes, these genes may function in similar biological processes.

2676F

Frequency of pathogenic germline mutations in *EZH2* in 776 children with neuroblastoma. *D.R. Stewart[1], J.P. Evans[2], T. Pugh[3], R. Patidar[4], Y. Song[4], K. Conkrite[2], K.A. Cole[2], N. Menghani[2], L. Farra[2], S. Sindiri[4], X. Wen[4], S. Zhang[4], G. Lopez[2], J. Guidry Auvil[5], D.S. Gerhard[5], S.J. Chanock[1], J.M. Maris[2], J. Khan[4], J. Wei[4], S.J. Diskin[2], Cancer Genomics Research Laboratory.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) Princess Margaret Cancer Centre, Toronto, Canada; 4) Center for Cancer Research, National Cancer Institute, Bethesda, MD; 5) Office of Cancer Genomics, National Cancer Institute, Bethesda, MD.

   BACKGROUND. Genetic susceptibility to neuroblastoma (NB) is complex, and the contribution of rare germline mutations beyond *ALK* and *PHOX2B* is unknown. Here, we report on our efforts to identify rare, deleterious germline variants that play a role in the initiation of sporadic NB. We initially focused our analysis on known cancer-predisposition genes and genes underlying syndromes associated with NB. **METHODS.** We conducted Complete Genomics whole-genome sequencing (106 high-risk tumor-germline) and Illumina exome sequencing (222 high-risk tumor-germline; 52 overlapping, 54 non-overlapping with whole-genome). 500-genes were sequenced by custom-capture in an independent validation cohort of 500 NB patients (germline-only) and compared with 1001 cancer-free controls sequenced/analyzed by the same platform/pipeline. Rare variants were annotated using the database for non-synonymous functional predictions (dbNSFP). Indels were called by bam2MPG and Platypus in the 222 exome cohort and in 1000 Genomes (1000G). We queried 1000G, Exome Sequencing Project (ESP) and Exome Aggregation Consortium (ExAC) for pathogenic variation. **RESULTS.** In 776 children with NB, there were 4 pathogenic (P) or likely pathogenic (LP) germline *EZH2* variants not observed in 1000G, ESP or ExAC: 1) c.G598A; p.D200N (missense, splice-site) in a male, now deceased, with unfavorable histology, 2) c.1732_1735del; p.T578fs in a male, now deceased, with ganglioneuroblastoma (GNB) and unfavorable histology, 3) c.-1_2delCAT; p.Met1del in a male with GNB and unfavorable histology but still alive and, 4) c.1184delG; p.Gly395fs in a male with NB and favorable histology, still alive. The latter tumor also harbored a somatic 5.9 kb 3-exon deletion in *EZH2*. No truncating *EZH2* variants are reported in ExAC. There were no truncating *EZH2* or HG-MD-reported variants in the 1001 cancer-free controls (P < 0.05). **DISCUSSION.** Germline mutations in *EZH2*, a member of the polycomb repressive complex 2, are associated with Weaver syndrome (WS), an autosomal dominant overgrowth disorder. Case reports of children with WS and/or *EZH2* mutations include leukemia (ALL/AML) and NB. We found 4 germline P/LP *EZH2* variants in 4/776 children with NB (0.5%), including one whose tumor also harbored a somatic multi-exon deletion in *EZH2*. **CONCLUSION.** We establish the frequency of pathogenic germline *EZH2* mutations in sporadic NB. Risk of malignancy in children with pathogenic germline *EZH2* mutations merits further investigation.

2677W

Pancreatic cancer as an inclusion factor for genetic testing of breast and colorectal cancer susceptibility genes. *B.A. Thompson[1,2], E.L. Young[1], M.A. Firpo[1,3], S.J. Mulvihill[1,3], T. Werner[1,4], R. Bell[1], K.R. Smith[5], J. Berger[5], A. Fraser[5], L. Neumeyer[6], D.E. Goldgar[7], W.K. Kohlmann[1], S.V. Tavtigian[1,8].* 1) Huntsman Cancer Institute, Salt Lake City, UT; 2) Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, Australia; 3) Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT; 4) Division of Oncology, Department of Medicine, University of Utah, Salt Lake City, UT; 5) Population Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 6) Department of Surgery and Arizona Cancer Center, University of Arizona, Tucson, AZ; 7) Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT; 8) Department of Oncological Sciences, University of Utah, Salt Lake City, UT.

   Genes associated with hereditary breast and ovarian cancer (HBOC) and colorectal cancer (CRC) susceptibility are known to play a role in pancreatic cancer susceptibility. Pancreatic cancer is still lacking adequate measures for prevention or early stage tumor detection. Germline genetic testing will be most beneficial for at-risk relatives of pancreatic cancer cases with pathogenic variants in established HBOC and CRC genes, but it is unclear what proportion of pancreatic cancer cases harbor pathogenic variants in those genes.In an initial experiment, a set of 66 pancreatic cancer cases diagnosed at the Huntsman Cancer Hospital (HCH), SLC, UT, unselected for age of onset or family history, were sequenced using a custom 34 gene panel including known HBOC and CRC genes. The proportion of cases with a pathogenic gene variant was estimated probabilistically. Cases with an established pathogenic variant were assigned weight=1.0. Cases with a Variant of Uncertain Significance (VUS) in *BRCA1/2* [MIM 113705, 600185] or a mismatch repair gene [MIM 276300] were assigned a weight equal to the VUSs probability of pathogenicity from curated publicly available databases. Carriers of VUS from other genes were assigned a weight from Align-GVGD, CADD, Polyphen-2, and MAPP scores. We found that 8.5% of these cases carried a pathogenic variant.To refine the carrier frequency estimate for unselected pancreatic cancer cases, a meta-analysis was conducted using the initial set, a second set from the HCH (n=95), the Cancer Genome Atlas (TCGA) cohort (n=154), and a published dataset from a similar screen of unselected patients (n=96). These were compared to the non-TCGA Exome Aggregation Consortium (ExAC) dataset, using Standardized Incidence Ratios (SIR).Overall, 11.9% of unselected pancreatic cancer cases carried a variant (including weighted VUS) in known HBOC and CRC susceptibility genes, that would alter the screening recommendations for at-risk relatives. High-risk genes contributed 5.4% (SIR 5.37, p<0.001), and moderate-risk breast cancer genes contributed 5.0% (SIR 2.51, p<0.001).We conclude that the frequency of mutation carriers among unselected pancreatic cancer cases is high enough to rationalize applying genetic panel tests that include both HBOC and CRC susceptibility genes to all newly diagnosed pancreatic cancer cases. This could further benefit at-risk relatives with cascade testing of healthy relatives for increased HBOC and CRC surveillance measures.

2678T

Effects of Western-style diet and genetic predisposition on the colon proteome. *S. Valo[1,2], D. Dermadi[2,3,4], S. Ollila[5], R. Soliymani[6], M. Pussila[2], L. Sarantaus[2], J. Linden[7], M. Baumann[6], M. Nyström[2].* 1) Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Biosciences, Genetics, University of Helsinki, Helsinki, Finland; 3) Laboratory of Immunology and Vascular Biology, Department of Pathology, Stanford University School of Medicine, Stanford, California, USA; 4) The Center for Molecular Biology and Medicine, Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA; 5) Research Program Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 6) Meilahti Clinical Proteomics Core Facility, Department of Biochemistry and Developmental Biology, Medicum, University of Helsinki, Helsinki, Finland; 7) Department of Basic Veterinary Sciences, University of Helsinki, Helsinki, Finland.

   Lifestyle and diet have a major effect on the development of colorectal cancer (CRC). Dietary habits of Western populations in particular are recognized as a risk factor for CRC. However, the mechanisms that mediate the effects of Western-style diet (WD) on colorectal tumor development are largely unknown. CRC develops via multiple steps which involve genetic changes, such as mutations in growth-regulatory genes, and epigenetic alterations, such as CpG island hypermethylation. These changes accumulate over time in the normal colonic epithelial cells due to aging and many environmental factors such as diet. Lynch syndrome (LS) is one of the most common inherited cancer susceptibility syndromes. It is caused by inherited defects of the DNA mismatch repair genes (MMR), which together with other genetic and epigenetic changes are known to accelerate tumorigenesis. Interestingly, the disease phenotype such as age of onset and tumor spectrum vary considerably between individuals that carry the same germline mutation, which suggests contribution of environmental effects. A long-term feeding experiment with mouse models for LS and sporadic CRC was conducted to characterize tumor-promoting changes in normal colonic mucosa caused by WD and/or genetic predisposition. Changes in the proteome of histologically normal colonic mucosa were monitored at different time points of feeding experiment (5 weeks, 12, 18 and 21 months) with two high-throughput proteomic methods (2D DIGE and LC-MS) followed by analysis of affected pathways. Overall, 27 out of 38 colonic tumors were detected in mice fed with WD. Data from the proteomic analysis indicated that the proteome was more consistently changed by diet and aging than by genotype. Proteomic study coupled with histological analysis of colon crypts indicated increased proliferation in the colon mucosa of mice fed with WD, which may promote colorectal tumorigenesis. Moreover, analysis of pathways indicated decreased apoptotic processes, disrupted lipid metabolism and increased oxidative stress in the normal-appearing tissue in association with WD consumption. In summary, the results offer new insights into the initiating molecular mechanisms through which Western-style diet contributes to colorectal carcinogenesis.

2679F

Elucidating the constitutional genetic causes of multiple primary tumours through next generation sequencing applied to a large patient series. *J. Whitworth[1], H. West[1], P. Smith[1], J.E. Martin-Rodriguez[1], J. Adlard[2], V.K. Ajith Kumar[3], J. Barwell[4], C. Brewer[5], D.G. Evans[6], L. Greenhalgh[7], A. Henderson[8], J. Hoffmann[9], L. Izatt[10], F. Lalloo[6], L. Side[3], A.B. Skytte[11], K. Snape[12], E. Woodward[6], M. Tischkowitz[1], E. Maher[1].* 1) Department of Medical Genetics, University of Cambridge, UK; 2) Department of Clinical Genetics, Chapel Allerton Hospital, Leeds, UK; 3) Department of Clinical Genetics, Great Ormond Street Hospital, London, UK; 4) Department of Clinical Genetics, University Hospitals Leicester, UK; 5) Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK; 6) Manchester Centre for Genomic Medicine, UK; 7) Department of Clinical Genetics, Liverpool Women's Hospital, Liverpool, UK; 8) Institute of Genetic Medicine, Newcastle, UK; 9) Department of Clinical Genetics, Birmingham Women's Hospital, UK; 10) Department of Clinical Genetics, Guys Hospital, London, UK; 11) Department of Clinical Genetics, Aarhus University Hospital, Denmark; 12) Department of Clinical Genetics, St Georges Hospital, London, UK.

Expansion in the number of inherited neoplasia genes (INGs) and capability to sequence those genes simultaneously and has provided extensive opportunities to diagnose neoplastia susceptibility syndromes. Referrals for assessment by cancer genetics services have historically been guided by suspicion of a specific syndrome/gene but a more agnostic approach, based on broad indicators of susceptibility, is likely to be beneficial in light of those advances. One such indicator is the occurrence of multiple primary tumours (MPT) in the same individual (particularly if they occur at younger ages) but a previous analysis of individuals with MPT referred to a large UK cancer genetic service showed that only around a quarter were identified as harbouring a relevant pathogenic variant in an ING[1]. In cases where a constitutional genetic cause is present, possible reasons for non-detection include an atypical phenotype leading to the relevant ING not being tested, more than one deleterious variant in the same individual, mosaicism, structural variants affecting INGs and involvement and novel INGs. In order to address these possibilities, we have established an undiagnosed MPT patient series (currently n≈550) and are applying NGS technologies to blood DNA samples. Central techniques include the Illumina TruSight cancer panel to ensure high sequencing depth (>100X) of 94 known INGs and whole exome/genome sequencing. We have achieved a molecular diagnosis in at least 10% of cases from panel sequencing. Review of these cases revealed that the diagnosis had often not been made previously because of atypical or under-recognised tumour associations with the relevant ING. Unusual combinations of multiple primary tumour phenotypes may also result from *multilocus inherited neoplastic alleles syndrome* (MINAS)[2] and we have established a genotype/phenotype database (http://databases.lovd.nl/shared/diseases/04296) to aid clinical management and prognostification in such cases. Investigations of the role of mosaicism for known INGs, CNVs/structural variants and novel INGs in the cases that are currently undiagnosed are ongoing.   1.Eur. J. Hum. Genet. 2015 23, 581-587 2.JAMA Oncol. 2016 2(3), 373-9.

2680W

Germline *ESR2* mutation predisposes to medullary thyroid carcinoma and causes up-regulation of RET expression. *E.R. Woodward[1,2,4], J. Smith[2], M. Read[3], J. Hoffman[4], R. Brown[5], B. Bradshaw[6], C. Campbell[1], T. Cole[4], J. Dieguez Navas[7], F. Eatock[8], J.S. Gundara[9], E. Lian[10], D. McMullan[6], N.V. Morgan[3], L. Mulligan[10], P.J. Morrison[11], M. Robledo[12], M.A. Simpson[13], V. Smith[3], S. Stewart[4], R.C. Trembath[14], S. Sidhu[9], F.S. Togneri[6], N.C. Wake[2], Y. Wallis[6], J.C. Watkinson[5], E.R. Maher[3,15], C. McCabe[3].* 1) Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre (MAHSC), Manchester M13 9WL, UK; 2) Centre for Rare Diseases and Personalised Medicine, University of Birmingham, Birmingham B15 2TT, UK; 3) School of Clinical and Experimental Medicine, University of Birmingham, B15 2TT, UK; 4) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham B15 2TG, UK; 5) Queen Elizabeth Hospital, Queen Elizabeth Medical Centre, Birmingham B15 2TH, UK; 6) West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital, Birmingham B15 2TG, UK; 7) Human Biomaterials Resource Centre, College of Medical and Dental Sciences, University of Birmingham, Vincent Drive, Edgbaston B15 2TT, UK; 8) Department of Endocrine Surgery, Belfast Health and Social Care Trust, Royal Victoria Hospital, Belfast, Northern Ireland, UK; 9) Cancer Genetics, Level 9, Kolling Building and Endocrine Surgical Unit, Royal North Shore Hospital, University of Sydney, Pacific Highway, St. Leonards, NSW, 2065, Australia; 10) Division of Cancer Biology and Genetics, Cancer Research Institute, Queen's University, Canada; 11) Centre for Cancer Research and Cell Biology, Queens University of Belfast, 97 Lisburn Road, Belfast BT9 7AE; 12) Hereditary Endocrine Cancer Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain and Centro de Investigación Biomédica en Red de Enfermedades Raras, Madrid, Spain; 13) Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London, UK; 14) Queen Mary University of London, Barts and The London School of Medicine and Dentistry, London, UK; 15) Department of Medical Genetics, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.

Familial medullary thyroid cancer (MTC) and its precursor, C cell hyperplasia (CCH), is associated with germline *RET* mutations causing multiple endocrine neoplasia type 2. However, some rare families with apparent MTC/CCH predisposition do not have a detectable *RET* mutation. To identify novel MTC/CCH predisposition genes we undertook exome resequencing studies in a family with apparent predisposition to MTC/CCH and no identifiable *RET* mutation. We identified a novel *ESR2* frameshift mutation, c.948delT, which segregated with histological diagnosis following thyroid surgery in family members and demonstrated loss of *ESR2* encoded ERβ expression in the MTC tumour. ERα and ERβ form heterodimers binding DNA at specific estrogen response elements (ERE) to regulate gene transcription. ERβ represses ERα mediated activation of the ERE and the *RET* promoter contains three ERE. *In vitro*, we showed that *ESR2* c.948delT results in unopposed ERα mediated increased cellular proliferation, activation of the ERE and increased RET expression. *In vivo*, immunostaining of CCH and MTC using an anti-RET antibody demonstrated increased RET expression. Together these findings identify germline *ESR2* mutation as a novel cause of familial MTC/CCH and provide important insights into a novel mechanism causing increased RET expression in tumourigenesis.

2681T

Germline hemizygous deletion of *CDKN2A-CDKN2B* **locus in a patient presenting with Li-Fraumeni syndrome.** *S. Chan[1], W.K. Lim[1,4], S. Michalski[2], J.Q. Lim[1], D. Ishak[1], M. Met-Domestici[1], C. Ng[1], K. Vikstrom[2], E. Epslin[2], J. Fulbright[2], M.K. Ang[1], J. Wee[1], K. Sittapalam[3], M. Farid[1], S. Lincoln[2], K. Itahana[4], S. Abdullah[5], B.T. Teh[1,4], J. Ngeow[1,4].* 1) National Cancer Centre Singapore, Singapore, Singapore; 2) Invitae, San Francisco, California, U.S.A. 94107; 3) Singapore General Hospital SIngapore; 4) Duke NUS Medical School, Singapore; 5) RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam.

Li-Fraumeni syndrome (LFS) is a rare cancer predisposition syndrome usually associated with *TP53* germline alterations. Its genetic basis in *TP53* wild-type pedigrees is less understood. Using whole genome sequencing (WGS), we identified a germline hemizygous deletion ablating *CDKN2A-CDKN2B* in a *TP53* wild-type patient presenting with high grade sarcoma, laryngeal squamous cell carcinoma and a family history suggestive of LFS. Quantitative reverse transcription PCR (RT-qPCR) analysis confirmed that basal gene expression of *p14[ARF]* and *p16[INK4A]* was 50% lower than in healthy controls (*p*<0.001 and *p*<0.01 respectively). Immunohistochemical analysis of both MPNST and laryngeal SCC tumors were clearly null for p14[ARF] and p16[INK4A] consistent with the genomic loss observed by WGS . Patient-derived LCLs demonstrated reduced basal gene and protein expression of the *CDKN2A*-encoded tumor suppressors p14[ARF] and p16[INK4A] with concomitant decrease in p21 and faster cell proliferation, implying potential deregulation of p53-mediated cell cycle control. Review of 13 additional patients with pathogenic *CDKN2A* variants suggested associations of germline *CDKN2A* mutations with an expanded spectrum of non-melanoma familial cancers. To our knowledge, this is the first report of a germline gross deletion of the *CDKN2A-CDKN2B* locus in an LFS family. These findings highlight the potential contribution of germline *CDKN2A* deletions to cancer predisposition and the importance of interrogating the full extent of *CDKN2A* locus in clinical testing gene panels.

2682F

Exome sequencing reveals germline gain-of-function *EGFR* **mutation in an adult with Lhermitte-Duclos Disease.** *S. Colby[1,2], L. Yehia[1,3], F. Niazi[1], J. Chen[1], J.L. Mester[1], C. Eng[1,4,5,6].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Case Western Reserve University School of Medicine, Cleveland, OH; 3) Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Taussig Cancer Institute, Cleveland Clinic Foundation, Cleveland, OH; 5) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 6) Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

Lhermitte-Duclos disease (LDD) is a rare hamartomatous lesion of the cerebellar cortex that is pathognomonic for Cowden syndrome (CS), a multisystem autosomal-dominant disorder characterized by elevated risk for benign and malignant neoplasms. Presently, the only known etiology for LDD is germline phosphatase and tensin homolog (*PTEN*) mutation, however there is a subset of LDD patients who are negative for germline alterations in all genes known to be associated with CS (*PTEN*, *SDHB-D*, *PIK3CA/AKT1*, *KLLN*). Therefore, despite the rarity of this disease entity (estimated at 1 in 1,000,000), LDD provides a valuable patient population for discovery of other CS-relevant predisposition genes, which, like *PTEN*, have far-reaching somatic implications in common sporadic neoplasias. Whole-exome sequencing and variant prioritization of eight unrelated LDD probands who are wildtype for the known CS susceptibility genes revealed a pathogenic missense heterozygous epidermal growth factor receptor (*EGFR*) variant (c.977G>T [p.Cys326Phe]) in one proband. This variant disrupts a disulfide bridge in the extracellular domain of the receptor. We show the *EGFR* p.Cys326Phe variant is an activating gain-of-function mutation resulting in increased EGFR autophosphorylation, consistent with enhanced receptor activation. We demonstrate increased downstream phosphorylation of ERK1/2 and AKT, consistent with pathway activation mimicking loss-of-function of PTEN. Our findings suggest that activating mutations in *EGFR* contribute to the pathogenesis of *PTEN* wildtype LDD. Finding the underlying molecular etiology of identical phenotypes facilitates more accurate risk assessment, specific medical management and correct choice of targeted therapy, or in the future, targeted prevention.

2683W

Results of a universal tumor screening program for Lynch syndrome: Do *PMS2* mutations account for more cases than previously estimated? *J.E. Hunter[1], K.R. Muessig[1], J.M. Zepp[1], K.A. Arnold[1], K.M. Bergen[1], J. Cook[1], J.V. Davis[1], A.F. Rope[1], L.S. Acheson[2], S.K. Peterson[3], S. Syngal[4], G.L. Weisner[5], J.A. Reiss[1], K.A.B. Goddard[1].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Departments of Family Medicine and Community Health, Reproductive Biology, and Oncology, Case Western Reserve University, Cleveland, OH; 3) Department of Behavioral Science, University of Texas MD Anderson Cancer Center, Houston, TX; 4) Dana-Farber Cancer Institute, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 5) Vanderbilt Hereditary Cancer Program, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN.

Lynch syndrome (LS) is associated with an increased risk of several cancers, including colorectal (CRC) and endometrial cancers (EC). LS is due to mutations in DNA mismatch repair genes with 90% of cases attributed to pathogenic mutations in *MLH1* and *MSH2*, 7-10% in *MSH6*, <5% in *PMS2*, and 1-3% in *EPCAM*. Until recently, most patients with LS were identified by selective screening of patients who met high-risk criteria based on personal and family history of cancer and early age at onset. However, not all patients with LS meet high-risk criteria for selective screening and not all patients who meet high-risk criteria get screened. To increase the identification of patients with LS, universal tumor screening (UTS) programs that screen all newly diagnosed cases of CRC or EC for LS, regardless of high-risk criteria, are being increasingly implemented. We report on a study of UTS implementation at Kaiser Permanente Northwest among newly diagnosed cases of CRC. Individuals not included in the UTS program were those who declined screening, had a prior diagnosis of LS or another hereditary cancer syndrome, or had already initiated screening for LS with the genetics department prior to contact. Of the 191 patients who consented to the UTS program, tumor tissue was screened with microsatellite instability (MSI) testing. Patients with a high MSI result, indicative of LS, were offered follow-up testing, including immunohistochemistry, *MLH1* hypermethylation, *BRAF* mutation, and, when appropriate, germline genetic testing. A diagnosis of LS was confirmed for 6 unrelated patients (1 *MLH1*, 1 *MSH2*, 1 *MSH6*, and 3 *PMS2*). The high proportion of *PMS2* mutations identified via UTS (50%) was unexpected because *PMS2* had been previously estimated to account for a low proportion of LS cases. Given that *PMS2* mutations are associated with lower penetrance and later age at onset compared to mutations in other LS-associated genes, patients with *PMS2* mutations may be less likely to be identified by selective screening because they do not meet the high-risk criteria and may be more likely to be identified by UTS programs. Literature review of previously published UTS programs indicated that *PMS2* mutation testing was often not performed or the sequencing was limited. Thus including *PMS2* germline testing in UTS programs is likely to increase the diagnosis of patients with *PMS2* mutations, which could lead to this gene accounting for a higher proportion of LS cases than previously estimated.

2684T

Evaluation of BRCA1 related DNA repair genes in breast and ovarian carcinogenesis. *L. Maresca[1], R. Scarpitta[1], L. Spugnesi[1], S. Lodovici[2], M. Vitiello[2], L. Poliseno[2], A. Galli[2], M.A. Caligo[1].* 1) Section of Medical Genetics, University of Pisa, Pisa, Italy; 2) Institute of Clinical Physiology, CNR, Pisa, Italy.

Breast cancer is the most frequently occurring neoplasia in women. 5-10% of all breast and ovarian cancers can be explained by mutations in *BRCA* genes. *BRCA1* is a tumor suppressor gene encoding a nuclear phosphoprotein involved in many cellular processes. *BRCA1* is a highly polimorfic gene. Some of these mutations lead to the production of a truncated non functional protein. In addition, many missense mutations, called VUS (variants of unknown significance), have been identified. These variants have no known pathological significance. BRCA1 interacts with several proteins involved in DNA repair including MSH2, MSH6, MRE11A, RAD50, RAD51. The aim of this study is to determine which, among BRCA1 partner genes involved in DNA repair, may contribute to breast and ovarian carcinogenesis. Previous experiments performed by our research group, suggested that *MSH2*, *MSH6*, *MRE11A*, *RAD50* and *RAD51* may have a role in genomic instability induced by BRCA1 in *S. Cerevisiae*. We also performed a *MSH2* mutational analysis in selected breast and ovarian tumors from *BRCA1* VUS carriers and we identified probably pathogenic mutations in 36% of cases. On the basis of these results, the purpose of this work is to assess the mutational status of *MSH6*, *MRE11A, RAD50* and *RAD51* genes in breast and/or ovarian tumor tissues. Our hypothesis is that mutations in these genes may contribute to BRCA1 mediated carcinogenesis. We selected 34 FFPE tumors from patients affected by HBOC: 14 *BRCA1* VUS carriers, 8 *BRCA1* mutated and 12 *BRCA1*wt. Mutational analysis was performed by NGS targeted resequencing approach on PGM Ion Torrent. Sequencing data were analyzed by Ion Reporter software with high stringency somatic analysis setting. We identified 46 variants: 20 unique and 26 present in more than one patient. Among them, we selected 17 interesting and rare variants. All of them were confirmed and assessed in lymphocytic DNA. In conclusion we identified 8 somatic missense mutations probably pathogenic according to the prediction tools interrogated. 4 variants were on *MSH6* gene, 2 on *MRE11A*, 1 on *RAD50* and 1 on *RAD51*. Interestingly, *MSH6* was found altered in 20% of tumors analyzed and affected by predicted pathogenic mutations in 11% of samples. Moreover, in order to validate yeast results we will perform a GFP-dependent homology directed repair assay in MCF7 cell line knock out for MSH2 gene. This gene will be inactivated by CRISPR Cas9 system.

2685F

Germline compound heterozygous poly-glutamine deletion in *KIAA2018/usf3* **gene causes a pro-migration cellular phenotype in heritable and sporadic epithelial thyroid carcinoma.** *Y. Ni[1,4,5], S. Seballos[1], B. Fletcher[1], T. Romigh[1], L. Yehia[1], J. Mester[1,2], L. Senter[6,8], M. Saji[7,8], M.D. Ringel[7,8], T. LaFramboise[1,4,9], C. Eng[1,2,3,4,9].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; 3) Stanley Shalom Zielony Nursing Institute, Cleveland Clinic, Cleveland, OH; 4) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 5) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 6) Division of Human Genetics, Ohio State University School of Medicine, Columbus, OH; 7) Division of Endocrinology, Diabetes and Metabolism, Department of Internal Medicine, The Ohio State University School of Medicine, Columbus, OH; 8) Thyroid Center Unit, Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 9) Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

   Cowden syndrome (CS, [MIM 158350]) is an autosomal dominant disorder that predisposes to breast, thyroid, and other epithelial cancers. Differentiated thyroid carcinoma (DTC), as one of the major component cancers of CS, is the fastest rising incident cancer in the USA, and the most familial of all solid tumors. To identify novel candidate genes of CS and potentially DTC, we analyzed a multi-generation CS-like family with papillary thyroid cancer (PTC), applying a combined linkage-based and whole-genome sequencing strategy and identified an in-frame germline compound heterozygous deletion, *del1Q/3Q* in *KIAA2018/USF3*. Among 49 unrelated CS/CS-like patients, 22% were found to have *KIAA2018del1Q/3Q*. Of 497 TCGA PTC patients, 138 (27%) were found to carry germline *KIAA201del1Q/3Q*, with somatically decreased tumor *KIAA2018* expression. We demonstrate an increased migration phenotype along with enhanced epithelial-to-mesenchymal transition (EMT) signature after *KIAA2018* knockdown or *KIAA2018del1Q/3Q* overexpression, which sensitizes cells to endoplasmic reticulum (ER) stress. Loss of *KIAA2018* function induced cell necrosis-like features and impaired respiratory capacity while providing a glutamine-dependent cell survival advantage, strongly suggests a metabolic survival and migration-favoring microenvironment for carcinogenesis. We identify a novel role for *KIAA2018/USF3* in thyroid tumorigenesis. Importantly, the results that glutamine-dependent survival and sensitivity to ER stress in KIAA2018-deficient cells provide avenues for therapeutic and adjunct preventive interventions for both sporadic cancer as well as cancer predisposition syndromes with similar mechanisms.

2686W

Identification of germline pathogenic alleles in 143 cancer predisposing genes in Mexican patients with hereditary breast cancer by massive parallel sequencing. *R. Quezada Urban[1], C.E. Díaz Velásquez[1], R. Gitler[2], M.P. Rojo Castillo[2], M. Sirota Toporek[2], A. Figueroa Morales[2], O. Moreno García[2], F. Mainero Ratchelous[3], L. Sabrina De Hoyos Arév[3], I. Delgado Enciso[4], V.H. Garzón Barrientos[5], G. Torres Mejía[7], L.I. Terrazas Valdés[1], F. Vaca Paniagua[1,6].* 1) Laboratorio Nacional en Salud: Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, Facultad de Estudios Superiores Iztacala, UNAM, México; 2) Fundación Alma, México; 3) Instituto Mexicano del Seguro Social, México; 4) Instituto Estatal de Cancerología de Colima, México; 5) Hospital General de Chilpancingo, Guerrero, México; 6) Instituto Nacional de Cancerología, México; 7) Instituto Nacional de Salud Pública, Cuernavaca, México.

   The neoplastic disease with the highest incidence and mortality worldwide is breast cancer and almost 10% of that incidence is due to inherited pathogenic variants in predisposing genes such as *BRCA1* and *BRCA2*, but they only account for nearly 50% of the cases of hereditary breast cancer and ovarian cancer syndrome (HBOC). The range of inherited breast cancer susceptibility that is not caused by these genes has not been explored in Mexicans. Therefore, it is of great importance to identify new inherited pathogenic variants that allow a detailed analysis of the molecular epidemiology of HBOC in this population. In this work we aim to analyze exonic regions and splice sites on 143 genes of 137 patients with suspected HBOC through massive parallel sequencing, to identify new variants associated with this syndrome. We selected Mexican female patients with cancer that had family cancer history. Inclusion criteria followed the guidelines of the National Comprehensive Cancer Network for Genetic/Familial High-Risk Assessment of Breast and Ovarian Cancer. Data from international databases of normal populations and cancer patients, as well as annotation information and technical parameters were used to define pathogenic variants. In these patients the genes with the highest frequency of pathogenic alleles (stopgain/loss, frameshift indels) were *BRCA2* (26.09%, 6/137), *BRCA1* (13.04%, 3/137), *FANCC* (8.7%), *MSR1* (8.7%), *FANCL* (4.35%), *SDHB* (4.35%) and *TSC2* (4.35%). New or rare missense variants with unknown clinical significance (VUS), but defined as pathogenic in ClinVar or by algorithms assessing evolutionary conservation and deleterious structural changes at protein level, were found in single patients in the genes *AIP, ANTXR1, APC, ATR, CD96, ERCC3, ERCC6, FANCA, FANCB, FANCE, LIG4, LYST, MSH6, MSR1, MTAP, PDE11A, PDGFRA, PMS2, POLE, PTCH1, RAD50, RHBDF2, RUNX1* and *WRN*. These patients did not have pathogenic alleles, suggesting a potential contribution of these VUS to disease susceptibility. This work contributes to identify new susceptibility alleles that can predispose to HBOC in the Mexican population. Further studies need to be conducted to define the clinical impact of the VUS identified here, which together with international efforts will better define genetic susceptibility to breast cancer.

2687T

Familial clustering of haematological malignances: Harbingers of wider germline cancer susceptibility. *H.S. Scott[1,2,3,4,5], A.L. Brown[1,2], D. Hiwase[6], A.W. Schreiber[5], J.H. Feng[5], M. Babic[1], N. Poplawski[7], G. Suthers[7], M. Fine[7], L. Rawlings[1], S. Branford[1,2,3,4], S. Fuller[8], I. Lewis[6], R.J. D'Andrea[2,6], C.N. Hahn[1,2,4].* 1) Department of Genetics & Molecular Pathology, Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia; 2) School of Pharmacy and Medical Sciences, Division of Health Sciences, University of South Australia, Adelaide, South Australia, Australia; 3) School of Molecular and Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia; 4) School of Medicine, University of Adelaide, Adelaide, South Australia, Australia; 5) ACRF Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, South Australia, Australia; 6) Department of Haematology, Centre for Cancer Biology, SA Pathology, Adelaide, South Australia, Australia; 7) SA Clinical Genetics Service, Directorate of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia; 8) Department of Medicine,Sydney Medical School Nepean, The University of Sydney, Sydney, NSW, Australia.

This year germline predisposits to haematological malignancy (HM) makes their debut in the WHO classification of leukaemias. It's been a long time coming, well over 100 years since the first phenotypic recognition of familial HM in 1861, and nearly 17 years since the first genetic evidence; the discovery that *RUNX1* causes thrombocytopenia predisposing to MDS/AML (Song et al 1999). Today, there are at least 12 known FHM predisposition genes. Our systematic collection of over 120 (and growing) HM families has led us to identify germline mutations in *GATA2, RUNX1, CEBPA* and *DDX41*, with approximately 50% of familial MDS/AML now attributable to known FHM genes. Phenotypic heterogeneity exists between families with different mutations in the same gene; *GATA2* mutations predispose to a range of myeloid malignancies as well as lymphoedema and immunodeficiency, and *de novo* germline *GATA2* mutations are an under-recognised but frequent event in paediatric MDS. Our *RUNX1* mutated families show phenotypic heterogeneity within families, and our data indicate that germline variants in additional genetic modifiers may influence phenotype and penetrance. We find that rare germline variants in FHM and HM genes are also enriched in sporadic HM cohorts suggesting a significant contribution to predisposition and possibly treatment response. The majority of our unsolved families have lymphoid malignancies, most commonly non-Hodgkin lymphoma and chronic lymphocytic leukaemia. Our finding of familial lymphoma associated *DDX41* mutations led us to identify segregating mutations in a number of *DDX/DHX* genes, including the Warsaw Breakage Syndrome gene, *DDX11*, implicated in dysregulated DNA damage response (DDR), in several families. Genetic alteration of DDR pathways in families that have a history of lymphoma, and also solid tumours (pan-cancer) included damaging mutations in a range of BRCA/FANC genes such as *PALB2,* which predisposes strongly to both lymphoma and breast cancer in one family. Collectively, these mutations may have a large impact on our understanding of germline cancer predisposition, a finding supported by recent genetic population studies on sporadic cohorts, and opening the door for the application of rational therapies as evidenced by the BRCA/PARP inhibitor paradigm. However, as each individual gene mutation is rare, international collaborative studies will be essential for accurate measurements of risk and prognostication to assist individuals in the clinic.

2688F

Novel germline *TP53* **mutations in the Swedish constitutional** *TP53* **cohort.** *E. Tham, S. Bajalica Lagercrantz, P. Kharaziha, C. Larsson.* Karolinska Institutet, Stockholm, Stockholm, Sweden.

Individuals with a constitutional *TP53* mutation have a high risk of cancer in childhood and/or early adulthood, but show a large phenotypic variation. In some families, pediatric tumors are more frequent while in others only breast cancer is seen. The underlying cause for this variation is unknown. However, this question is of utmost importance for which surveillance should be offered to mutation carriers. It is therefore a large need for further molecular characterization of the genomic landscape with regard to both the TP53-mutations, as well as of genetic modulating factors, in order to understand the genotype-phenotype correlation.   In the Swedish cohort of 32 families, we found 25 different mutations in the *TP53* gene. Nine have not been reported previously as germline mutations according to the HGMD or IARC databases. Five of these have not even been reported as somatic mutations in the IARC or cosmic databases and are thus truly novel. We have generated cDNA expression constructs for all nine novel germline *TP53* mutants. All constructs have been sequenced to confirm inserted mutations. Transfection in p53 null H1299 cells have demonstrated a wide range of activities among the constructs in terms of induction of apoptosis in response to 12 G irradiation, as well as variable impact on the cell cycle regulation.   We showed that all novel mutation were express stably in cell lines and decrease spontaneous apoptosis and have a poorer response to therapy which indicate that affected individual has higher risk of tumorigenesis and less response to radio therapy.

2689W

Lynch syndrome family with co-occurrence of germline pathogenic splice site mutations of *MSH2* and *MSH6*. *I. Vogelaar[1], Y. Hu[1], F. Wang[1,2], M. Fairbank[3], R. Rapkins[4], J. Balcom[5], A. Leininger[6], M. Hitchins[1].* 1) Department of Medicine (Oncology), Stanford Cancer Institute, Stanford University, Stanford, CA 94305, USA; 2) School of Public Health, Harbin Medical University, Harbin, People's Republic of China; 3) Lynch Syndrome Australia; 4) Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia; 5) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 6) Minnesota Oncology, Woodbury, MN, USA.

Lynch syndrome (LS) is caused by an autosomal dominant heterozygous germline (epi-)mutation of one of the DNA mismatch repair (MMR) genes (*MLH1, MSH2, MSH6* or *PMS2*). Disease penetrance, tumor spectrum, and mean age of onset vary depending on the gene mutated. Germline *MLH1 or MSH2* mutation carriers have especially high risks of developing colorectal cancer (CRC) and endometrial cancer (EC). *MSH2* mutation carriers also have a high risk of urinary tract cancer. Germline *MSH6* and *PMS2* mutation carriers have a substantially lower risk for CRC and age at diagnosis is approximately 10 years later compared to *MLH1* and *MSH2* mutation carriers. In *MSH6* mutation carriers, the risk for EC is comparable to *MLH1* and *MSH2* mutation carriers. Recently, a new classification system for LS was proposed based on the MMR gene mutated (e.g. *MLH1*-Lynch syndrome), with tailored clinical surveillance recommendations. We describe the first family with co-occurrence of two pathogenic germline MMR mutations, one in *MSH2* (c.2006G>T) and the other in *MSH6* (c.3936_4001+8dup). *MSH2* and *MSH6* are neighboring genes located on chr2p21. Pedigree analysis showed the two mutations segregated independently, indicating they are located on separate chr2p21 alleles; some family members carried both mutations, others carried one or the other, whilst others were negative for both. Functionally, RNA analyses showed both are splice mutations that result in out-of-frame exon skipping, ultimately resulting in protein truncation. The *MSH2* mutation leads to skipping of exon 13 and the *MSH6* mutation results in skipping of exon 9. The identification of two pathogenic mutations in one family has serious implications in terms of cancer risk stratification and genetic counselling. Firstly, the risk of cancer will vary between family members according to their mutation status, given that cancer risks vary by MMR gene mutated and number of MMR genes affected. Secondly, since both alleles are affected, double-mutation carriers are predicted to have a 100% risk of transmitting either one of the mutations to their offspring, whereas single-mutation carriers have a 50% risk of transmission. This family illustrates the need for caution when limiting genetic screening to site-specific mutation testing in families with LS (and other high-risk cancer syndromes). The possibility of a second cancer-predisposing mutation should be considered when the known familial mutation is not found in cancer-affected family members.

2690T

Germline heterozygous *SEC23B* variants in Cowden syndrome and apparently sporadic thyroid cancer: ER stress resistance and survival of the fittest. *L. Yehia[1,2], F. Niazi[1], Y. Ni[1], J. Ngeow[1,3], M. Sankunny[1], Z. Liu[1], W. Wei[1], J.L. Mester[1], R. Keri[4], B. Zhang[1], C. Eng[1,5,6,7].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH; 3) Division of Medical Oncology, National Cancer Centre, Singapore; 4) Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH; 5) Taussig Cancer Institute, Cleveland Clinic Foundation, Cleveland, OH; 6) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 7) Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

Cancer-predisposing genes associated with inherited cancer syndromes help explain mechanisms of sporadic carcinogenesis and often inform normal development. Cowden syndrome (CS) is an autosomal-dominant disorder characterized by high lifetime risks of epithelial cancers, such that ~50% of affected individuals are wildtype for known cancer-predisposing genes. Using whole-exome and Sanger sequencing of a multi-generation CS family wildtype for all known CS-associated genes and affected by thyroid and other cancers, we identified a pathogenic missense heterozygous *SEC23B* variant (c.1781T>G [p.Val594Gly]) that segregates with the phenotype. We also found germline heterozygous *SEC23B* variants in 3/96 (3%) unrelated mutation-negative CS probands with thyroid cancer and in The Cancer Genome Atlas (TCGA) apparently sporadic thyroid, breast, and endometrial cancers. We note that the TCGA thyroid cancer dataset is enriched with unique germline deleterious *SEC23B* variants associated with a significantly younger age of onset. *SEC23B* encodes Sec23 homolog B, a component of coat protein complex II (COPII), which transports proteins from the endoplasmic reticulum (ER) to the Golgi apparatus. Interestingly, germline homozygous or compound-heterozygous *SEC23B* mutations cause an unrelated disorder, congenital dyserythropoietic anemia type II, and SEC23B-deficient mice suffer from secretory organ degeneration due to ER-stress-associated apoptosis. By characterizing the p.Val594Gly variant in a non-malignant thyroid cell line, we show that it is a functional alteration that results in ER-stress-mediated survival, cell-colony formation, and overgrowth - reflecting major hallmarks of cancer. We subsequently show via RNA-seq of untreated and Tunicamycin-treated (for ER stress induction) *SEC23B*-Wildtype and *SEC23B*-p.Val594Gly thyroid cells that SEC23B affects gene expression of ribosomal proteins and multiple mitochondrial respiratory chain enzymes. We also show that SEC23B exists in the nucleus in addition to its known presence in ER-Golgi, and localizes within nucleoli of mutant cells. Our findings suggest a non-classic role for SEC23B, whereby germline heterozygous variants associate with cancer-predisposition potentially mediated by ER stress "addiction.".

2691F

Rare germline variants in *PALB2* and *BRCA2* in familial and sporadic chordoma. *R. Yang[1], Y. Xiao[1], B. Zhu[1], H. Bennett[1], M. Rotunno[1], N. Liebsch[2], K. Jones[3], A. Vogt[3], L. Burdette[3], B. Zhu[3], B. Hicks[3], S.J. Chanock[1], M. Yeager[3], M.A. Tucker[1], M.J. Kelley[4], D.M. Parry[1], A.M. Goldstein[1].* 1) Division of Cancer Epidemiology & Genetics, NCI/NIH, Bethesda, MD; 2) Massachusetts General Hospital, Boston, MA; 3) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD; 4) Duke University, Durham, NC.

Chordoma is a rare bone cancer that is believed to originate from notochordal remnants. It is typically sporadic, but occasionally occurs in a familial setting, usually in an autosomal dominant pattern. We previously identified germline *T* duplication as a major susceptibility mechanism in several chordoma families. However, genetic causes in the majority of sporadic cases and some chordoma families remain unknown. The goal of this study was to identify additional susceptibility genes in *T* duplication-negative families (five families with 19 cases) and sporadic cases (N=137) using whole-exome sequencing. We focused on rare exonic variants that were present in ≤0.1% in the 1000 Genomes Project (1,092 subjects), ESP (6,500 subjects), 500 healthy in-house population-based controls of European ancestry, and £ 1 family from our in-house database of >900 cancer-prone control families (excluding chordoma families and sporadic chordoma cases). We identified a missense non-synonymous (NS) variant in *PALB2* (c.1042C>A, p.Gln348Lys) that was present in two chordoma cases, one obligate gene carrier, and one case with juvenile pilocytic astrocytoma in one chordoma family. This variant was reported in 0.00023 and 0.00015 of individuals of European ancestry in ESP and The Exome Aggregation Consortium (ExAC), respectively. Five additional rare NS variants in *PALB2* were identified, each in a single sporadic case. *PALB2* is a binding partner of *BRCA2* and functions as a tumor suppressor gene. One of the top biological processes involving *BRCA2* is chordate embryonic development whose specific outcome is the progression of the embryo over time, from zygote formation through a stage including a notochord and neural tube until birth or egg hatching. Interestingly, we also identified 11 rare variants in *BRCA2* in 10 sporadic cases. Among these variants, one was a stopgain, one was a frameshift, and the others were NS missense variants. Two of these variants were classified as DM (disease-causing mutation) and one variant was classified as uncertain DM in the Human Gene Mutation Database for breast cancer and/or ovarian cancer. Results from the burden test showed that chordoma cases had a significantly higher number of rare exonic variants in *PALB2* (*P*=0.028) and *BRCA2* (*P*=0.025) compared to 500 population controls. We plan to validate all promising genes/variants using targeted sequencing and design experimental assays to evaluate the functional relevance of these variants.

2692W

***De novo* status and family history of cancer in individuals with germline likely/pathogenic *PTEN* variants.** *J. Mester, S. Keilman, R. Klein, E. Butler, V. Zvereff, S. Warren, K. Hruska.* GeneDx, Gaithersburg, MD.

PTEN Hamartoma Tumor syndrome (PHTS) describes individuals with germline mutation of the *PTEN* tumor suppressor gene (MIM +601728), with a frequency of *de novo* mutation estimated at almost 50%. While a clinical tool exists to estimate *a priori PTEN* mutation risk based on personal history, family history is not included. Thus we reviewed family history and testing data from individuals with pathogenic or likely pathogenic *PTEN* variants to understand whether family history may serve as a predictor of *PTEN* mutation status. Since 2000 our clinical laboratory has identified approximately 375 individuals with pathogenic or likely pathogenic *PTEN* variants through single-gene, panel-based, array, whole exome (WES), or targeted variant analysis. For the current study, patients were excluded if personal/family history or permission for research inclusion were not provided. Among 65 eligible probands, age at testing was not significantly higher for the 39 females (34 yrs) than the 26 males (28 yrs, p=0.19). Among 18 probands, familial testing revealed the variant to be inherited from a parent in 7, assumed *de novo* (parentage unconfirmed) in 8, and proven *de novo* in 3 (parentage confirmed by WES), leading to a *de novo* frequency of 61% (11/18). Age at testing was no different for those whose variant was inherited vs. *de novo* (p=0.29). Twenty-eight of 65 probands (43%) reported no first- or second-degree relatives with a PHTS-spectrum tumor, including 4 of the 7 whose variants were inherited. Nevertheless, all four were pediatric cases, placing relatives at an age below that for which tumor presentation may occur. Among relatives with cancer, median ages at diagnosis for thyroid (32yrs), endometrial (40yrs), and breast cancer (48yrs) were lower than seen in the general population per SEER data (50, 62, and 61 yrs respectively), although numbers affected were too low to derive meaningful statistical comparisons. These data imply that family history of early-onset thyroid, endometrial, or breast cancer may be a helpful predictor of *PTEN* mutation status; however, lack of cancer family history should not dissuade consideration of a PHTS diagnosis given the high likelihood of *de novo* mutation, or in pediatric cases where family members may not be old enough for tumors to occur.

2693T

Can we predict which pediatric solid tumor and brain tumor patients carry germline cancer susceptibility mutations? Results from the BASIC3 study. *S.E. Plon[1,2,3,4], S. Scollon[1,4], K. Bergstrom[1,4], T. Wang[5], R.A. Raesz-Martinez[1,4], U. Ramamurthy[1,6], D.M. Muzny[3], A. Roy[7,8], S.G. Hilsenbeck[5], Y. Yang[2,9], C.M. Eng[2,9], R.A. Gibbs[2,3], D.W. Parsons[1,2,3,4].* 1) Dept of Pediatrics, Baylor College of Medicine, Houston, TX; 2) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Texas Children's Cancer Center, Texas Children's Hospital, Houston, TX; 5) Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, TX; 6) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX; 7) Department of Pathology, Baylor College of Medicine, Houston, TX; 8) Department of Pathology, Texas Children's Hospital; 9) Baylor Miraca Genetics Laboratories, Houston, TX.

   Background: Recent studies have suggested that approximately 10% of pediatric cancer patients carry cancer susceptibility mutations. We determined whether demographic features, family history or tumor type are predictive of diagnostic findings from whole exome sequencing (WES). **Methods:** The BASIC3 study includes clinical germline and tumor WES in an ethnically diverse (49% Hispanic) cohort of sequentially diagnosed children with CNS and non-CNS solid tumors. At subject entry, a cancer genetics professional specifies whether genetic tests would be recommended based on cancer type, age and family history. These data were compared with the results of clinical germline WES. The associations between diagnostic findings and other factors were analyzed using Fisher's exact tests and between variants of uncertain significance (VUS) and other factors by Wilcoxon rank-sum tests or Kruskal-Wallis tests. **Results:** Among 255 enrolled patients (median age 6.5 years; range 0.1-17.9) there were 90 patients with CNS and 165 with non-CNS tumors. Diagnostic pathogenic or likely pathogenic cancer susceptibility variants in 19 genes were found in 25 (9.8%) patients (24 heterozygous dominant variants and 1 homozygous recessive variant). No gene was reported in more than 3 patients (3 each for *VHL* and *TP53*). Pathogenic variants were identified in genes previously associated with both adult and pediatric cancers. There was no significant association of a diagnostic finding with age, gender, race, ethnicity, or CNS versus non-CNS location. For the more common (>9 subjects each) pediatric tumor types e.g. Wilms tumor, there was no significant association between tumor type and diagnostic result. Only the "other" category including rare tumors, e.g. pheochromocytoma, had a significantly higher diagnostic rate (8/24; 33%; p=0.0007). There was a median of 3 VUS in cancer genes per subject (range 0-10) with no significant differences other than African American race (median 4; p=0.0004). In only 13 of 25 cases (52%) did the genetics professional recommend testing at study entry for the gene reported as a diagnostic finding. **Conclusions:** Diagnostic cancer susceptibility mutations were reported in 10% of unselected children with CNS and non-CNS solid tumors. Current clinical practice using age, tumor type and family history for genetic testing decisions is likely to miss a substantial number of patients and families that warrant increased cancer surveillance. Supported by 1U01HG006485.

2694F

The *PTEN* Gene ClinGen Expert Panel: A model for creating a framework for gene specific criteria using ACMG/AMP guidelines. *C. Eng[1,8], J. Mester[2], L. Milko[3], R. Ghosh[4], T. Pesaran[5], H. Costa[6], R. Karam[5], R. Huether[5], J. Ngeow[7], E. Sorokin[6], J. Barnholtz-Sloan[8], K. Hruska[2], L. Zhang[9], K. Lachlan[10], M. Hegde[11].* 1) Genomic Med Inst, Cleveland Clinic, Cleveland, OH; 2) GeneDx Laboratories, Gaithersburg, MD; 3) University of North Carolina, Durham, NC; 4) Baylor College of Medicine, Houston, TX; 5) Ambry Genetics, Aliso Viejo, CA; 6) Stanford University School of Medicine, Stanford, CA; 7) National Cancer Centre Singapore, Singapore; 8) Case Comprehensive Cancer Center, Cleveland, OH; 9) Memorial Sloan Kettering Cancer Center, New York, NY; 10) Princess Anne Hospital, Southampton, UK; 11) Emory Genetics Laboratory, Atlanta, GA.

   ClinVar is a publicly available resource of genetic variants along with evidence supporting each assertion submitted by clinical and research laboratories. Different groups using their own approaches and available evidence for variant interpretation often results in conflicting assertions. To help ClinVar users understand the level of evidence behind each assertion, ClinGen developed a hierarchical 4-star rating system where more stars indicate a greater level of review, with ClinGen-designated Expert Panels at 3 stars. For some genes an expert group already existed; for others, ClinGen is facilitating their development. PTEN, associated with Cowden, Bannayan-Riley-Ruvalcaba, and other syndromes caused by germline PTEN mutation, is the first gene in the Hereditary Cancer domain for which an expert panel has been developed within the ClinGen framework. The group's first meeting in Dec 2014 included ClinGen staff as well as experts in PTEN clinical care and research. The group was tasked with creating an expert panel drawn from both research and diagnostic groups and developing gene-specific criteria for PTEN based on the 2015 ACMG/AMP Variant Interpretation Guidelines, which will be submitted as part of the group's formal ClinVar Expert Panel application. To develop the criteria for PTEN, working groups were assembled around the following evidence types: population frequency, splicing, computational/predictive, functional, phenotype, and segregation/de novo data. The group decided to first define benign criteria, then pathogenic, and will then curate a test set of variants. Here we present proposed PTEN-specific benign criteria the group has decided upon to date. Stand-alone: allele frequency $\geq$ 1%. Strong: allele frequency 0.1-1%, homozygous (hmz) observation in 1 unaffected individual with hmz status confirmed, 2 without confirmation. Supporting: 1 observation in trans with a pathogenic variant or 3 observations in cis/phase unknown with different pathogenic variants, 2 hmz observations without clinical data. The computational/predictive working group was challenged by the paucity of known benign PTEN missense variants, and was unable to validate use of *in silico* predictors for this purpose. A formal evaluation of criteria for pathogenic variants is in progress. We hope this outline of our process and challenges for creating gene-specific criteria will be helpful to other expert panels as they develop and establish their variant interpretation guidelines.

2695W

A multiscale survey of inflammatory diseases and prostate oncophe-notypes. *K. Shameer[1,2], K.K. Yadav[3], L. Li[1,2], S.S. Yadav[3], J. O'Connor[3], J. Li[3], B.S. Glicksberg[1,2], K.W. Johnson[1,2], M.A. Badgeley[1,2], C. Elaiho[1,2], B. Read-head[1,2], B.A. Kidd[1,2], A. Kasarskis[2], A.K. Tewari[3], J.T. Dudley[1,2,4].* 1) Institute of Next Generation Health, Mount Sinai Health System, New York, NY, USA; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Department of Urology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 4) Department of Health Evidence and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Prostate cancer (PCa) is the most common cancer detected in men (181,000 annual cases), and nearly 26,000 American men die each year due to PCa and related complications. Reports have shown that PCa is more aggressive when its comorbid with inflammatory diseases. However, the correlations and risks attributed to inflammatory diseases of the abdominal cavity and onco-phenotypes are not known. To investigate how inflammatory pathways and PCa genes induce aggressive oncophenotypes in the setting of PCa, we have compiled a multiscale survey that includes data from surgical observations, inflammatory phenotypes, clinical registries, biomarkers and mouse models. We applied an integrative informatics approach with experimental validation to understand the associations between inflammatory diseases (e.g., Crohn's disease, ulcerative colitis, collagenous colitis, indeterminate colitis, ischemic colitis, diverticulitis, hernia, etc.) and PCa. We found distinct patterns of shared molecular features---gene sets, pathways, and networks---and comorbidities across inflammatory disease and PCa. For example, we found that diverticuli-tis tend to increase inflammation in the abdominal cavity and could potentially lead to aggressive prostate oncophenotypes. To test abdominal inflammation and PCa correlation, we induced inflammation in a mouse model of hiatus hernia, which resulted in an increase in the expression of the combined markers of inflammation and PCa (*TGFB, TNFA, IL6,* and *IL2*). Evaluation of pathology stage, Gleason scores and physical attributes of previous inflamma-tion observed during robotic prostatectomy surgery also reveals trend towards aggressive tumor characteristics with an increase in inflammation. Gene-set overlap analyses showed that several inflammatory disease and prostate can-cer genes share genetic modules. Collectively, our findings provide the first set of computational, experimental and clinical evidence to recommend clinicians to evaluate the impact of inflammatory disease induced oncophenotypes in patients with PCa. Given that 1.3 million patients undergo prostate-specific antigen (PSA)-triggered invasive trans-rectal biopsy, the present findings in combination with PSA could facilitate the identification patient subset with aggressive cancer. Stratifying patients at risk for prostate cancer who are un-dergoing surgical interventions of abdominal cavity for inflammation diseases could also evaluate other non-surgical or therapeutic strategies.

2696T

Impact of SNP-SNP interactions on neurocognitive outcomes in survivors of hematopoietic cell transplantation (HCT) for hematologic malignancies. *N. Sharafeldin[1], A. Bosworth[2], S.K. Patel[2], Y. Chen[2], P. Singh[1], X. Wang[1,3], J. Richman[1], E. Morse[2], M. Mather[2], C. Sun[2], L. Francisco[1], S.J. Forman[2], L. Wong[2], S. Bhatia[1].* 1) University of Alabama at Birmingham, Birmingham, AL; 2) City of Hope, Duarte, CA; 3) University of Wisconsin-Mil-waukee, Milwaukee, WI.

Impaired cognition – an increasingly recognized concern after hematopoietic cell transplantation (HCT) – has significant potential to impact societal reinte-gration. Variability in patients' risk of cognitive impairment post-HCT suggests the need to explore genetic susceptibility. Cognitive function, however, is a complex phenotype which emphasizes the importance of considering non-ad-ditive genetic variance effects on its heritability. Here we present results from a prospective longitudinal study of 277 hematologic cancer survivors (58.5% males; 68.6% non-Hispanic whites) up to 3 years post-HCT. Patients' cognitive function was assessed using a 2-hour battery of standardized neurocognitive tests pre-HCT, 6 months, 1, 2, and 3 years post-HCT. The primary outcome was the Global Deficit Score (GDS), a widely accepted indicator of cognitive impairment, which was computed at each time point. Germline DNA was obtained pre-HCT and used to genotype 56 candidate genes related to the blood-brain barrier, telomere homeostasis, neural repair, neurotransmission, and DNA repair. We used logic regression with cross-validation to identify SNP-SNP interactions by searching for Boolean combinations (AND, OR, and NOT) of SNPs within each gene. Associations of these interaction terms with post-HCT cognitive impairment were estimated using generalized estimating equation models, adjusted for age, sex, race/ethnicity, cognitive reserve, and pre-HCT GDS. We found statistically significant associations of SNPs within 5 genes after accounting for multiple testing (Bonferroni $p \leq 8.9 \times 10^{-4}$): telomere homeostasis *TEP1* (odds ratio (OR) =5.85, 95%CI: 2.50, 13.69, p-val-ue=4.7×10^{-5}) and *TERF2* (OR=4.90, 95%CI: 2.05, 11.72, p-value=3.5×10^{-4}); and DNA repair *DCLRE1C* (OR=5.07, 95%CI: 2.28, 11.24, p-value=6.7×10^{-5}), *PMS2* (OR=4.17, 95%CI: 1.96, 8.87, p-value=2.1×10^{-4}), and *EXO1* (OR=4.59, 95%CI: 1.98, 10.65, p-value=3.8×10^{-4}). Using mechanism-specific cutoff p-values identified 2 additional genes: blood-brain barrier transporter gene (total 2 genes with p cutoff=0.025): *ABCB1* (OR=3.77, 95%CI: 1.62, 8.78, p-value=0.002) and DNA repair gene (total 40 genes with p cutoff=1.25×10^{-3}): *MRE11A* (OR=4.89,95%CI: 1.89, 12.66, p-value=1.1×10^{-3}).This study is the first to show an association between genetic variants that could be pathoge-netically involved with cognitive impairment. These findings, when replicated, could help identify vulnerable populations that could benefit from remediation.

2697F

Defining the genetic architecture of lung cancer etiology. *C.I. Amos[1], R. Hung[2], Y. Han[1], Y. Bossé[3], X. Xiao[1], Y. Li[1], J.K. Field[4], H. Bickebõller[5], X. Zong[2], D.C. Christiani[6], P. Brennan[7], M.T. Landi[8], J.D. McKay[7] OncoArray Lung Cancer Consortium.* 1) Biomedical Data Science, Dartmouth, Lebanon, NH, USA; 2) Lunenfeld-Tanenbaum Research Institute of Mt. Sinai Hospital, Toronto, Canada; 3) Department of Molecular Medicine, Laval University, Québec, Canada; 4) University of Liverpool, United Kingdom; 5) University Medical Centre Göttingen, Göttingen, Germany; 6) Harvard School of Public Health, Boston, MA, USA; 7) International Agency for Res. on Cancer, Lyon, France; 8) Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, MD, USA.

Lung cancer is considered an archetypical environmentally induced disease because of the high risk from exposure to tobacco smoke. However, family studies clearly identified strong familial aggregation. Family studies and genome wide association studies have identified selected variants influencing lung cancer risk but have been underpowered to provide a more comprehensive assessment of genetic architecture. Here, we present the largest genome-wide scan of lung cancer susceptibility, in European-descent populations, comprising data derived from 29,863 patients and 55,586 controls. Analyses were adjusted for age, sex and the first three principal components for overall lung cancer and stratified by histology (adenocarcinoma, squamous carcinoma, and small cell carcinoma) and ever or never smoking. We found 24 loci influencing lung cancer risk for loci with minor allele frequencies of 0.5% or higher, of which 14 had not previously reached genome-wide significance levels. Effect sizes ranged from 0.37 for the functional variants rs17879961 in *CHEK2* to 2.12 for[LMT([1] rs11571833 in *BRCA2*, indicating the role of selected uncommon variant in strongly influencing lung cancer risk. Among the novel variants identified, 6 influenced overall lung cancer risk, 6 were specific to adenocarcinoma and 1 each influenced never and ever smokers respectively. Aside from previously described variants in the *CHRNA5* and *CYP2A6* regions that influence cigarette consumption, all other loci showed significant heterogeneity among histologies. Array based analysis found little shared heritability between adenocarcinoma and squamous carcinoma,. Evaluation of eQTL results derived from normal lung tissue identified consistent cis-acting effects of the variant rs77468143 influencing expression of *SECISBP2L*, rs6920364 influencing the extracellular ribonuclease *RNASET2*, and rs146729428 in the *EPHX2,* involved in inflammation. Several loci that influenced adenocarcinoma alone are involved in telomere maintenance or cell cycle, while several of the squamous cancer-specific loci are involved in DNA recombination and repair. These results indicate striking variation among histological subtypes of lung cancer. Further studies are ongoing to identify the impact that smoking has on lung cancer risk for these loci. These analyses provide a comprehensive assessment of genetic effects on lung cancer risk and will elucidate how these genetic variants interact with smoking behavior.

2698W

Noncoding somatic and inherited single-nucleotide variants converge to promote *ESR1* expression in breast cancer. *S.D. Bailey[1,2], K. Desai[3], K.J. Kron[1,2], P. Mazrooei[1,2], N.A. Sinnott-Armstrong[4], A.E. Treloar[1,2], M. Dowar[1], K.L. Thu[5], D.W. Cescon[1,5], J. Silvester[5], S.Y.C. Yang[1,2], S. Wu[1], R.C. Pezo[1], B. Haibe-Kains[1,2,6], T.W. Mak[2,5], P.L. Bedard[1,7], T.J. Pugh[1,2], R.C. Sallari[8], M. Lupien[1,2,9].* 1) Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada; 2) Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada; 3) 3.Department of Genetics, Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, NH, USA; 4) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 5) Campbell Family Institute for Breast Cancer Research, University Health Network, Toronto, ON, Canada; 6) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 7) Division of Medical Oncology, Department of Medicine, University of Toronto, Toronto, ON, Canada; 8) Massachusetts Institute of Technology, Cambridge, MA, USA; 9) Ontario Institute for Cancer Research, Toronto, ON, Canada.

Sustained expression of the oestrogen receptor alpha (ESR1) drives two-thirds of breast cancer (BrCa) and defines the ESR1-positive subtype. This transcription factor engages enhancers upon oestrogen stimulation setting an oncogenic expression program in BrCa cells. Somatic copy number alterations involving *ESR1* occur in approximately 1% of ESR1-positive BrCa, implying that other mechanisms underlie the persistent expression of *ESR1*. We report the significant enrichment of somatic mutations within the set of regulatory elements (SRE) converging on *ESR1* in 7% of ESR1-positive BrCa. These mutations regulate *ESR1* expression by modulating transcription factor binding to the DNA. The SRE includes a recurrently mutated enhancer whose activity is also affected by the functional inherited single nucleotide variant (SNV) rs9383590 accounting for the risk associated with the rs2046210, rs3734805 and rs9383938 BrCa risk-loci. Overall, our work highlights the importance of considering the combinatorial activity of regulatory elements as a single unit to delineate the impact of genetic alterations on single genes in cancer.

2699T

Pleiotropy among non-Hodgkin lymphoma subtypes. *S.I. Berndt[1], Z. Wang[2], A. Nieters[3], N. Camp[4], C.F. Skibola[5], J. Vijai[6], L.M. Morton[1], K.E. Smedby[7], A.R. Brooks-Wilson[8], L.R. Teras[9], S.S. Wang[10], R. Vermeulen[11], P. Cocco[12], G.G. Giles[13], C.M. Vajdic[14], A. Monnereau[15], D. Albanes[1], B.K. Link[16], E. Roman[17], B.M. Birmann[18], A. Zeleniuch-Jacquotte[19], Y. Zhang[20], S. Chanock[1], N. Chatterjee[21], S. de Sanjose[22], X. Wu[23], K. Offit[6], J.R. Cerhan[24], S.L. Slager[24], N. Rothman[1].* 1) Division of Cancer Epidemiology & Genetics, National Cancer Institute, Rockville, Maryland, USA; 2) St. Jude Children's Research Hospital, Memphis, Tennessee, USA; 3) Center of Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg, Germany; 4) Division of Hematology and Hematologic Malignancies, Department of Internal Medicine, Huntsman Cancer Institute and University of Utah School of Medicine, Salt Lake City, Utah, USA; 5) Department of Epidemiology, School of Public Health and Comprehensive Cancer Center, Birmingham, Alabama, USA; 6) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA; 7) Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 8) Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 9) Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA; 10) Department of Cancer Etiology, City of Hope Beckman Research Institute, Duarte, California, USA; 11) Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; 12) Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Monserrato, Cagliari, Italy; 13) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia; 14) Centre for Big Data Research in Health, University of New South Wales, Sydney, New South Wales, Australia; 15) Epidemiology of Childhood and Adolescent Cancers Group, Inserm, Center of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), Paris, France; 16) Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa, USA; 17) Department of Health Sciences, University of York, York, United Kingdom; 18) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 19) Department of Population Health, New York University School of Medicine, New York, New York, USA; 20) Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA; 21) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 22) Cancer Epidemiology Research Programme, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain; 23) Department of Epidemiology, MD Anderson Cancer Center, Houston, Texas, USA; 24) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA.

Non-Hodgkin lymphoma (NHL) is comprised of multiple subtypes with distinct morphologic and clinical features. Genome-wide association studies (GWAS) have identified multiple loci for specific subtypes, but the scope of pleiotropy among subtypes has not been investigated. We utilized the analytical approach, Association analysis based on SubSETs (ASSET), to evaluate pleiotropy among four common NHL subtypes. ASSET is a statistical method for pleiotropy that explores all possible combinations of subsets, choosing the subset with the maximal test statistic and then evaluating its significance with adjustment for multiple testing. Data from eight GWAS, including 3,100 chronic lymphocytic leukemia (CLL) cases, 2,847 follicular lymphoma (FL) cases, 3,857 diffuse large B-cell lymphoma (DLBCL) cases, 825 marginal zone lymphoma (MZL) cases, and 9,504 controls of European ancestry, were utilized in the analysis. A total of 17 loci (defined as +/- 500kb from the most significant SNP) reached genome-wide significance ($P < 5 \times 10^{-8}$), most of which were driven almost exclusively by one subtype and had been previously reported. Two or more subtypes contributed to seven of the loci, and all four subtypes contributed to risk at two loci, 3p24.1 (*EOMES*) and 6p21.32 (HLA Class II region), suggesting some shared heritability near genes involved in adaptive immunity. Four potential novel loci from the ASSET analysis were taken forward for replication in 4,468 NHL cases and 2,185 controls of European ancestry. One locus, identified from the subset of 3 out of 4 subtypes (CLL, FL, and MZL), reached genome-wide significance in a joint meta-analysis of the discovery and replication sample sets (10q23.33, $P = 3.27 \times 10^{-9}$). Interestingly, this SNP resides in a transcription binding site for *IRF4*, a gene harboring variants associated with risk of CLL, and near *HHEX*, a gene involved in hematopoietic differentiation. Further related pleiotropy and pathway analyses are ongoing; however, results to date suggest that the genetic architecture of common NHL subtypes is largely distinctive, with only very modest pleiotropy.

2700F

Spectrum of somatic mutations in a patient population of Cyprus with unilateral vestibular schwannoma (VS). *R. Birkenhager, M.S. Stadler, A. Neuhold, S. Arndt, A. Aschendorff, R. Laszig.* Department of Otorhinolaryngology and Head and Neck Surgery, University Medical Center Freiburg, Killianstrasse 5, D-79106 Freiburg, Germany.

A vestibular schwannoma (VS) (acoustic neuroma) is a benign tumors originated from Schwann cells (SC) of the vestibular nerve and located in the cerebellopontine angle or the inner auditory canal. The tumor results from an over proliferation of SC; these cells wrap themselves around nerve fibres, often causing gradual hearing loss, tinnitus and dizziness. It can also affect with the facial nerve causing paralysis by compression. Early detection of the tumor is sometimes difficult, because the symptoms may be subtle and may not appear in the beginning. There are two types of VS: Bilateral and unilateral. Unilateral VS account for approximately 8 % of all cranial tumors. The exact cause of unilateral VS is unknown, most occur spontaneously. Until now there is only one single gene known for bilateral VS, that leads to neurofibromatosis type 2 (NF2). Our study group consists of 14 patients of Cyprus; all of them are clinically well characterized. Genetically a mutation analysis of the NF2 gene [MIM 607379] was performed at the genomic and tumor level; by PCR based individual exon and intron transitions sequencing of the corresponding gene. Additionally a haplotyping of the corresponding chromosomal region 22q12.2 was performed to identify LOH (loss of heterozygosity) areas, which may indicate genomic and / or somatic deletions. So far we identified 8 different missense/nonsense and splice mutations in NF2 gene, on somatic level exclusively in tumor cells, which all lead to a loss of function of the gene product Merlin and the appearance of VS, Mutations in the genomic level were excluded. In 5/14 samples genomic regions were identified on chromosome 22q12.2 with suspected LOH areas. In one case, a novel heterozygous mutation [c.1231C> T, p.Arg411Cys] was detected in the NF2 gene, which has a functional relevance and leads to loss of function of the corresponding gene Merlin. In addition to somatic changes and mutations there are possible indications of a genomic predisposition for unilateral vestibular schwannoma.

2701W

Novel regulators of ESR1 activity. *T.M. Campbell, B.A.J. Ponder, K.B. Meyer.* Cancer Research UK Cambridge Institute, Cambridge, United Kingdom.

Studying transcription factor (TF) interactions and gene regulatory networks in breast cancer, we have recently identified two distinct and opposing clusters of TFs associated with enhanced breast cancer risk, defining estrogen receptor-positive and estrogen receptor-negative breast cancer, respectively. These two clusters of TFs both regulate genes linked to risk loci identified by genome-wide association studies (GWAS), suggesting commonality of mechanisms. The relative activity of these two groups of TFs has a dramatic effect on patient outcomes and is likely to influence the phenotypic plasticity observed in breast cancer.[1] Here, we examine the molecular mechanisms underlying the opposing functions of the two networks. We identified novel TFs that interact with the estrogen receptor (ESR1) using Rapid Immunoprecipitation Mass Spectrometry of Endogenous Proteins (RIME), co-immunoprecipitation and microscopy experiments. The effect of identified TFs on ESR1 transcriptional activity was analysed using qRT-PCR, proteome/transcriptome analysis and luciferase reporter assay experiments. We demonstrate for the first time that two novel TFs associated with estrogen receptor-negative disease progression interact with the ESR1-FOXA1 TF complex and are able to influence the transactivational potential of ESR1. Moreover, signalling through FGFR2, a known risk factor in breast cancer development, appears to augment this interaction and further repress ESR1 target gene expression. This finding is consistent with our novel observation that risk SNPs in the *FGFR2* gene lead to a decrease in FGFR2 expression and activity, and is associated with greater estrogen responsiveness.[2] In conclusion, we show that members of two opposing clusters of risk TFs, associated with estrogen receptor-positive and -negative breast cancer, respectively, physically interact. Two TFs associated with estrogen receptor-negative disease associate with the ESR1-FOXA1 TF complex in order to repress ESR1 activity and drive cells towards a less estrogen-dependent cancer phenotype. [1]Castro *et al.* (2016) *Nature Genetics* 48: 12-21 [2]Campbell *et al.* (2016) *Carcinogenesis* PMID: 27236187.

2702T

Association of +331G/A polymorphism of the progesterone receptor (*PgR*) gene with susceptibility to breast cancer patients from Mexican population. *D.I. Carrillo-Moreno[1,5], L.E. Figuera[2], G.M. Zúñiga[3], A.M. Puebla[4], M.P. Gallegos-Arreola[5].* 1) Universidad de Guadalajara, Guadalajara, Jalisco,. México; 2) División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Instituto Mexicano del Seguro Social (IMSS);; 3) 3Laboratorio de Inmunofarmacología, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara (U de G); 4) Laboratorio de Mutagénesis , CIBO, IMSS; 5) Laboratorio de Genética Molecular, División de Genética, CIBO, IMSS. Guadalajara, Jalisco. México.

The polymorphism *+331G/A* of the progesterone receptor (*PgR*) gene, has been associated with susceptibility to breast cancer (BC) in different populations. The PgR is a sex steroid hormone receptor that binds progesterone and play to important role in the normal breast development. The *+331G/A* (rs10895068) promoter polymorphism are associated with cancer risk possibly by altering the expression of progesterone receptor B isoform. This study determined the association of *+331G/A* polymorphism with susceptibility of BC patients from Mexican population DNA of 162 BC patients and 64 healthy controls from general population was extracted. The *+331G/A* polymorphism was determined by polymerase chain reaction (PCR). The PCR fragments were digested with NlaIV restriction enzyme; the products were separated on a 6% polyacrylamide gel. The genotype frequencies of *+331G/A* polymorphism were 92% and 91% of *G/G*; 7.4% and 9% of *G/A*; and 0.6% and 0% of *A/A* from patients and controls respectively. The comparative analysis of association none shown associated with susceptibility to BC patients [odds ratio OR 0.78 Confidence intervals CI95% 0.26-2.3), p=0.77]. Conclusion: Our results indicate that the *+331G/A* polymorphism of *PgR* non shown association with to susceptibility to BC patients from Mexican population.

2703F

Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *D. Chubb[1], P. Broderick[1], S.E. Dobbins[1], M. Frampton[1], B. Kinnersley[1], S. Penegar[1], A. Price[1], Y.P. Ma[1,5], A.L Sherborne[1], C. Palles[2], M.N. Timofeeva[3], D.T Bishop[4], M. Dunlop[3], I.P Tomplinson[2], R.S Houlston[1].* 1) Molecular and Population Genetics, The Institute of Cancer Research, London, UK; 2) Molecular and Population Genetics Laboratory,Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) Centre for Population Health Sciences, University of Edinburgh, Scotland, UK; 4) Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, St James's University Hospital, Leeds, UK; 5) Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada,.

Colorectal cancer (CRC) displays a complex pattern of inheritance. It is postulated that much of the missing heritability of CRC is enshrined in high-impact rare alleles, which are mechanistically and clinically important. In this study, we assay the impact of rare germline mutations on CRC, analysing high-coverage exome sequencing data on 1,006 early-onset familial CRC cases and 1,609 healthy controls, with additional sequencing and array data on up to 5,552 cases and 6,792 controls. We identify highly penetrant rare mutations in 16% of familial CRC. Although the majority of these reside in known genes, we identify POT1, POLE2 and MRE11 as candidate CRC genes. We did not identify any coding low-frequency alleles (1–5%) with moderate effect. Our study clarifies the genetic architecture of CRC and probably discounts the existence of further major high-penetrance susceptibility genes, which individually account for >1% of the familial risk. Our results inform future study design and provide a resource for contextualizing the impact of new CRC genes.

2704W

Cancer risks are elevated among relatives of *MRE11A, RAD50* and *NBN* mutation carriers but similar to mutation-negative families in a laboratory-based cohort. *N.M. Chun[1], D. Qian[2], S. Li[2], V. Speare[2], H. LaDuca[2], H.M. Lu[2], A.W. Kurian[1,3].* 1) Medicine, Stanford University, Stanford, CA; 2) Ambry Genetics; 3) Health Research and Policy, Stanford University, Stanford, CA.

BACKGROUND: Genes for which cancer risks are not well characterized are included on clinical next-generation sequencing-based multi-gene panel tests (MGPT), including *MRE11A*, *RAD50*, and *NBN* (MRN). The protein products of these genes form the MRN complex and play an integral role in meiotic recombination and telomere maintenance. We aimed to further characterize cancer risks of MRN mutation carriers (MRN+) through family history analysis. METHODS: MRN+ were identified through hereditary cancer MGPT at a commercial laboratory. Complete family cancer histories were obtained from ordering clinicians as approved in an IRB reviewed protocol (n=180). Standardized Incidence Ratios (SIRs) relative to population incidences of breast, ovarian and other cancers were calculated for relatives of MRN+. A comparison cohort consisted of 180 families with a proband negative for MRN mutations (MRN–). RESULTS: Among 6277 relatives of MRN+, SIRs were greater than two times the population risk for breast, ovarian, pancreatic, gastric cancer and melanoma when MRN complex genes were combined (Table 1). However, similar risks were also observed for relatives of MRN– compared to the general population (Table 1). SIRs for ovarian, gastric and pancreatic cancers trended higher for relatives of MRN+ compared to MRN–, but were not statistically significant. CONCLUSIONS: Relatives of MRN+ have substantially elevated cancer risks compared to the general population; however, these risks are similarly increased among relatives of MRN–, likely reflecting an ascertainment bias in this laboratory cohort. As such, we are unable to determine whether the increased risks in relatives of MRN+ are attributable to the MRN mutation or other factors increasing familial risk. Future studies should investigate cancer risks among relatives of unselected MRN+.

Table 1. SIRs in relatives by mutation carrier status of index proband*

	MRE11A+	RAD50+	NBN+	MRN+	MRN-
Cancer	SIR (95% CI)	SIR (95% CI)	SIR (95% CI)	SIR (95% CI)	SIR (95% CI)
Breast	**3.8 (3.0-4.8)**	**3.5 (3.0-4.1)**	**4.6 (3.9-5.4)**	**3.9 (3.6-4.3)**	**3.8 (3.4-4.2)**
Ovarian	**6.1 (3.1-10.5)**	**6.9 (4.8-9.6)**	**7.6 (4.9-11.3)**	**6.9 (5.4-8.8)**	**4.8 (3.5-6.4)**
Colon	1.5 (0.9-2.4)	1.1 (0.8-1.6)	**2.8 (2.1-3.6)**	**1.7 (1.4-2.1)**	**1.8 (1.5-2.2)**
Gastric	2.7 (0.9-6.3)	**4.6 (2.9-6.9)**	**3.7 (1.9-6.5)**	**4.0 (2.8-5.4)**	**2.7 (1.7-4.0)**
Melanoma	1.7 (0.8-3.2)	**1.8 (1.2-2.6)**	1.1 (0.5-1.9)	**1.6 (1.1-2.0)**	1.0 (0.7-1.5)
Pancreatic	**3.8 (2.0-6.7)**	**2.2 (1.3-3.4)**	**2.6 (1.4-4.4)**	**2.6 (1.9-3.5)**	**1.7 (1.1-2.4)**

* CI: confidence interval; bold type: CI does not include 1.

2705T

A genetic contribution to prostate cancer incidence in men of African ancestry. *D.V. Conti, C.A. Haiman on behalf of the African Ancestry Prostate Cancer and ELLIPSE GAME-ON Consortia.* Preventive Medicine, Univ Southern California, Los Angeles, CA.

The greater incidence of prostate cancer observed in men of African ancestry remains one of the most important unanswered health disparities globally. Reasons for this disparity by race/ethnicity are unknown, but likely include a multitude of factors, including social aspects such as access to health care, exposures to environmental and lifestyle factors and host/biological factors including inherited susceptibility. There is very strong evidence to suggest genetic differences in the allelic architecture of prostate cancer across populations. Admixture, genome-wide association (GWAS) and fine-mapping studies have identified multiple regions of the genome that harbor genetic risk alleles that are only identified in men of African ancestry, which may account, in part, for the greater disease incidence in this population. In an attempt to quantify the underlying genetic contribution to the greater incidence of prostate cancer in men of African ancestry, we combined genetic association results from all existing large-scale GWAS of prostate cancer globally in men of African ancestry totaling 10,155 cases and 10,857 controls. A total of 19.5 million genotyped and imputed SNPs, as well as insertion/deletion variants ≥1% frequency were tested for association with prostate cancer risk. Outside of the known susceptibility regions, markers at chromosomes 13q34 and 22q12, were associated with prostate cancer risk at a genome-significant level. At these susceptibility regions, the risk-associated alleles were found only in men of African ancestry. At 13q, the most associated marker (2.2% frequency in controls), was associated with an OR of 1.56 ($p=4.9 \times 10^{-12}$). This marker is located within a cluster of associated alleles 5' of the gene *IRS2* (insulin receptor S -2) and 3' of a long non-coding RNA (LINC00676). At 22q, the most notable signal was a missense variant in the *CHEK2* gene (1.5% frequency in controls; OR=1.61, $p=1.33 \times 10^{-9}$). These findings provide further support for population-specific alleles in contributing to the greater incidence of prostate cancer in men of African ancestry.

2706F

Somatic and germline mutational heterogeneity in high hyperdiploid acute lymphoblastic leukemia. *A.J. de Smith[1], K.M. Walsh[2], I. Smirnov[2], E. Sanders[1], H.M. Hansen[2], C. Metayer[3], J.L. Wiemels[1,2].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States of America; 2) Division of Neuroepidemiology, Department of Neurological Surgery, University of California San Francisco, San Francisco, CA, United States of America; 3) School of Public Health, University of California Berkeley, Berkeley, California, United States of America.

High hyperdiploid acute lymphoblastic leukemia (HD-ALL), characterized by a non-random pattern of chromosomal gains (51-67 chromosomes), is one of the most common childhood cancers. Cure rates are high, yet significant treatment-related morbidity and mortality warrant more investigations into the etiology and molecularly-targeted therapies of this disease. Deep-sequencing of 538 cancer-related genes was performed using the UCSF500 Cancer Gene Panel in 57 HD-ALL tumors from California Childhood Leukemia Study patients. To enrich for discovery of novel driver genes, patients were selected without overt *KRAS* and *NRAS* hotspot mutations or common ALL deletions. A Combined Annotation Dependent Depletion (CADD) Phred score ≥20 was used to filter predicted damaging mutations. To remove polymorphisms, we retained only mutations with allele frequency <0.01% in the Exome Aggregation Consortium (ExAC). Novel damaging somatic mutations were discovered in epigenetic regulatory genes, including *DOT1L* (n=4), with >50% of patients harboring mutations in this pathway. Receptor tyrosine kinase (RTK)/Ras/MAPK signaling pathway mutations were found in ~80% of patients, including novel mutations in *ROS1* (n=4), which mediates phosphorylation of *PTPN11*, and in the VEGF-receptor *KDR* (n=4). We detected an extraordinary level of tumor heterogeneity, with microclonal (mutant allele fraction, MAF <10%) *KRAS*, *NRAS*, *FLT3* or *PTPN11* hotspot mutations evident in 31/57 (54.4%) patients. Multiple *KRAS* and *NRAS* codon 12 and 13 microclonal mutations significantly co-occurred within tumor samples (P=4.8x10^{-4}), suggesting ongoing formation of, and selection for, *Ras* mutations. We also discovered a high frequency of likely causal germline mutations, confirmed via Sanger sequencing. At least 25% of HD-ALL patients carried one or more rare (<0.01% allele frequency in ExAC) and predicted damaging germline mutations, including in known ALL predisposition genes, DNA repair genes, or within known hotspot mutation loci that had previously been reported only in tumor genomes. Future work is required to investigate whether tumor microheterogeneity should impact therapeutic regimens and to elucidate the biological function of epigenetic dysregulation in HD-ALL. Whole-exome sequencing of more patients will be required to comprehensively assess high-penetrance predisposition variants underlying HD-ALL.

2707W

Biomarkers elucidating epithelial-mesenchymal plasticity of breast cancer stem cells. *D. Dhawan[1,2].* 1) Institute of Life Sciences, Ahmedabad University, Ahmedabad, Gujarat, India; 2) Pangenomics, Ahmedabad, Gujarat, India.

The continued expression of adhesion molecules and the intact basal lamina results in the cancerous epithelial cells being confined to a primary site. However, as the cancer progresses some cells are believed to undergo an epithelial-mesenchymal transition (EMT) event, leading to increased motility, invasion and, ultimately, metastasis of the cells from the primary tumor to secondary sites within the body. Like stem cells, these disseminated cancer cells need the ability to self-renew in order to establish and maintain a heterogeneous metastatic tumor mass. Identification of the specific subpopulation of cancer stem cells amenable to the process of metastasis is highly desirable. In this study, we have isolated and characterized cancer stem cells from luminal and basal breast cancer cell lines (MDA-MB-231, MDA-MB-453, MDA-MB-468, MCF7 and T47D) on the basis of cell surface markers CD44 and CD24; as well as Side Populations (SP) using Hoechst 33342 dye efflux. The isolated populations were analysed for epithelial and mesenchymal markers like E-cadherin, N-cadherin, Sfrp1 and Vimentin by Western blotting and Immunocytochemistry. MDA-MB-231 cell lines contain a major population of CD44+CD24- cells whereas MCF7, T47D and MDA-MB-231 cell lines show a side population. We observed higher expression of N-cadherin in MCF7 SP as compared to MCF7 NSP (Non-side population) suggesting that the SP cells are mesenchymal like cells and hence express increased N-cadherin with stem cell-like properties. There was an expression of Sfrp1 in the MCF7 NSP as compared to no expression in MCF7 SP, which suggests that the Wnt pathway is expressed in the MCF7 SP. The wound healing assays showed a significant difference in the time taken for wound closure between the MCF7 SP and the MCF7 NSP suggesting the ability for metastasis in the SP of MCF7. The mesenchymal marker Vimentin was expressed only in MDA-MB-231 cells. Hence, understanding the breast cancer heterogeneity would enable a better understanding of the disease progression and therapeutic targeting.

2708T

RAD51C and *RAD51D* **mutation screening in South African breast and/or ovarian cancer families.** *C.M. Dorfling, J. Eloff, E.J. van Rensburg.* Genetics, University of Pretoria, Pretoria, South Africa.

Most known breast cancer susceptibility genes have important roles in DNA repair by homologous recombination. Pathogenic mutations in *BRCA1* and *BRCA2* have been shown to occur in ~70% of South African high-risk breast and/or ovarian cancer families. The remaining families may be explained by other genes in the DNA repair pathway that also confer increased breast and ovarian cancer risk, such as *RAD51C* and *RAD51D*. To investigate the role of *RAD51C* and *RAD51D* in breast and ovarian cancer predisposition in South African breast and/or ovarian cancer families, we screened 74 patients from 65 families (51 breast cancer specific, 12 breast and ovarian cancer, two ovarian cancer specific) for germline mutations in these genes. Direct Sanger sequencing was performed on all coding regions of *RAD51C* and *RAD51D*, including the 5'UTR and the exon-intron boundaries. Six previously reported *RAD51D* mutations were identified; three intronic, one synonymous and two non-synonymous variants. Seven previously reported *RAD51C* variants were detected; one frameshift, two non-synonymous, three intronic and one 5'UTR variant. The truncating *RAD51C* mutation (c.93delG, p.Gly31GlyfsX9) was identified in a breast cancer only family. This truncating mutation was also described in breast and ovarian cancer families from Finland. Originally pathogenic *RAD51C* mutations were identified in families with both breast and ovarian cancer. Later studies suggested that *RAD51C* mutations contribute to a smaller fraction of breast and ovarian cancer families compared to ovarian cancer only families. Thus, our finding of a pathogenic *RAD51C* mutation in a breast cancer specific family is rare. In conclusion, no pathogenic *RAD51D* mutations were identified whereas pathogenic *RAD51C* mutations were identified in ~2% (1/51) of breast cancer specific families and overall 1.5% (1/65) in high-risk cancer families from South Africa.

2709F

Association of plasma exosome miRNAs with survival in advanced clear cell type kidney cancer. *M. Du[1], K. Giridhar[2], J. Zhu[1], M. Kohli[2], L. Wang[1].* 1) Department of Pathology and Cancer Center, Medical College of Wisconsin, Milwaukee, WI, USA; 2) Department of Oncology, Mayo Clinic, Rochester, MN, USA.

Background. Clinical characteristics are currently used for prognostication of advanced stage kidney cancer. Since these can be inaccurate in individual patients, we evaluated plasma exosomal miRNAs for prognosis in clear cell type renal cell cancer (ccRCC). **Methods.** ccRCC patients were enrolled on a tertiary hospital based repository between 9/2011 and 12/2015 and plasma was collected and processed uniformly for all patients. Exosomal RNA was isolated from plasma and RNA sequencing was performed in 43 ccRCC patients. Cox regression and Kaplan-Meier survival analyses were performed to evaluate the association of miRNAs with overall survival. Expression for significantly associated candidate exosomal miRNAs with overall survival were evaluated in TCGA ccRCC dataset with 508 patients. **Results.** RNA sequencing of plasma in the 43 ccRCC patients generated approximately 3.72 million mappable reads per patient. Of those with normalized read counts>8 RPM (reads per million), 93.6% were mapped to miRNAs for a total of 322 miRNAs, 6.2% were mapped to piwiRNAs, and 0.2% was other RNAs including mRNA, tRNA and rRNA. Cox regression analysis identified 20 exosomal miRNAs in plasma that were associated with overall survival (p < 0.05). Of these, lower expression of miR-190 was significantly associated with poor overall survival in plasma (p=0.0007, HR=0.051) and in tumor tissues from TCGA (p=1.05E-06, HR=0.67). Lower expression of miR-30d was also associated with poor overall survival in plasma (p=0.034, HR=0.31) and in TCGA tumor tissues (p=0.022, HR=0.75). In addition, we found that at least three miRNAs were associated with overall survival in plasma but not in TCGA dataset. Higher expression of miR-133b and miR-378h were associated with poor overall survival (p=0.032, HR=2.27 and p=0.006, HR=4.48, respectively). Lower expression of miR-26a-1-3p was associated with poor overall survival (p=0.001, HR=0.018). **Conclusion.** Multiple plasma exosomal miRNAs were determined as promising prognostic biomarkers for advanced ccRCC patients. These markers, if validated, could be incorporated to current clinicopathologic parameters to improve prognostic information on individual patients. Prospective validation is needed for further evaluation of these candidate miRNAs.

2710W

Routes to prostate cancer progression and identification of new drug targets: Results from the CRUK-ICGC prostate initiative. *R. Eeles[1], D.C. Wedge[2], G. Gundem[2], T.J. Mitchell[2], M. Ghori[2], J. Zamora[2], A. Butler[2], H. Whitaker[3], Z. Kote-Jarai[1], C. Massie[3], L.B. Alexandrov[3], A. Warren[4], C. Verrill[5], D. Berney[6], N. Dennis[1], S. Merson[1], M. Fraser[7], P. Boutros[8], R. Bristow[7], C. Sander[9], F. Hamdy[10], U. McDermott[2], B. Al-Lazikani[1], P. Workman[1], A.G. Lynch[3], G.S. Bova[11], C. Foster[12], D.S. Brewer[13], D. Neal[2], C.S. Cooper[13].* 1) Div of Genetics & Epidemiology, Inst Cancer Res, Sutton, Surrey, United Kingdom; 2) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 3) Cancer Research UK, Cambridge Institute, Cambridge, United Kingdom; 4) Cambridge University Hospitals NHS Foundation Trust, Addenbrookes Hospital, Cambridge, United Kingdom; 5) Oxford University Hospital NHS Foundation Trust, Oxford, United Kingdom; 6) Barts Cancer Centre, Barts and the London School of Medicine, London, United Kingdom; 7) Princess Margaret Cancer Centre, Toronto, Canada; 8) Ontario Institute for Cancer Research /MaRS Centre, Toronto, Canada; 9) Dana-Farber Cancer Institute, Boston, USA; 10) University of Oxford, Oxford, United Kingdom; 11) University of Tampere & FimLab Laboratories, Tampere, Finland; 12) HCA Healthcare, London, United Kingdom; 13) University of East Anglia, Norwich, United Kingdom.

Prostate cancer (PCa) is a challenge to treat because, while some cases are indolent, others progress and cause morbidity and mortality. We sequenced the whole genomes of 112 primary and metastatic PCas to 50X and their associated germline DNA to 30X through the CRUK - ICGC prostate cancer project. We then combined these data with publically available sequence data from prostate cancers in the TCGA. We identified 18 novel putative driver genes. Through the temporal dissection of aberrations and the stratification of tumours by ETS status and primary/metastatic site, we identified driver mutations specifically associated with steps in the progression of PCa. Using the CanSAR analysis of mutations which is a bioinformatic tool to predict which products may be susceptible to current drug targets, we identified 8 targets of current drugs, 6 targets of investigational drugs and 26 compounds that may be active and should be considered as trial candidates in future clinical trial design in this disease.

2711T

Morning fatigue in cancer patients receiving chemotherapy is associated with differential expression and methylation of regions associated with immune function/inflammation and energy metabolism. *E. Flowers[1], C. Miaskowski[1], J. Mastick[1], B. Cooper[2], S. Paul[2], K. Kober[1].* 1) Physiological Nursing, University of California, San Francisco, San Francisco, CA; 2) University of California, San Francisco, San Francisco, CA.

Fatigue is the most common symptom associated with cancer treatment. However, due to the lack of understanding of the mechanisms that underlie fatigue, no effective interventions are available. Investigation of the molecular mechanisms that are associated with higher levels of fatigue in oncology patients undergoing chemotherapy (CTX) may guide the development of more effective treatments. The objective of this study is to examine differential gene expression and methylation in a subset of breast cancer patients experiencing lower (n=14) versus higher (n=30) levels of morning fatigue. Oncology outpatients (n=582) were assessed over two CTX cycles. Morning fatigue was assessed using the Lee Fatigue Scale (LFS). In this sample, we previously described three latent classes of patients with distinct morning fatigue trajectories over two cycles of CTX (i.e., low (32%), high (51%), and very high (17%)). Gene expression and methylation were assayed from peripheral blood for 44 breast cancer patients prior to a dose of CTX with the Illumina HumanHT-12 and HumanMethylation450 arrays. No differences were found in demographic or clinical characteristics between the low (n=14, mean LFS score 4.5) and higher fatigue (n=30, mean LFS score 6.1) groups. Differences in gene expression were found between the low and higher fatigue groups for 97 genes and 135 differentially perturbed KEGG pathways. The pathways are involved in inflammation and immunity, energy metabolism, and circadian rhythm. In addition, 1585 preliminary differentially methylated positions (pDMPs) were found between the low versus higher fatigue groups. Genes annotated for these pDMPs (n=1188) were significantly enriched in 63 KEGG pathways. Three pathways overlapped with those enriched in the gene expression analysis, including chemokine signalling pathway, antigen processing and presentation, and circadian rhythm. These findings suggest that the severity of morning fatigue in oncology patients is associated with changes in gene expression and epigenetic markers, and support currently hypothesized mechanisms for fatigue. The development of interventions that target these mechanisms may be useful to treat oncology patients undergoing CTX.

2712F

Association of *2R/3R* **polymorphism of the thymidylate synthase gene with non-response to chemotherapy in colon cancer patients from Mexican population.** *K. Gómez-Mariscal[1,5], G.M. Zúñiga[2], L.E. Figuera[3], A.M. Puebla[4], M.P. Gallegos-Arreola[5].* 1) Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Laboratorio de Mutagénesis , Centro de Investigación Biomédica de Occidente (CIBO),, Instituto Mexicano del Seguro Social (IMSS); 3) División de Genética, CIBO, IMSS; 4) Laboratorio de Inmunofarmacología, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), UdeG; 5) Laboratorio de Genética Molecular, División de Genética, CIBO, IMSS. Guadalajara, Jalisco, México.

The polymorphism *2R/3R* of Thymidylate Synthase gene (*TYMS*) has been associated with susceptibility to colon cancer (CRC) in different populations. Also has been a play an important role in the conversion of dUMP to dTMP, which in turn may result in differences in the outcome of CRC chemotherapy. This study determined the association of *2R/3R* polymorphism with non-response to chemotherapy in CRC patients from Mexican population. DNA of 47 patients with CRC with 5-fluorouracil or capecitabine with chemotherapy treatment (5-FU) was extracted. The *2R/3R* polymorphism was determined by polymerase chain reaction (PCR) and wild genotype (*2R/2R*) was identified by 210bp band, the heterozygous genotype (2R/3R) by 210 and 238bp; and polymorphic genotype (*3R/3R*) by 238 bp. the PCR products were separated on a 6% polyacrylamide gel. Patients were sub-grouped by chemotherapy status: response (RC), non-response (NRC), and partial response (PRC). Results: genotype 3R*/3R* of *VNTR* polymorphism no was associated with NRC and PRC (p>0.05). Conclusion: Our results indicate that the *2R/3R TYMS* polymorphism no was associated with non-response chemotherapy in CRC patients of analyzed sample from Mexican population.

2713W

Association of *-31T<C* **polymorphism of the IL-1β gene with susceptibility to colon cancer patients from Mexican population.** *M.P. Gallegos-Arreola[1], L.E. Figuera[2], A.M. Puebla-Pérez[3], G.M. Zúñiga[4].* 1) Laboratorio de Genética, División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), IMSS,; 2) División de Genética, CIBO, IMSS; 3) Laboratorio de Inmunofarmacología, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara (U de G); 4) Laboratorio de Mutagenesis, CIBO, IMSS. Guadalajara, Jalisco. México.

In different kind of cancers have been investigated that the cytokine-mediated immune and inflammatory responses to play an important role in their pathogenesis. The present study investigated the association of *-31T<C* polymorphism of interleukin-1β (IL1β) with colon cancer (CRC) patients from Mexican population. DNA of 230 CRC patients and 220 healthy controls from general population was extracted. The *-31T<C* polymorphism in the 5′-untranslated region (5′-UTR) was determined by polymerase chain reaction (PCR) and the 272bp fragments of PCR were digested with AluI restriction enzyme, the products were separated on an 6% polyacrylamide gel. The genotype frequencies of *-31T<C* polymorphism were 23% and 29% of *T/T*; 62% and 51% of *T/C*; and 15% and 20% of C/C from patients and controls respectively. The comparative analysis of association shown that heterozygous genotype was associated with susceptibility of CRC patients [odds ratio OR 1.5 Confidence intervals CI95% 1.1-2.2), p=0.02]. Conclusion: Our results indicate that the *-31T<C* polymorphism of *IL1β* was associated with CRC patients from Mexican population.

2714T

Rare germline loss-of-function variants in melanoma susceptibility genes in familial melanoma. *A.M. Goldstein[1], Y. Xiao[1], H. Bennett[1], M. Rotunno[1], K. Jones[1,2], A. Vogt[1,2], L. Burdette[1,2], B. Zhu[1,2], M. Yeager[1,2], S.J. Chanock[1], M.A. Tucker[1], X.R. Yang[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

Approximately 10% of cutaneous malignant melanoma (CMM) cases occur in a familial setting and known high-risk CMM genes account for melanoma risk in less than 40% of melanoma-prone families, suggesting the existence of additional high-risk genes or a polygenic mechanism involving multiple genetic modifiers such as variants in intermediate-risk and low-risk genes. The goal of this analysis was to systematically characterize rare germline loss-of-function (LOF) variants in established melanoma high-risk, intermediate-risk, and low-risk genes among CMM patients in 76 American melanoma-prone families without known mutations (144 CMM cases) using data from whole-exome sequencing. Rare variants were defined as <0.1% in the 1000 Genomes Project (1,092 subjects), ESP (6,500 subjects), 500 healthy in-house population-based controls of European ancestry, and ≤1 family from our in-house database of >900 cancer-prone control families (excluding CMM and pancreatic cancer families). Among the 42 known melanoma genes evaluated, we identified nine LOF variants. Six of them remained of interest after technical validation and further evaluation in about 800 population-based CMM cases and 800 controls, including two variants in *TYRP1* and one each in *TYR*, *PLA2G6*, *POT1*, and *TINF2*. Of particular interest, a stopgain variant in *TYR* was present in 5 of the 6 sequenced CMM cases/obligate gene carriers in one family and a single sporadic CMM case. The variant was only observed in 3 out of 66,390 alleles among non-Finish Europeans in The Exome Aggregation Consortium (ExAC) database and not seen in any other public control databases or our population-based controls. This stopgain variant is classified as a disease-causing mutation in the Human Gene Mutation Database with the associated phenotype being *OCA1* albinism. A start gain variant in the 5'UTR region of *PLA2G6* was seen in all three examined cases in one family. It was not present in any evaluated public databases or our in-house controls. A stopgain variant in *POT1* was seen in one family in a single CMM case and a family member with dysplastic nevi. A frameshift variant in *TINF2* was observed in two of three cases in a family and the two *TYRP1* variants (one stopgain and one frameshift) were each observed in a single case in two different families. Our results suggest that rare LOF variants in known CMM susceptibility genes may contribute to CMM predisposition in some melanoma-prone families without known genetic causes.

2715F

Loss of selective constraints drives damage to mitochondrial genomes in cancer. *S. Grandhi[1], B. Colleen[1], M. Wesley[1], N. Ying[1,2], L. Thomas[1,2,3].* 1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio, USA; 2) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA; 3) 3Center for Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA.

Although mitochondrial genomes (mtDNA) accumulate elevated levels of mutations in cancer cells, the origin and functional impact of these mutations remain controversial. Here, we queried whole-genome sequence data from 1,916 patients across 24 cancer types to characterize patterns of mtDNA mutations and elucidate the selective constraints driving their fate. Tumor mitochondrial genomes show a distinct mutational signature and are disproportionately enriched for protein-altering changes, in contrast to the strong purifying selection found in the human germline and somatic normal cells. Moreover, protein-altering mtDNA variants that are initially present at low frequencies in normal cells preferentially expand in the altered tumor environment, suggesting selective advantage. Dramatic shifts in selective pressures suggest that tumor cells exploit the dynamic and transient nature of normal mitochondria by tuning levels of damaged mitochondrial genomes to facilitate the tumorigenic process.

2716W

Pathogenic/likely pathogenic variants in non-high risk genes in persons with multiple primary cancers. *S.A. Jackson, L.R. Susswein, J.L. Mester, K.A. Barker Grill, R.T. Klein, K.S. Hruska.* GeneDx, Gaithersburg, MD.

One criterion for hereditary cancer susceptibility testing is a personal history of multiple primary cancer diagnoses. With the evolution from single gene to multiplex testing, moderate-risk and newer cancer susceptibility genes are often included on panels with historical high-risk genes. As a clinical diagnostic laboratory, we aimed to characterize the distribution of pathogenic or likely pathogenic variants (P/LPVs) across genes and risk categories among individuals with multiple primary cancers undergoing multi-gene cancer panel testing. We retrospectively reviewed personal history and testing results from approximately 11,000 individuals who underwent Comprehensive Cancer Panel testing in our diagnostic laboratory between August 2013 and March 2016. This next-generation sequencing and deletion/duplication panel includes 17 genes associated with highly penetrant hereditary cancer syndromes, 3 causing a moderate increase in cancer risk, and up to 12 recently described in association with cancer risk. Cancers of unlikely hereditary etiology (cervical, lung, non-melanoma skin) were excluded from analysis. Cases were also excluded if prior analysis of one or more genes on this panel had been performed at this or another laboratory, or if an individual only had a single P/LPV in *MUTYH*. Of the 127 total individuals with at least one P/LPV and multiple primary cancers, 62 (48.8%) harbored a variant in a newer or moderate-risk gene. Among 121 individuals heterozygous for a single P/LPV, 52.1% (63/121) occurred in a high-risk gene, with BRCA1/2 accounting for over half (37/63, 58.7%); 38.8% (47/121) occurred in a moderate-risk gene, with most (29/47, 61.7%) in CHEK2; and 9.1% (11/121) involved a newer cancer risk gene. In addition, six individuals were found to have two P/LPVs: two with P/LPVs in high risk genes and four involving one high- and one moderate-risk gene. Although a diagnosis of multiple primary cancers may appear more consistent with a highly-penetrant hereditary cancer syndrome as the underlying etiology, a large portion of P/LPVs were identified in newer or moderate-risk genes. Furthermore, four individuals with a high-risk P/LPV also had a moderate-risk P/LPV, which would have gone undetected had testing been limited to analysis of high-risk genes. These results strongly support evaluation of moderate-risk and newer cancer susceptibility genes in individuals with a personal history of multiple primary cancers.

2717T

A microRNA target site polymorphism in *SLC7A6* confer risk of hepatitis B virus-related hepatocellular carcinoma. *D. Jiang[1,2,3,4], X. Ma[2], Q. Xiao[5], C. Conran[1], S. Zheng[1], L. Yu[2,6], J. Xu[1,2,3,4,5,7].* 1) Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, IL, Affiliate of University of Chicago Pritzker School of Medicine; 2) State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China; 3) Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; 4) Center for genetic Epidemiology, School of Life Sciences, Fudan University, Shanghai, China; 5) Center for Genomic Translational Medicine and Prevention, School of Public Health, Fudan University, Shanghai, China; 6) Institute of Biomedical Science, Fudan University, Shanghai, China; 7) Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China.

Purpose: Single nucleotide polymorphisms (SNPs) in microRNA (miRNA) binding sites within 3′-untranslated regions (UTRs) of miRNA target genes may influence miRNA–mRNA interaction and contribute to carcinogenesis. In the present study, we aimed to perform a systematic evaluation of the associations between SNPs in putative miRNA binding sites across the whole genome and risk of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). **Methods:** We used our recent genome-wide association study (GWAS) on HBV-related HCC including 2,514 chronic HBV carriers with (n=1,161) or without (n=1,353) HCC (Nat Genet 2013;45:72-5) as a discovery stage. We extracted 32,074 SNPs within miRNA target sites from the GWAS data and investigated their associations with risk of HBV-related HCC. Seven promising SNPs were then replicated in four independent case-control populations totaling 6,922 chronic HBV carriers, including 3,161 HCC cases and 3,761 controls. A meta-analysis of the discovery stage and subsequent replications was performed to identify significant SNPs ($P\square1.56\times10^{-6}$ after a Bonferroni correction for multiple tests of 32,074 SNPs). We assessed the correlations between significant SNPs and mRNA expression of genes that harbor the significant SNPs both in tumor and corresponding adjacent normal tissues from 103 HBV-related HCC patients using real-time PCR. We also assessed the binding affinity of predicted miRNAs to their target sites possessing different alleles of the significant SNPs using miRNA mimics and reporter gene assays. **Results:** We found that, among the 32,074 SNPs, rs11350 C>T in the 3′ UTR of *SLC7A6* was the only significant SNP. The C allele of rs11350 was significantly associated with increased HBV-related HCC risk under an additive model (OR=1.40, 95% CI=1.23-1.59, $P_{meta}=2.09\times10^{-7}$) and a dominant model (OR=1.42, 95%CI=1.24-1.62, $P_{meta}=1.82\times10^{-7}$). The rs11350 C allele carriers had higher *SLC7A6* mRNA expression levels in both HCC tissues and corresponding adjacent non-tumor tissues. Additionally, rs11350 was predicted to be within the miR-548d-3p and miR-548x seed binding regions. Further functional analyses showed that the rs11350 C allele weakened the binding capacity of miR-548d-3p to the 3′-UTR of *SLC7A6* compared to the T allele. **Conclusion:** Our findings indicate that rs11350, located at the miR-548d-3p binding site, may alter the expression of *SLC7A6* and contribute to the development of HBV-related HCC in Chinese populations.

2718F

Homologous recombination and the SWI/SNF chromatin remodeling pathways are necessary for the repair of formaldehyde-induced double strand breaks. *E. Juarez, A.K. McCullough.* Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR.

The SWI/SNF complex was originally identified as an important player in cancer development when biallelic truncating mutations were identified in malignant rhabdoid tumors, a highly aggressive pediatric cancer. Since then, mutations in other subunits of the SWI/SNF complex have been identified in 20% of cancers, suggesting a vital role in neoplasia. The mechanisms by which mutations in this complex specifically drive tumorigenesis are still unclear. Given the current evidence that SWI/SNF deficiency plays a role in carcinogenesis, and growing evidence that it may be acting as a bono-fide tumor suppressor, we chose to investigate the role of the SWI/SNF complex in DNA repair and genome stability. Tumor suppressor genes are often DNA repair-associated genes and protect the cell from genetic instability that favors tumor progression. Formaldehyde was used to induce DNA-protein crosslinks (DPCs) and subsequent double-strand breaks (DSB), and these studies were performed in yeast cells due to their genetic versatility. Our lab, and others, have identified cellular pathways critical for cell survival following DPC induction. Specifically, we demonstrated via a screen of the *S.cerevisiae* gene deletion library, that components of the Nucleotide Excision Repair (NER) pathway are necessary for cell survival after high-dose acute formaldehyde exposure. Conversely, components of the Homologous Recombination (HR) pathway, and not NER, were necessary for survival following low-dose chronic exposure. In this study we show that components of the SWI/SNF complex are exquisitely sensitive to formaldehyde exposure under chronic conditions, and that double mutants harboring mutations in both the HR and SWI/SNF pathways show an epistatic relationship. Deletion of the SWI/SNF ATPase core subunit, SNF2*,* also resulted in an accumulation of DSBs likely due to a delayed DNA-damage response as supported by abnormal upregulation of MEC1 (ATR) and TEL1 (ATM) gene expression. Moreover, persistent DSBs resulted in a cell cycle arrest in G2/M 24 hours after initial chronic exposure. Finally, formaldehyde exposure also resulted in failed cytokinesis resulting in multinucleated yeast cells. Therefore, we propose that the SWI/SNF complex is working with the HR repair pathway to maintain genome integrity and cell cycle progression, and believe that this interplay may be contributing to its role as a tumor suppressor in cancer development and progression.

2719W

Genomic differences between pure ductal carcinoma *in situ* **and synchronous ductal carcinoma** *in situ* **with invasive breast cancer.** *S.Y. Kim[1,3], S.H. Jung[1,3], M.S. Kim[2], I.P. Baek[1,3], S.H. Lee[4], T.M. Kim[5], Y.J. Chung[1,3], S.H. Lee[2].* 1) Microbiology, The Catholic University of Korea, Seoul, South Korea; 2) Pathology, The Catholic University of Korea, Seoul, South Korea; 3) Integrated Research Center for Polymorphism, The Catholic University of Korea, Seoul, South Korea; 4) Hospital Pathology, The Catholic University of Korea, Seoul, South Korea; 5) Medical Informatics, The Catholic University of Korea, Seoul, South Korea.

Although ductal carcinoma in situ (DCIS) precedes invasive ductal carcinoma (IDC), the related genomic alterations remain unknown. To identify the genomic landscape of DCIS and better understand the mechanisms behind progression to IDC, we performed whole-exome sequencing and copy number profiling for six cases of pure DCIS and five pairs of synchronous DCIS and IDC. Pure DCIS harbored well-known mutations (e.g., TP53, PIK3CA and AKT1), copy number alterations (CNAs) and chromothripses, but had significantly fewer driver genes and co-occurrence of mutation/CNAs than synchronous DCIS-IDC. We found neither recurrent nor significantly mutated genes with synchronous DCIS-IDC compared to pure DCIS, indicating that there may not be a single determinant for pure DCIS progression to IDC. Of note, synchronous DCIS genomes were closer to IDC than pure DCIS. Among the clinicopathologic parameters, progesterone receptor (PR)-negative status was associated with increased mutations, CNAs, co-occurrence of mutations/CNAs and driver mutations. Our results indicate that although pure DCIS has already acquired some drivers, more changes are needed to progress to IDC. In addition, IDC-associated DCIS is more aggressive than pure DCIS at genomic level and should really be considered IDC. Finally, the data suggest that PR-negativity could be used to predict aggressive breast cancer genotypes.

2720T

Genetic predisposition to breast cancer due to mutations other than *BRCA1* and *BRCA2* founder alleles among Ashkenazi Jewish women. *M.C. King, J.B. Mandell, S. Casadei, S. Gulsuner, M.K. Lee, T. Walsh.* Depts of Genome Sciences and Medicine, University of Washington, Seattle, WA.

Among Ashkenazi Jewish (AJ) breast cancer patients who do not carry one of the three founder mutations of *BRCA1* or *BRCA2*, the likelihood of carrying some other pathogenic mutation in *BRCA1*, *BRCA2*, or another breast cancer gene is not well established. We undertook this assessment in a cohort of 1007 AJ breast cancer patients, not selected for family history or age at diagnosis; previous analyses had identified a *BRCA1* or *BRCA2* founder mutation in 10% (104) of these patients (Science 302:643, 2003). For the present project, we sequenced genomic DNA from these 1007 patients for 28 known and candidate breast cancer genes using BROCA. Of the 903 patients with no founder mutation, 0.8% (7) carried rare truncating mutations in *BRCA1* or *BRCA2*. Of these seven mutations, one is apparently private; the other six appear in one or more European populations. Family history was suggestive for four of the seven mutation carriers. Frequencies of non-founder *BRCA1* and *BRCA2* mutations were higher among patients diagnosed before age 40 (2.9%) compared to those diagnosed at ages 40-49 (1.1%), ages 50-59 (0.4%), or age 60 or older (0.4%). The gene panel also revealed that 3.5% (32/903) of patients carried a pathogenic mutation in another known or candidate breast cancer gene: 29 in *CHEK2* (entirely explained by 1100delC, p.S428F, and c.444(+1)G>A), and one each in *BRIP1, GEN1*, or *NBN*. Frequencies of mutations in *CHEK2, BRIP1, GEN1*, and *NBN* were higher among patients diagnosed before age 60 (4.1%) than at age 60 or older (2.2%), and among patients with a first or second degree relative with breast cancer (4.4%) compared to no such family history (2.5%). These results suggest that among AJ breast cancer patients with none of the *BRCA1* or *BRCA2* founder mutations, about 1% carry another actionable mutation in *BRCA1* or *BRCA2* and 3% to 4% carry an actionable mutation in another breast cancer gene. Finally, we estimate that of all inherited predisposition to breast cancer identified thus far in the AJ population, 73% is due to a *BRCA1* or *BRCA2* founder allele, 5% to another *BRCA1* or *BRCA2* mutation, and 22% to a mutation in another gene. (Supported by the Breast Cancer Research Foundation.).

2721F

DDI2, a novel candidate against sporadic colorectal cancer. *L. Lei[1,2], Y. Yang[1,2], X.J. Zhao[1,2], R.T. Liu[1,2], E.F. Chen[1,2], H.J. He[1,2], J. Dong[1,2], J. Yang[1,2].* 1) Northwest University, Xi'an, ShaanXi, China; 2) Key Laboratory of Resources Biology and Biotechnology in Western China, Ministry of Education.

Identifying and understanding genomic changes in sporadic colorectal cancer (sCRC) is essential for early diagnosis and targeted treatment. Here we analysed 3 cases of sCRC samples in the Chinese Han population by exome sequencing and identified 15 gene candidates, including DNA-damage inducible protein 2 (DDI2). DDI2 is likely to be involved in multiple cellular processes connected to the cell cycle control through ubiquitin-mediated degradative pathways. It is reported that the mRNA expression of DDI2 is lower in malignant than in benign thyroid lesions. However, whether DDI2 is associated with sCRC and the underlying mechanisms have yet to be investigated. In this study, we detected one missense mutation in deep sequencing group of 50 sCRC cases and illustrated lower expression of DDI2 in adenocarcinoma than adjacent normal tissue by immunohistochemistry. Further experiments showed exogenous introduction of DDI2 into LoVo cells suppressed migration, invasion, and cell proliferation in vitro as well as tumor growth in vivo. Taken together, our results suggest DDI2 may serve as a novel candidate against tumor growth for sCRC.

2722W

Further dissection on germline defects in Singapore Lynch Syndrome patients without obvious defects in common mismatch repair genes. *Y. Liu, M.H. Chew, C.T.T. Loi, C.L. Tang.* Colorectal Surgery, Singapore General Hospital, Singapore, Singapore.

   Background: Lynch Syndrome (LS) is an autosomal,dominantly inherited syndrome associated with substantial risks for malignancies in colon, rectum, endometrium, stomach, ovaries, small bowel, ureter, billary tract, renal pelvis, brain and skin. LS accounts for approximately 5% of all colorectal cancer cases in Singapore and we have previously identified disease-causing defects in common mismatch repair (MMR) genes (*MLH1* and *MSH2*) in one-third of local LS families. **Objectives**: To verify the association of 9 copy number variation (CNV) regions and 5 single or dinucleotide loci found in previous genome-wide studies with increased risks of LS. **Methodology**: Two sets of case-control samples were used in this study. CNV was analysed through the Nanostring nCounter® System while genotyping was performed using the Sequenome MassARRAY iPLEX™ platform. **Results**: Five out of the 9 CNV regions showed variation in the copy numbers. Within these 5 regions, we identified a handful of potentially implicated genes including *UGT2B4*, *TRIM5* and *ADAM32*, while the others were gene desert areas. As regards genotyp-ing, 4 loci were verified as novel mutations in *MLH1* and *MSH6* genes that were observed only in LS patient group. The remaining locus is a polymorphic site located within the miRNA binding site at 3' UTR of *PRKD3* gene. The patient group exhibited significantly higher minor allele frequency as compared to 185 matched healthy controls (*P*=0.0035). **Discussion**: The susceptibil-ity loci validated in this study will provide insights into pathogenic defects underlying LS in families who do not harbour obvious defects in the common MMR genes. The ultimate benefits will facilitate development of genetic testing for local LS families and subsequent disease management for LS cases and disease monitoring for defect-carriers among family members.

2723T

Association of vitamin D levels and risk of ovarian cancer: A Mendelian randomization study. *S. Macgregor[1], J. Ong[1], A. Berchuck[2], P.D.P. Pharoah[3], G. Chenevix-Trench[1], P. Gharahkhani[1], R. Neale[1], P. Webb[1] On behalf of the Ovarian Cancer Association Consortium.* 1) QIMR Berghofer Medical Re-search Institute, Brisbane, Australia; 2) Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA; 3) The Centre for Cancer Ge-netic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

   In vitro and observational epidemiological studies suggest that vitamin D may play a role in cancer prevention. However, the relationship between vitamin D and ovarian cancer is uncertain, with observational studies gen-erating conflicting findings. A potential limitation of observational studies is inadequate control of confounding and reverse causality. To overcome this problem, we used Mendelian randomization (MR) to evaluate the association between single nucleotide polymorphisms (SNPs) associated with circulating 25-hydroxyvitamin D (25(OH)D) concentration and risk of ovarian cancer. We employed two SNPs with well-established associations with 25(OH)D concen-tration as instrumental variables for MR: rs7944926 (*DHCR7*) and rs12794714 (*CYP2R1*). We included 31 719 women of European ancestry (10 065 cases, 21 654 controls) from the Ovarian Cancer Association Consortium, who were genotyped using customized Illumina Infinium iSelect (iCOGS) arrays. A two-sample (summary data) Mendelian randomization approach was used, and analyses were performed separately for all ovarian cancer (10 065 cases) and for high-grade serous ovarian cancer (4 121 cases).   The odds ratio for epithelial ovarian cancer risk (10 065 cases) estimated by combining the indi-vidual SNP associations using inverse variance weighting was 1.28 (95% con-fidence interval: 1.03 to 1.58) per 20nmol/L decrease in 25(OH)D concentra-tion. The estimated odds ratio for high-grade serous epithelial ovarian cancer (4 121 cases) was 1.63 (1.20, 2.22). Genetically lowered 25-hydroxyvitamin D concentrations were associated with higher ovarian cancer susceptibility in Europeans. These findings suggest that increasing plasma vitamin D levels may reduce risk of ovarian cancer.

2724F

***MCPH1* is a novel hereditary breast cancer susceptibility gene.** *T. Mantere[1], R. Winqvist[1], S. Kauppila[2], M. Grip[3], A. Jukkola-Vuorinen[4], A. Tervasmäki[1], K. Rapakko[5], K. Pylkäs[1].* 1) Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit and Biocenter Oulu, Northern Finland Laboratory Centre NordLab Oulu, University of Oulu, Oulu, Finland; 2) Department of Pathology, Oulu University Hospital and University of Oulu, Oulu, Finland; 3) Department of Surgery, Oulu University Hospital and University of Oulu, Oulu, Finland; 4) Department of Oncology, Oulu University Hospital and University of Oulu, Oulu, Finland; 5) Laboratory of Genetics, Northern Finland Laboratory Centre NordLab Oulu, Oulu, Finland.

Breast cancer is strongly influenced by hereditary risk factors, a majority of which still remain unknown. Here, we performed a targeted next-generation sequencing of 796 genes implicated in DNA repair in 189 Finnish breast cancer cases with indication of hereditary disease susceptibility and focused the analysis on protein truncating mutations. A recurrent heterozygous mutation (c.904_916del, p.Arg304ValfsTer3) was identified in early DNA damage response gene, *MCPH1*, significantly associating with breast cancer susceptibility both in familial (5/145, 3.4%, P = 0.003, OR 8.3) and unselected cases (16/1150, 1.4%, P = 0.016, OR 3.3). A total of 21 mutation positive families were identified, of which one-third exhibited also brain tumors and/or sarcomas (P = 0.0007). Mutation carriers exhibited significant increase in genomic instability assessed by cytogenetic analysis for spontaneous chromosomal rearrangements in peripheral blood lymphocytes (P = 0.0007), suggesting an effect for MCPH1 haploinsufficiency on cancer susceptibility. Furthermore, 40% of the mutation carrier tumors exhibited loss of the wild-type allele. These findings collectively provide strong evidence for *MCHP1* being a novel breast cancer susceptibility gene, which warrants further investigations in other populations.

2725W

Case-case analysis of the genetic signature associated with inherited aggressive prostate cancer risk. *M. Mijuskovic[1], D.A. Leongamornlert[1], E.J. Saunders[1], S. Stephens[1], I. Whitmore[1], C. Cieza-Borrella[1], T. Dadaev[1], K. Govindasami[1], C. Haiman[2], D. Conti[2], R.A. Eeles[1], Z. Kote-Jarai[1].* 1) The Institute of Cancer Research, London, United Kingdom; 2) University of Southern California, Los Angeles, USA.

Prostate cancer (PrCa) is the most common malignancy diagnosed in men living in the developed world, with 250,000 deaths per year worldwide. Family history is a strong risk factor for the disease, however only a modest number of contributing moderate penetrance gene variants (e.g. in *BRCA2*) has been described to date. We hypothesized that additional rare germline variants exist that may be relevant to the patient outcome and can improve clinical management of the disease. To enrich the presence of rare variants associated with poor outcome, we performed extreme phenotype sampling of patients from the UK Genetic Prostate Cancer Study (UKGPCS), in which ~90% of patients had clinically presenting disease at diagnosis. We selected 144 of the aggressive (metastatic, age < 60 at diagnosis) and 144 non-aggressive (tumor grade ≤ T2b, no metastasis or nodal involvement, Gleason score < 7, age at diagnosis ≥ 60) cases for whole exome sequencing. All cases were of European ancestry, unrelated and had clinical follow-up data available. Samples were block randomized based on aggressive status and DNA extraction method and sequenced using Agilent SureSelectXT V5 to 20X median target coverage. Variants were called using GATK 3.5 and annotated against the GENCODE v19 transcript set. Rigorous data cleaning was performed that included the assessment of genotype quality, contamination, batch quality, sample identity and case-case confounding. The final filtered dataset included 139 aggressive and 141 non-aggressive cases with a total of 151,562 exonic/splice variants. We identified 97,889 rare variants (minor allele frequency ≤ 1%), among which 5,307 were predicted to be protein truncating (1,613 frameshift indels, 1,189 nonsense mutations, 2,494 splice site mutations, 8 start codon deletions and 3 stop codon insertions). 29 variants were found in 60 genes from the BROCA cancer risk panel, and they were overrepresented among samples with the aggressive phenotype (p=0.0031). The most frequently truncated genes from this set were *BRCA2* (1.8%), *ATM* (1.4%) and *NBN* (1%). Besides these, we found an additional 61 rare protein truncating variants in genes with a role in various DNA repair pathways. We furthermore describe the exome-wide gene and pathway-based association analysis of germline signature of aggressiveness which will aid future genetic profiling and screening for clinically significant PrCa. .

2726T

LGR5 **is a potential prognostic marker in young Egyptian rectal cancer patients.** *H. Morsy[1], A. Gaballah[2], M. Samir[3], M. Shamseya[4], H. Mahrous[1], A. Ghazal[2], W. Arafat[5], M. Hashish[1].* 1) Human Genetics Department , Medical Research Institute, Alexandria, Egypt; 2) Microbiology Department, Medical Research Institute, Alexandria, Egypt; 3) Clinical and Experimental Surgery Department, Medical Research Institute, Alexandria, Egypt; 4) Clinical and Experimental Internal Medicine Department, Medical Research Institute, Alexandria, Egypt; 5) Clinical Oncology and Nuclear Medicine Department, Faculty of Medicine, Alexandria University.

Objectives: The aim of the study was to delineate the gene expression profile of *LGR5* in rectal cancer (RC) tissues among young Egyptian patients. The study aimed at investigating the possible link between this CSCs related gene and clinical outcome, including response to neoadjuvant chemo-radiotherapy (CRT). Methods: The study was conducted on 30 young Egyptian RC patients, who were recommended to undergo neoadjuvant CRT. Paired tumor and non-tumor adjacent mucosal tissues were obtained from patients by routine biopsy techniques. Total RNA was extracted followed by reverse transcription, then quantitative PCR was performed using SYBR green. Expression levels of *LGR5* in tumor relative to adjacent non-tumor tissues were calculated using the comparative Cq method after normalization for the expression of *ACTB*. Patients were followed up for assessment of response to neoadjuvant CRT based on revised RECIST 1.1. Results: Overexpression of *LGR5* in human RC tissues compared to non- tumor tissues was detected.Furthermore, correlation analysis showed that higher expression of *LGR5* was correlated with more depth of tumor invasion, LN metastasis, advanced cTNM stage and mesorectal fascia involvement. On the other hand, it was not correlated with gender, age, tumor site nor pathological features. These results suggest that *LGR5* may not only play an important role in the progression of RC but also serve as a potential unfavorable prognostic biomarker for RC. In addition, high *LGR5* expression level was associated with poor response to CRT. ROC curve analysis showed that *LGR5* expression is an excellent predictor for response to neoadjuvant therapy. Conclusions: 1- High *LGR5* expression is, most likely, indicative of poor prognosis among young Egyptian RC patients. 2- *LGR5* expression level is proposed to be a novel predictive marker of resistance to neoadjuvant CRT in young Egyptian RC patients.

2727F

Latinas and breast cancer: Discovery of mutations in non-*BRCA1/2* **high and moderate penetrance breast cancer susceptibility genes.** *S.L. Neuhausen[1,4], J.N. Weitzel[1,4], A. Adamson[1], S. Tao[4], T.P. Slavin[1,4], C. Ricker[3], B. Nehoray[4], M. Rosenblatt[2], D. Senitzer[4], S. Sand[1,4], L. Steele[1], D. Hu[2,5], S. Huntsman[2,5], D. Castillo[1], E. Ziv[2,5].* 1) Beckman Research Institute of City of Hope, Duarte, CA; 2) UCSF Hellen Diller Comprehensive Cancer Center, San Francisco, CA; 3) USC Norris Comprehensive Cancer Center, Los Angeles, CA; 4) City of Hope Comprehensive Cancer Center, Duarte, CA; 5) UCSF Institute for Human Genetics, San Francisco, CA.

Latinos represent the largest and fastest-growing minority population in the US, yet they have been under-studied in terms of genetic susceptibility to breast cancer. Mutations in *BRCA1* and *BRCA2* (*BRCA*) explain approximately 25% of their familial breast cancer. The benefit of sequencing other breast cancer susceptibility genes in Latinas has not been defined. Seven hundred and seven *BRCA*-negative Latina women with breast cancer diagnosed under age 50 years, bilateral breast cancer, or diagnosed with both breast and ovarian cancers, and/or with at least two first- or second-degree relatives diagnosed with breast cancer under age 70 years were selected from the City of Hope Clinical Cancer Genomics Research Network and UCSF clinical cancer genetics program. Whole exome sequencing was performed on an Illumina 2500HT using customized Agilent Clinical Exome Capture libraries with a target of 70x coverage. Sequencing data were aligned to HG-19 and variants annotated with ANNOVAR. This report focuses on 13 known (*ATM, BARD1, BRIP1, CDH1, CHEK2, MRE11A, NBN, PALB2, PTEN, RAD51C, RAD51D, STK11, TP53*) and 9 (*BLM, FANCC, FANCM, NF1, RAD50, RECQL, RINT1, XRCC2*) suspected high- and moderate-penetrance breast cancer susceptibility genes. We identified pathogenic mutations in 45 women in 17 genes including *PALB2* (N=14); FANCM (N=6); *BARD1* (N=4), FANCC (N=3); *ATM, BLM, CHEK2, RAD50,* RECQL (N=2)*;* and *BRIP1, CDH1, MRE11A, NF1, PTEN, STK11, SLX4,* and *XRCC2* (N=1). Of the 30 pathogenic mutations identified, 5 were splicing, 15 were frameshift leading to truncation, and 10 were nonsense; 25 (83%) had not been reported previously in Clinvar. No likely pathogenic mutations were observed in *NBN, RAD51C, RAD51D, RINT1,* and *TP53*. Two recurrent mutations were identified, including a frameshift mutation in PALB2 (n=8; 1.1%) and a nonsense mutation in FANCM (n=4; 0.6%). In addition, of the non-synonymous variants in these genes, 63 were classified as damaging by a majority (5 of 8) of mutation classification algorithms, with 9 classified as damaging by all 8 callers. Overall, excluding non-synonymous mutations, 7.2% of Latinas carried a known or likely pathogenic mutation in 17 of the 22 genes. These mutation carriers may benefit from receiving this additional risk information to enhance decision making.

2728W

Whole-genome analysis of paediatric cancer from cases with a family cancer history: Early insights from Victoria, Australia. *T. Nguyen-Dumont[1], B. Feng[2], H. Tsimiklis[1], D. Amor[3], E. Algar[4], E. Milne[5], L.G. Spector[6], J. Ng[7], E. Smibert[8], A. Lonie[9], B.J. Pope[9,10], D.J. Park[1,9], G.S. Dite[11], M.C. Southey[1], J.A. Heath[1,12], J.L. Hopper[11].* 1) Genetic Epidemiology Laboratory, The University of Melbourne, Melbourne, Australia; 2) Department of Dermatology and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 3) Murdoch Childrens Research Institute, Melbourne, Australia; 4) Monash Health and Hudson Institute, Clayton, Australia; 5) Telethon Kids Institute, University of Western Australia, Perth, Australia; 6) Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota, USA; 7) The Royal Children's Hospital and The Royal Women's Hospital, Melbourne, Australia; 8) The Royal Children's Hospital, Melbourne, Australia; 9) Victorian Life Sciences Computation Initiative, The University of Melbourne, Melbourne, Australia; 10) Department of Computing and Information Systems, The University of Melbourne, Melbourne, Australia; 11) Centre for Epidemiology and Biostatistics, University of Melbourne, Melbourne, Australia; 12) Sidra Medical and Research Center, Doha, Qatar.

Approximately 1000 children and adolescents are diagnosed with cancer in Australia every year. However, less than 10% of all childhood tumours are associated with a known cancer predisposition syndrome. We hypothesized that whole genome analysis of DNA from affected children from families with multiple occurrences of early-onset cancers may be a valuable approach to identifying candidate genes or pathways that are associated with increased risk for childhood cancers. We performed whole genome sequencing (WGS) on 30 probands enrolled in the Victorian Paediatric Cancer Family Study (VPCFS), a population-based sample of 379 children diagnosed with cancer under the age of 15 years recruited from the two paediatric cancer hospitals for the state of Victoria, Australia. The probands for WGS were selected on the basis of sample availability and having one or more relative diagnosed with early onset cancer. Cancer in the selected probands included leukaemia (ALL and AML), Wilms tumours and brain tumours, and were diagnosed at a median age of 5 years (range, 1 to 14 years). Fifteen cancers were reported in first degree relatives and 166 cancers in second and third degree relatives. Of 181 cancers observed in the families, 17, 37 and 59 were diagnosed before the age of 30, 40 and 50 years, respectively. WGS was performed on the HiSeq XTen using blood-derived DNA. The analysis pipeline was based on the GATK best practices v3.3. We applied the PERCH (Polymorphism Evaluation, Ranking and Classification for a Heritable trait) framework for prioritisation of findings from WGS. Preliminary analysis of selected candidate genes, including *TP53, BRCA1, BRCA2, APC, DNMT3A, CTR9, NSD1, PMS2, NF1, RB1* and *RUNX1*, did not reveal any potentially pathogenic variant as defined by ClinVar. Comprehensive data analysis, including copy number variation and large rearrangements, is in progress. Germline DNA was extracted from all remaining VPCFS participants who provided a blood sample (n=745) for further screening by targeted-sequencing of genes of interest, identified from the WGS analysis and from the literature. Fresh frozen and paraffin embedded pathology material for all the cancers (where a biopsy or surgical resection was undertaken) are available for follow-up analyses. We anticipate that this dataset will reveal important insights into cancer aetiology and lead to improved cancer prevention, effective cancer screening and optimal clinical care.

2729T

Genetic and epigenetic changes in breast cancer tumors with BRCAness profile. *D. Plaseska-Karanfilska[1], M. Jakimovska[1], I. Maleva Kostovska[1], K. Popovska-Jankovic[1], K. Kubelka-Sabit[2], M. Karagjozov[2], L. Stojanovska[3], A. Arsovski[3].* 1) Macedonian Academy of Sciences and Arts, Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", Skopje, R. Macedonia; 2) Acibadem-Sistina Hospital, Skopje, R. Macedonia; 3) Re-Medika General Hospital, Skopje, R. Macedonia.

BRCAness refers to shared tumor characteristics between sporadic and BRCA-mutated cancers. Different genetic and epigenetic mechanisms might be responsible for the development of *BRCA*ness phenotype. Recently, it was proposed that BRCAness is associated with the presence of homologous recombination repair defect in the tumor in the absence of germline *BRCA1* or *BRCA2* mutations. In a subset of BRCAness *BRCA1* promoter hypermethylation is present, but the molecular causes in the remaining tumors are unknown. We aimed to investigate the relevance of copy number alterations (CNAs) and promoter methylation (PM) in BRCAness breast cancers. We studied a total of 251 fresh frozen breast tumor tissues. The samples were assessed for BRCAness, CNAs and PM by Multiplex Ligation-dependent Probe Amplification (MLPA). Twenty six BC tissues (10,4%) showed BRCAness profile. Triple negative tumors represented 41,7% of BRCAness group and 2,8% of non-BRCAness group (p=1x10^{-6}). *BRCA1/2* germline mutations were present in 23,1% (3/13) of the BRCAness tumors. *BRCA1* promoter methylation was observed in 7.69% (2/26) of *BRCA*ness tissues. *TIMP3* gene promoter was methylated in 23% of *BRCA*ness tumors and 1.8% in non-BRCAness tumors (p=1,2x10^{-4}). Statistically significant positive associations were observed between BRCAness profile and duplications and/or amplifications of *CCNE1, EMSY, MTDH* and *EGFR* and deletion of *ADAM9* gene. Amplification of *FGFR1* gene was negatively associated with BRCAness profile. Our study showed that *BRCA1/2* mutations and *BRCA1* hypermethylation explain around 30% of breast cancers with BRCAness profile. It also suggests that CNAs and PM of other cancer genes might be involved in the development of BRCAness profile.

2730F

Association of *1494del6* **polymorphism of the thymidylate synthase gene with non-response to chemotherapy in breast cancer patients from Mexican population.** *A.M. Puebla-Pérez[1], G.M. Zúñiga[2], L.E. Figuera[3], M.P. Gallegos-Arreola[4].* 1) Departamento de Farmacología, Laboratorio de Inmunofarmacología, Universidad de Guadalajara; 2) Laboratorio de Mutagénesis, Centro de Investigación Biomédica de Occidente (CIBO), Instituto Mexicano del Seguro Social (IMSS); 3) División de Genética, CIBO, IMSS; 4) Laboratorio de Genética Molecular, División de Genética, CIBO, IMSS. Guadalajara, Jalisco, México.

The polymorphism *1494del6* of Thymidylate Synthase gene (*TYMS*) has been associated with susceptibility to breast cancer in different populations despite to decreased levels of mRNA. Also plays an important role in the conversion of dUMP to dTMP, which in turn may result in differences in the outcome of breast cancer chemotherapy. This study determined the association of *1494del6* polymorphism of the *TYMS* gene with non-response to chemotherapy in breast cancer patients from Mexican population. DNA of 518 patients with breast cancer with FEC chemotherapy treatment (5-fluorouracil- epirubicin-cyclophosphamide) was extracted. The *1494del6* polymorphism (6 bp insertion/deletion) in the TS 3′-untranslated region (3′-UTR) was determined by polymerase chain reaction (PCR) and the 152bp fragments of PCR were digested with DraI restriction enzyme, the products were separated on an 8% polyacrylamide gel. Patients were sub-grouped by chemotherapy status: response (RC), non-response (NRC), non-response by recurrence (NRRC) and partial response (PRC). Results: genotype *Del/Del* of *1494del6* polymorphism was associated with NRC [odds ratio OR 2.54 Confidence intervals CI95% 1.26-5.12), p=0.0089]; NRRC [OR 2.19 (CI95% 1.09-4.66), p=0.0214)] and PRC [OR 3.59 (IC95% 1.52-8.47), p=0.0035)] to respect at RC. Conclusion: Our results indicate that the *1494del6* polymorphism of *TYMS* was associated with non-response chemotherapy in breast cancer patients from Mexican population.

2731W

Exome sequencing and methylation analysis in monozygotic twins discordant for acute lymphocytic leukemia. *A. Russo[1,2], E. Casalone[1,2], S. Guarrera[1,2], K. Mareschi[3,4], F. Fagioli[3], G. Matullo[1,2].* 1) Human Genetics Foundation, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Italy; 3) Stem Cell Transplantation and Cellular Therapy Laboratory, Paediatric Onco-Haematology, City of Science and Health of Turin, Italy; 4) Department of Public Health and Paediatrics, University of Turin, Italy.

Acute Lymphocytic Leukemia (ALL) is the most common type of cancer diagnosed in children. Although the 11q23 rearrangement is the most common cytogenetic abnormality in infant ALL, studies demonstrate that this rearrangement is not sufficient for leukemogenesis. In order to characterize somatic mutational and epigenetic profiles of one patient with infant ALL at diagnosis, blood DNA was collected from the affected child at diagnosis and remission, from her healthy monozygotic twin sister, and from their healthy parents. Five DNA samples underwent DNA methylation analysis on the Illumina HumanMethylation450 BeadChip. Data were analyzed according to standard procedures (MethyLumi, Bioconductor). Whole exome sequencing was also performed using Agilent's SureSelectQXT kit and the Illumina NextSeq500 platform. Reads were aligned to the human reference genome h19 and the Genome Analysis Toolkit was used for variant calling. SnpEff and Variant Effect predictor were used to annotate and prioritize variants. We searched for stretches of contiguous CpGs (A-Clustering method). We identified about 80 differentially methylated regions (DMRs) with methylation "deltas" between the co-twins ranging from 20% and 80% at an individual CpG level. The affected child showed consistent hypermethylation as compared to the sister: out of the ~80 DMRs, only a 3 CpGs region on chromosome 15 was less methylated in the proband. Interestingly, when comparing the affected child with herself at remission, the previously observed hypermethylation (methylation levels at diagnosis) was reverted to the levels observed for the healthy sister, for all the loci. Preliminary analysis of exome sequencing data showed 12 *de novo* single nucleotide variants (SNVs) with a population frequency less than 1%, in heterozygous state in the diseased child that were absent in both parents and healthy sister. Two out of 12 SNVs were missense variants annotated by the SIFT and PolyPhen predictors as "deleterious/damaging". The two SNVs were respectively located in the *PADI4* gene, which encodes for a protein involved in a process that may be important for tumorigenesis, and in *BNIP3L,* a gene that was found to be differentially expressed between sensitive and resistant blasts in acute myeloid leukemia. These preliminary results provide insight into new somatic mutations that together with DMRs may be potentially involved in ALL onset.

2732T

Breast cancer risk factors: Focusing on men. *R. Scarpitta[1], L. Maresca[1], L. Spugnesi[1], I. Zanna[2], D. Palli[2], M.A. Caligo[1].* 1) Section on Medical Genetics, University of Pisa, Pisa, Pisa, Italy; 2) Molecular and Nutritional Epidemiology Unit Cancer Prevention and Research Institute (ISPO), Florence, Italy.

Breast cancer (BC) is a complex multi-factorial disease. Recent molecular insights have evidenced the strong complexity of breast cancer genetics. Currently known susceptibility alleles have different frequencies and confer different disease-associated risk. The strongest risk factor is represented by inherited high-penetrance mutations involving BRCA1/2 genes, which can explain only up to 15% of all cases. Moderate- and low-penetrance polymorphic variants should importantly contribute to BC risk, probably interacting with environmental and life style factors, particularly occupational and dietary carcinogen exposures. The main goal of this project is to contribute to clarify the etiology of breast cancer by investigating how occupational and dietary carcinogen exposures might modify the penetrance of rare alleles in DNA repair genes.Our hypothesis is that the combined effects of various factors can be better investigated in male breast cancer (MBC), which represents a clean model of BC since they lack in reproductive-history related variables.Fifty MBC cases were recruited and information on their cancer family history, lifestyle habits, and working history were collected using standardized questionnaires (EPIC). DNA samples were subjected to Next Generation Sequencing for a panel of 24 genes involved in DNA repair and breast cancer predisposition.Our preliminary results show that MBCs are mostly accounted for high-penetrance mutations in BRCA genes. A rare synonymous and three predicted pathogenic missense variants were detected in BRCA1. BRCA2 is the most frequently mutated gene: 9 variants were detected (4 SNVs, 2 frameshift deletion, 2 splicing variants, 1 stop-gain). A novel frameshift mutation, one pathogenic missense mutation and two predicted pathogenic missense variants were found in BRIP1. PALB2 was found mutated only in a case (1 SNV). These variants could act as moderate penetrance alleles influencing the interaction between BRIP1 and PALB2 and their functional partner, BRCA1 and BRCA2, respectively. In low- penetrance genes, 11 additional VUS (variant of unknown significance) have been identified in 8 genes: *CDH1* (1 SNV), *CHEK2* (1 SNV), *ERCC1* (1 SNV), *MLH1* (2 SNVs), *MSH2* (1 SNV), *MUTYH* (1 SNV), *NBN* (2 SNVs), *PMS2* (2 SNVs).Although these data must be confirmed statistically, and correlated to the environmental variable, those results may help to further improve our knowledge and provide new insights into the characterization of BC.

2733F

A family-based, genome-wide association study of young-onset breast cancer: Inherited variants and maternally mediated effects. *M. Shi[1], K. O'brien[1], D. Sandler[2], J. Taylor[2], D. Zaykin[1], J. Keller[3], A. Wise[1], C. Weinberg[1].* 1) Biostatistics Br, NIEHS, Res Triangle Park, NC; 2) Epidemiology Br, NIEHS, Res Triangle Park, NC; 3) Westat Inc., Durham, NC, USA.

Young-onset breast cancer shows certain phenotypic and etiologic differences from older-onset breast cancer and may be influenced by some distinct genetic variants. Few genetic studies of breast cancer have targeted young women and no studies have examined whether maternal variants influence disease in their adult daughters through prenatal effects. We conducted a family-based, genome-wide association study of young-onset breast cancer (age at diagnosis <50 years). A total of 602 188 single-nucleotide polymorphisms (SNPs) were genotyped for 1279 non-Hispanic white cases and their parents or sisters. We used likelihood-based log-linear models to test for transmission asymmetry within families and for maternally mediated genetic effects. Three autosomal SNPs (rs28373882, P=2.8 × 10[-7]; rs879162, P=9.2 × 10[-7]; rs12606061, P=9.1 × 10[-7]) were associated with risk of young-onset breast cancer at a false-discovery rate below 0.20. None of these loci has been previously linked with young-onset or overall breast cancer risk, and their functional roles are unknown. There was no evidence of maternally mediated, X-linked, or mitochondrial genetic effects, and no notable findings within cancer subcategories defined by menopausal status, estrogen receptor status, or by tumor invasiveness. Further investigations are needed to explore other potential genetic, epigenetic, or epistatic mechanisms and to confirm the association between these three novel loci and youngonset breast cancer.

2734W

Genetic progression of high grade prostatic intraepithelial neoplasia to prostate cancer. *S. Shin[1,2], S.H. Jung[1,2], H.C. Park[1,2], H. Cho[1,2], S.H. Lee[3], Y.J. Chung[1,2].* 1) Microbiology, The Catholic University of Korea, Seoul, South Korea; 2) Integrated Research Center for Genome Polymorphism, The Catholic University of Korea, Seoul, South Korea; 3) Pathology, The Catholic University of Korea, Seoul, South Korea.

Although high grade prostatic intraepithelial neoplasia (HGPIN) is considered a neoplastic lesion that precedes prostate cancer (PCA), the genomic structures of HGPIN remain unknown. To identify the genomic landscape of HGPIN and the mechanisms for its progression to PCA, we analyzed 20 regions of paired HGPIN and PCA from six patients using whole-exome sequencing and array-comparative genomic hybridization to compare somatic mutation and copy number alteration (CNA) profiles of paired HGPIN and PCA. The number of total mutations and CNAs of HGPINs were significantly fewer than those of PCAs. Mutations in FOXA1 and CNAs (1q and 8q gains) were detected in both HGPIN and PCA ('common'), suggesting their roles in early PCA development. Mutations in SPOP, KDM6A, and KMT2D were 'PCA-specific', suggesting their roles in HGPIN progression to PCA. The 8p loss was either 'common' or 'PCA-specific'. In-silico estimation of evolutionary ages predicted that HGPIN genomes were much younger than PCA genomes. Our data show that PCAs are direct descendants of HGPINs in most cases that require more genomic alterations to progress to PCA. The nature of heterogeneous HGPIN population that might attenuate genomic signals should further be studied. HGPIN genomes harbor relatively fewer mutations and CNAs than PCA but require additional hits for the progression. Our results provide a clue to explain the long latency from HGPIN to PCA and provide useful information for the genetic diagnosis of HGPIN and PCA.

2735T

***FANCM* c.5791C>T mutation in Italian *BRCA1/2* mutation negative male breast cancer cases.** *V. Silvestri[1], P. Rizzolo[1], A.S. Navazio[1], V. Valentini[1], V. Zelli[1], M.G. Tibiletti[2], L. Varesco[3], A. Russo[4], S. Tommasi[5], G. Giannini[1], D. Calistri[6], A. Viel[7], L. Cortesi[8], M. Montagna[9], D. Palli[10], P. Radice[11], P. Peterlongo[12], L. Ottini[1].* 1) Department of Molecular Medicine, Sapienza University of Rome, Italy; 2) Unit of Pathology, Ospedale di Circolo, Varese, Italy; 3) Unit of Hereditary Cancers, IRCCS AOU San Martino – IST, Genoa, Italy; 4) Section of Medical Oncology, Department of Surgical and Oncological Sciences, University of Palermo, Italy; 5) Molecular Genetics Laboratory, Istituto Tumori "Giovanni Paolo II", Bari, Italy; 6) Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori, Forlì, Italy; 7) Unit of Functional onco-genomics and genetics, CRO Aviano, National Cancer Institute, Aviano (PN), Italy; 8) Department of Oncology and Haematology, University of Modena and Reggio Emilia, Modena, Italy; 9) Immunology and Molecular Oncology Unit, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy; 10) Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy; 11) Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy; 12) IFOM, The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy.

Male breast cancer (MBC) is a rare disease, whose etiology appears to be largely associated with genetic factors. Inherited high- and moderate- penetrance mutations in known predisposing genes, including *BRCA1*, *BRCA2* and *PALB2*, accounts for only 10-15% of all cases. Thus, a fraction of MBCs is expected to be caused by genetic risk factors yet to be identified. Recently, a case-control study showed that the *FANCM* c.5791C>T nonsense mutation (rs144567652) was associated with familial breast cancer (BC) risk, supporting the hypothesis that *FANCM* could be a novel moderate risk factor for the disease (Peterlongo P. *et al.*, *Hum Mol Genet.* 2015;24(18):5345-55). Interestingly, among the mutation positive families, one presented both female and male BC cases, suggesting that the c.5791C>T might be associated with MBC risk. Aimed at testing whether this mutation may explain a fraction of MBCs, we performed a case-control study including 489 *BRCA1/2* mutation negative MBCs, of which 33% had a first-degree family history of breast/ovarian cancer, and 800 controls from the ongoing Italian Multicenter Study on MBC. Genotyping of *FANCM* c.5791C>T mutation in peripheral-blood DNA samples was carried out using a custom TaqMan SNP genotyping assay. The *FANCM* c.5791C>T mutation was found in 1out of 800 controls (0.13%), a man aged 75 and with no personal history of cancer, and in none of the 489 MBC cases. Overall, our results, based on a country-based series, suggest that *FANCM* c.5791C>T mutation have not a major role in MBC predisposition in cases unselected for family history. Although based on the largest series of *BRCA1/2* mutation negative MBC reported to date, this study may be statistically underpowered to detect association between a rare mutation and risk of disease in MBC. Additional studies based on large MBC series from other countries, possibly enriched for familial cases, could clarify further the role of *FANCM* c.5791C>T mutation in MBC predisposition.   Study supported by *Italian Association for Cancer Research* (AIRC IG16933) to L.O.

2736F

Exome sequencing of oral squamous cell carcinoma reveals molecular subgroups and novel therapeutic strategies. *S. Su[1], W. Fan[1], W. Chung[1], S. Yang[2].* 1) Chang Gung Memorial Hospital, Keelung, Taiwan; 2) Chung Shan Medical University, Taichung, Taiwan.

Oral squamous cell carcinoma (OSCC), an epithelial malignancy affecting a variety of subsites in oral cavity, is prevalent in Southeast Asia. The survival rate of OSCC patients has not improved over the past decades due to its heterogeneous etiology, genetic aberrations, and treatment outcomes. Improvement in therapeutic strategies and tailored treatment options is an unmet need. To unveil the mutational spectrum, whole-exome sequencing of 120 OSCC from male individuals in Taiwan was conducted. Analyzing the contributions of the five mutational signatures extracted from the dataset of somatic variations identified four groups of tumors that are significantly associated with demographic and clinical parameters. In addition, known (*TP53, FAT1, EPHA2, CDKN2A, NOTCH1, CASP8, HRAS, RASA1,* and *PIK3CA*) and novel (*CHUK* and *ELAVL1*) genes that are significantly and frequently mutated in OSCC were discovered. Further analyses of mutation status with cancer progression revealed that *PIK3CA* mutations tend to appear at tumors with large size. Through defining the catalog of targetable genomic alterations (somatic mutations and copy number variations), 58% of the tumors were found to carry at least one aberrant event potentially aimed by US Food and Drug Administration (FDA)–approved agents. Strikingly, while targeting the P53-cell cycle pathway (*TP53* and *CCND1*) by the drugs studied in phase I-III clinical trials, those possibly actionable tumors are predominantly located at the tongue, suggesting that tumors at the specific anatomical site are likely more sensitive to current targeted therapies. Our work revealed molecular OSCC subgroups that reflect etiological and prognostic correlation as well as defined the extensive landscape of major altered events in the coding regions of OSCC genome. These findings provide insights into design of clinical trials for targeted therapies and genomically driven stratification of OSCC patients with differential therapeutic efficacy. .

2737W

Genomic analysis of breast cancer genes in the Women's Health Initiative: Creation of a public genomics resource. *T. Walsh[1], S. Gulsuner[1], M. Lee[1], S. Casadei[1], G. Anderson[2], J. Van den Akker[3], D. DeSloover[3], R. O'Connor[3], T. Sittler[3], E. Gil[3], M-C. King[3].* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Color Genomics, Burlingame, CA, 94010.

BRCA1, BRCA2, and multiple other genes are known to harbor mutations that significantly increase risk of breast cancer, with risks and allele frequencies that vary by gene and by mutation. A powerful way to distinguish pathogenic from benign variants in these genes is to characterize variant frequencies among appropriate controls; that is, among older women who have not developed breast cancer, then to compare variant frequencies between controls and breast cancer cases. Although public databases include exome sequences of many thousands of individuals, very few of those sequences are from older women known to be free of breast and ovarian cancer. We have sequenced germline DNA from 7000 European American (EA) and 3000 African American (AA) women who were participants in the Women's Health Initiative (WHI) and were at least 70 years old and had not had cancer as of last follow-up contact. Twenty-eight known and candidate breast and/or ovarian cancer genes were sequenced to high depth of coverage (>200x) and SNVs, indels, copy number variants, and complex rearrangements identified. Analysis of the first 4300 individuals reveals premature truncations, canonical splice site mutations, and experimentally validated loss-of-function missenses at a combined carrier frequency <0.001 for each of BRCA1 (one individual with M1775R), CHEK1, CTNNA1, GEN1, MRE11A, PALB2, RAD51B, RAD51C, RAD51D, RINT1, TP53 (one individual with R175C) and XRCC2; between 0.001 and 0.005 for each of ATM, ATR, BARD1, BRCA2, BRIP1, FAM175A, NBN, and SLX4; and >0.005 for each of CHEK2, FANCM, and RECQL. Thus far, unambiguously pathogenic mutations have not been identified in BAP1, CDH1, PTEN, or STK11. Our goal is to create a public resource of frequencies of all variants, pathogenic or benign, for both EA and AA populations, from these 10,000 cancer-free WHI participants. Investigators can then compare allele frequencies among these controls to frequencies among breast cancer patients in any EA or AA case series. The resource is also intended to provide investigators, health care providers, and consumers with background frequencies, among older cancer-free women, of variants detected in these genes by genetic testing from any source.

2738T

Synthetic lethality between *ATRX* mutations and *MYCN* amplification in neuroblastoma. *M. Zeineldin[1], J. Jeon[1], L. Griffith[1,2], S.M. Federico[1], X. Chen[1], J. Easton[1], J. Wu[1], S. Mao[1], Y. Liu[1], A. Naranjo[3], A. Pappo[1], M.D. Hogarty[4], M.A. Dyer[1,2].* 1) St. Jude Children's Research Hospital, Memphis, TN; 2) Howard Hughes Medical Insitute; 3) University of Florida, Gainesville, FL; 4) Children's Hospital of Philadelphia, Philadelphia, PA.

Neuroblastoma (NB) is a devastating pediatric tumor that is responsible for 15% of cancer related death in children. The chromatin remodeler ATRX is frequently mutated in NB and in a range of other pediatric and adult tumors. The tumorigenecity of *ATRX* mutations are not completely understood. Interestingly, *ATRX* mutations seem to be mutually exclusive with *MYCN* amplification; the later is also recurrently detected in NB. However, most of genetic studies of NB were conducted using a, relatively, small number of tumors. To study the genetic basis of NB and the mutual exclusivity between *ATRX* mutations and *MYCN* amplification, we did whole exome sequencing for 475 tumors and the matching normal tissues available from 353 patients from different age and clinical stage groups. In addition, intragenic *ATRX* deletions were examined by deep sequencing using custom-designed probe capture. We found *MYCN* amplification in 28 (5.9%) tumors and *ATRX* mutations in 19 (4%) other tumors, with no single tumor with both *MYCN* amplification and *ATRX* mutations. We also tested the mutual exclusivity between *ATRX* mutations and *MYCN* amplification in cultured cell lines. We found that induction of *MYCN* overexpression results in cell death and cell cycle changes in cell lines with *ATRX* mutations. We also detected reduced colony formation in two NB cell lines with *MYCN* amplification when ATRX was down-regulated using *ATRX*-specific *shRNA*. In addition, we tried to knock out *ATRX* in NB cell lines with MYCN amplification using CRISPR-Cas9 technology. In one of these cell line, IMR-32, no single clone with *ATRX* deletion was detected out of 142 clones screened. In another cell line, SK-N-BE(2), only one clone with *ATRX* mutation was detected out of 42 clones generated. Notably, in SK-N-BE(2) cells with CRISPR-induced *ATRX* mutations, the MYCN protein level was significantly reduced and the cells stopped growing and differentiated. We also mapped the epigenetic landscape of 8 different cell lines with and without *ATRX* mutations, 10 xenograft NB tumors orthotopically maintained in immune-compromised mice and a primary tumor with mutated *ATRX* using ChIP-Seq and bisulfite sequencing. Collectively, these data confirm synthetic lethality between *ATRX* mutations and *MYCN* amplification in NB cells and shed light on the molecular pathogenesis of *ATRX* mutations in NB. .

2739F

Absence of selection for somatic mutations in mitochondrial DNA of diffuse large B-cell lymphomas. *A.G.X. Zeng[1], A.C.Y Leung[2], A.R Brooks-Wilson[1,3].* 1) Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 2) Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada.

Diffuse Large B-Cell Lymphoma (DLBCL) is an aggressive and heterogeneous cancer. A subset of DLBCLs shows a gene expression pattern consistent with enhanced oxidative phosphorylation (oxphos) activity, suggesting that mitochondrial metabolism may play a role in this cancer. Mitochondrial DNA encodes subunits of oxphos complexes. Mitochondria have been linked to carcinogenesis primarily through aberrant production of reactive oxygen species (ROS). We characterized the mitochondrial genomes of 39 DLBCL cases using whole-genome data from the Cancer Genome Characterization Initiative, accessed through dbGaP. We analyzed germline variants not accounted for by haplogroup status (private variants) and somatic mutations present in tumor but not in peripheral blood DNA. Prevalence of heteroplasmies was assessed by comparing the heteroplasmic fractions (HF) of private variants (n = 137) and somatic mutations (n = 63) with a two sample t-test. Private variants were predominantly homoplasmic with some low-level heteroplasmies; somatic mutations were almost exclusively heteroplasmic (p < 0.001). While the frequency of G:C > T:A substitutions – the main consequence of ROS-mediated DNA damage – is higher in somatic mutations than private germline variants, they constituted only 8% of somatic mutations, suggesting that ROS plays a limited role in mtDNA mutagenesis in DLBCL. The mutational load of each functional region within the mitochondrial genome was compared using a chi-square test of homogeneity: private germline variants were more common within the D-Loop than expected by chance (p = 0.01), while the distribution of somatic mutations across the mitochondrial genome did not differ from that expected by chance. The HF of each heteroplasmy in the peripheral blood DNA was compared to the HF of its corresponding position in the tumor using paired t-tests. Heteroplasmic variants with a HF above 5% (n = 27) in peripheral blood showed a lower HF in the corresponding tumor (p < 0.001) and display a general trend towards fixation of the major allele regardless of whether it was a reference or variant allele, showing that changes in heteroplasmy during tumorigenesis appear random. In contrast with some studies on other cancers, we do not see evidence for positive selection for somatic mutations in DLBCL. Together, our observations suggest that somatic mutations in mitochondrial DNA are passengers rather than drivers, and that oxphos function is maintained in DLBCL.

2740W

Genetic characterization and prognostic implications of Korean patients with Lynch syndrome. *K. Park[1], J. Park[2], D. Chang[3], H. Kim[4], J. Kim[5].* 1) Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 2) Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Korea; 3) Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 4) Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 5) Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea.

　　Background: Lynch syndrome (LS) is caused by a genetic mutation that disrupts the mismatch repair (MMR) pathways. The overall survival in patients with LS seems to be different according to the mutation type in MMR genes. Currently, it is unclear whether the mutation type influences the tumor-free interval (TFI) in LS. The aims of this study were to investigate the mutation features and epidemiologic trends of LS in Korean and to evaluate the prognostic impacts of different mutation types. **Methods:** In total, 286 unrelated patients with cancers were enrolled from 341 consecutive individuals underwent a genetic workup (*MLH1, MSH2,* and *MSH6*) in the Cancer Genetics Clinic of the Samsung Medical Center in Korea from 2004 to 2015. Histopathologic findings and microsatellite instability data as well as clinical information such as tumor spectrum, family history and survival time were collected. Epidemiologic time trends of LS in Korea were evaluated. **Results:** A total of 118 patients were found to have mutations (41%). We identified 76 different mutations (47 in *MLH1*, 24 in *MSH2*, and 5 in *MSH6*) including 20 novel mutations and one founder mutation (c.1024delinsAA in *MSH2*). Truncation mutations such as frameshift and nonsense mutations were associated with shorter TFI compared to other mutation types (median of 3,129 days vs. 5,877 days, HR = 2.313, 95% CI = 1.101-4.858, *P* = 0.0396). There have been the epidemiologic changes in Korean LS patients between the period of 1995-2004 and that of 2004-2005; advanced age at diagnosis (*P* < 0.0001), presence of history of multiple tumor manifestations (*P* = 0.0019), increased mutation detection rate (*P* = 0.0022), and increased nonsense mutations and splicing mutations (*P* < 0.0001) **Conclusion:** This study shed light on the genetic features of Asian patients with LS as well as prognostic implications of different mutation types. This study will provide good guidance for genetic counseling and stratification of the surveillance program depending on mutation types.

2741T

Evaluation of copy number variants as modifiers of breast and ovarian cancer risk for *BRCA1* pathogenic variant carriers. *L.C. Walker[1], L. Marquart[2], J.F. Pearson[1], G.A.R. Wiggins[1], T.A. O'Mara[2], G. Chenevix-Trench[2], A.C. Antoniou[3], F.J. Couch[4], A.B. Spurdle[2], Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA).* 1) University of Otago, Christchurch, New Zealand; 2) QIMR Berghofer Medical Research Institute, Queensland, Australia; 3) University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; 4) Mayo Clinic, Rochester, Minnesota, USA.

　　Genome-wide studies of patients carrying pathogenic variants (mutations) in *BRCA1* or *BRCA2* have reported strong associations between single nucleotide polymorphisms (SNPs) and cancer risk. To conduct the first genome-wide association analysis of copy number variants (CNVs) with breast or ovarian cancer risk in a cohort of 2500 *BRCA1* pathogenic variant carriers, CNV discovery was performed using multiple calling algorithms and Illumina 610k SNP array data from a previously published genome-wide association study. Our analysis, which focused on functionally disruptive genomic deletions overlapping gene regions, identified a number of loci associated with risk of breast or ovarian cancer for *BRCA1* pathogenic variant carriers. Despite only including putative deletions called by at least two or more algorithms, detection of selected CNVs by ancillary molecular technologies only confirmed 40% of predicted common (>1% allele frequency) variants. These include four loci that were associated (P<0.05) with breast cancer (*GTF2H2, ZNF385B, NAALADL2, PSG5*), and two loci associated with ovarian cancer (*CYP2A7, OR2A1*). The most significant finding was an association of a validated CNV deletion at the *CYP2A7* locus (19q13.2) with decreased ovarian cancer risk (Relative Risk=0.50, P=0.007). Genomic analysis found this deletion coincides with a region displaying strong regulatory potential in ovarian tissue, but not in breast epithelial cells. This study provides evidence that CNVs can modify risk of breast or ovarian cancer in *BRCA1* pathogenic variant carriers.

2742F

Increased number in rare of germline variants in melanoma suscepti-bility genes in Spanish Giant Congential Melanocytic nevi patients. *C. Badenas[1], G. Tell[2], V.P. Martins da Silva[2], K. Jagirdar[3], M. Potrony[2], D.L. Duffy[4], H.P. Soyer[1,5], S. Puig[2], R.A. Sturm[3], A. Vicente[6], M.A. Gonzalez-Ensenyat[6], J. Malvehy[2], J.A. Puig-Butille[1].* 1) Molecular Genetics Service. Melanoma Units. CIBERER, Hospital Clinic de Barcelona, Barcelona, Spain; 2) Dermatology Service. Melanoma Units. CIBERER. Hospital Clinic de Barcelona, Barcelona, Spain; 3) Dermatology Research Centre, School of Medicine, The University of Queensland, Translational Research Institute, Brisbane, QLD, Australia; 4) QIMR-Berghofer Medical Research Institute, Brisbane, QLD, Australia; 5) Department of Dermatology, Princess Alexandra Hospital, Brisbane, QLD, Australia; 6) Clinic of Dermatology, Hospital Universitari Sant Joan de Déu, Barcelona, Spain.

   Introduction: Giant congenital melanocytic nevi (GCMN) are rare melano-cytic benign tumors associated with presence of neurocutaneous melanosis and higher risk of melanoma and other malignancies. The genetic cause of GCMN is likely to be the acquisition of postzygotic alterations. However, there are evidences of predisposition for the GCMN development. We hypothesized that germline variants in melanoma susceptibility genes (*MC1R, ATM* and *MITF*) have a role on this predisposition. **Materials and Methods**: Analysis of *MC1R* gene was performed in 21 patients by Sanger sequencing. Evaluation of ATM variants and p.E318-MITF variant was performed in 16 patients by CoreExome genotype platform (Illumina). *ATM* variants were evaluated by its ExAC Browser European frequency (EB-E freq) and its functional effect (Polyphen and Shift information) **Results**: *MC1R* variants were detected in 57.1% (12/21) GCMN patients. Statistically significant differences between patients and control population (N=296) in terms of number of alleles or type (Red hair vs non-Red hair alleles) were not detected. However, we observed a higher prevalence of patients harboring at least one *MC1R* variant (19%) compare to control population (9.8%). Moreover, 18.7% (3/16) GCMN patients carry rare *ATM* coding variants such as p.L1420F (EB-E freq= 0.019; benign/tolerated), p.V410A (EB-E freq= 0.003; benign/deleterious) and p.S333F (EB-E freq= 0.001; possibly damaging/deleterious). In contrast, the p.E318K-MITF was not detected in the set of GCMN patients. **Conclusions: t**he study suggests that GCMN patients have a higher incidence of *MC1R* variants and rare *ATM* variants. Study of a large number of patients is necessary to confirm such findings.

2743W

The role of CI-994, a histone deacetylase inhibitor, on the differentiation of acute myelogenous leukemia cell lines. *N. Bhambra[1,2], M.A. Blanco[2].* 1) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 2) Department of Newborn Medicine, Boston Children's Hospital, Boston, MA.

   Histone deacetylase (HDAC) inhibitors cause chromatin remodeling and inhibit the expression of otherwise highly expressed genes. HDAC inhibitors hold great promise as a cancer treatment due to their proven ability to induce differentiation in proliferative and cancerous cell lines. The induction of cellular differentiation in cancerous cells is of interest, because it reduces the ability of these cells to proliferate, and thus limits tumor growth. HDAC inhibitors and other drugs shown to induce cellular differentiation hold great promise as cancer therapies, as they cause significantly less cell death and are less toxic to patients than conventional chemotherapies. One such HDAC inhibitor is CI-994. Using flow cytometry and cell viability counts, our research shows that after treatment of human and mouse acute myelogenous leukemia (AML) cell lines with CI-994, cell populations show increased levels of cellular differen-tiation and a minimal reduction in cell viability. Some cell lines show a good balance between differentiation and cell viability. Although our results are not uniformly successful across all cell lines tested, our findings are promising. Our findings indicate that further research needs to be done on the exact mechanism of CI-994 in cellular differentiation. Future experiments will exam-ine the role of CI-994 on the remodeling of chromatin in these and other cell lines, and examine specifically the genes it may be suppressing.

2744T

Novel rare variants underpin prostate cancer susceptibility genome wide. *N. Emami[1], T. Hoffmann[2,3], J. Mefford[4], K. Lindquist[6], C. Cario[1], C. Tai[2], E. Wan[3], S. Wong[3], J. Gollub[6], A. Finn[6], D. Aaronson[7], J. Presti[7], L. Habel[8], J. Shan[8], D. Ranatunga[8], C. Chao[9], N. Ghai[9], E. Jorgenson[8], L. Sakoda[8], M. Kvale[3], P. Kwok[3], C. Schaefer[8], N. Risch[1,2,3,4,8], S. Van Den Eeden[5,8], J. Witte[1,2,3,4,5].* 1) Program in Biological and Medical Informatics, UCSF, San Francisco, CA; 2) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA; 3) Institute for Human Genetics, UCSF, San Francisco, CA; 4) Program in Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA; 5) Department of Urology, UCSF, San Francisco, CA; 6) Affymetrix, Inc., Santa Clara, CA; 7) Department of Urology, Kaiser Oakland Medical Center, Oakland, CA; 8) Division of Research, Kaiser Permanente Northern Califoria, Oakland, CA; 9) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena CA.

The impact of rare genetic variants on common disease susceptibility remains an area of wide interest and controversy. Although it has been hypothesized that rare variants with high effect sizes may contribute substantially to the missing heritability of complex human traits, the difficulty of tagging and confidently imputing rare variants has hindered the investigation of this hypothesis in genome wide association studies (GWAS). Furthermore, while next generation sequencing technologies have enabled us to bypass the challenge of rare variant imputation, the financial cost of sequencing seriously limits our ability to broadly pursue the discovery of trait-associated rare variants, which require many thousands of subjects to be sequenced in order to be detected at a genome-wide significance level. Previously we described the design of a custom Affymetrix Axiom microarray for directly assaying putatively functional rare variants and fine mapping loci previously associated with prostate cancer (PrCa), a highly heritable, common disease with appreciable missing heritability. Using our custom array, 416,047 markers were genotyped in a population of 8,480 prostate cancer cases and 8,128 controls nested within the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), ProHealth and California Men's Health Study. This approach revealed dozens of novel PrCa susceptibility variants, exonic, intronic, and intergenic, in the rare (MAF less than 1%) and intermediate (MAF greater than 1% and less than 5%) minor allele frequency ranges. Many of these variants have been uncovered at loci not previously associated with PrCa risk, including 3p14, 5q14, and 14q21. Furthermore, novel rare susceptibility variants were also discovered at previously reported PrCa loci, including 11q23 and 19q13, where the prostate-specific antigen gene KLK3 is located. Conditional analyses simultaneously modeling both novel rare variants and previously published GWAS variants suggest the presence of multiple independent signals of PrCa risk. While further study may clarify the extent to which rare variants drive heritability and synthetic associations at PrCa susceptibility loci genome-wide, our results represent preliminary evidence that rare variants may explain a significant fraction of the missing heritability for PrCa risk.

2745F

Whole exome sequencing in multigenerational mixed cancer families identifies a putative risk variant in the *PDIA2* gene. *R.M. Jones[1], P.E. Melton[1], A. Rea[1], E. Ingley[2], M.L. Ballinger[3], D.J. Wood[4], D.M. Thomas[3], E.K. Moses[1].* 1) The Curtin UWA Centre for Genetic Origins of Health and Disease, Faculty of Health Sciences, Curtin University and Faculty of Medicine, Dentistry & Health Sciences, The University of Western Australia; 2) Harry Perkins Institute of Medical Research, Western Australia; 3) The Kinghorn Cancer Centre, Garvan Institute of Medical Research, New South Wales, Australia; 4) School of Surgery, University of Western Australia.

The use of whole exome sequencing (WES) in families represents an optimal study design for the identification of rare genetic variants involved in the risk of cancer. We investigated 3 multi-generational cancer-cluster families from the International Sarcoma Kindred Study in order to identify putative deleterious exonic genetic variants that predispose individuals for sarcoma and other cancers. We performed WES using Ion Ampliseq Exome RDY at 100X on genomic DNA extracted from peripheral blood samples from 9 family members with cancer and 10 unaffected family members. Sequences were aligned to hg19 and called using the Torrent Variant Caller. We identified 94,449 SNVs across all 19 germline samples including 47,296 intronic, 19,295 missense (nonsynonymous, stop gain and stop loss), and 20,270 synonymous variants. To reduce multiple testing, variants from WES were filtered to prioritise those detected in 101 known cancer genes plus 25kb upstream and downstream (1,216 variants). These variants, along with pedigree and phenotype information, were imported into SOLAR and analysed using measured genotypes. We investigated all 3 families together and each family independently. Outcome variables used were sarcoma, cancer, age of onset of sarcoma, and age of onset of cancer. Our analysis revealed an association between a variant (rs45614840, c.1464C>G / p.T119R) in *PDIA2* at 16p13.3 and onset of all cancer patients in a single family, (p = 0.00028). This variant is only present in 4 family members with cancer (1 sarcoma, 1 prostate, 2 melanomas) and not found in any unaffected family members or members of the other families. Bioinformatic interrogation revealed that this variant has an allele frequency of 0.0479 in 1,000 Genomes and is located in exon 2 of *PDIA2*. Variants in *PDIA2* have been previously reported in 45 cancer cases in the Cancer Genome Atlas, with 12 in melanoma cases and the rs45614840 variant has been previously reported in two lymphoma cases. Our results demonstrate the first association of *PDIA2* with familial cancer. Our findings support the hypothesis that familial clustering of cancer is caused by heritable factors that can influence multiple cancer types.

2746W

Whole exome sequencing in 75 high-risk families with validation and replication in two case-control studies identifies three previously unknown prostate cancer susceptibility genes. *D.M. Karyadi[1], M. Geybels[2], E. Karlins[1,3], B. Decker[1], L. McIntosh[2], A. Hutchinson[3], S. Kolb[2], S.K. McDonnell[4], B. Hicks[3], S. Baheti[4], L.M. FitzGerald[5], M.S. DeRycke[6], S. Riska[4], M. Yeager[3], D.J. Schaid[4], S.J. Chanock[3], S.N. Thibodeau[6], S.I. Berndt[3], J.L. Stanford[2,7], E.A. Ostrander[1].* 1) Cancer Genetics and Comparative Genomics Branch, NHGRI/NIH, Bethesda, MD; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, MD; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 5) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia; 6) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 7) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA.

Prostate cancer (PCa) is genetically heterogeneous and predicted to be caused by a continuum from common, low-penetrant to rare, high-penetrant variants. While genome-wide association studies have been successful in defining at least 100 loci associated with PCa risk in Europeans, these loci likely represent common, low-penetrant variants and, to date, only a few moderately to highly penetrant PCa risk variants have been implicated and validated. To identify low-frequency variants with moderate penetrance, we leveraged whole exome sequence (WES) data with array-based SNP haplotypes from 75 hereditary PCa families. We performed rigorous disease model-based variant filtering, choosing variants with population frequency ≤ 2% that are predicted to be protein damaging, which reduced the candidate variants to 22,242. Selecting variants that are more frequent in the high-risk families versus the general population, and also segregate in at least three families reduced eligible variants to 998. We then varied three metrics: high-risk family frequency ratio, number of families segregating the variant, and average affected carrier frequency in order to identify the most compelling candidates for follow-up. Analysis of these 341 candidate risk variants in a population-based, case-control study of 2,495 men identified nine variants nominally associated with PCa risk with predicted protein consequences in *SWSAP1*, *HOXB13*, *D2HGDH*, *CHAD*, *EPHA8*, *TANGO2*, *BRD2*, *PPP6R2* and *OR5H14*. These nine variants were taken forward to test for replication in an independent nested case-control study of 7,121 men, which showed that two variants were associated with PCa in subset analyses. Meta-analysis combining the two case-control studies highlighted four variants that were significantly associated with risk, only one of which was previously identified. Co-occurrence of at least two of the variants increased risk by 5.47-fold (P = 0.02). Unlike other cancers, where high-risk families define one set of risk alleles and case-control studies another, our study clearly shows that PCa is unique; the complete mutation profile is a continuum. We conclude that well-defined, disease model-based filtering methods, followed by validation and replication in multiple case-control studies can successfully identify moderately penetrant mutations, which until now remained largely undiscovered.

2747T

A genome-wide association study of carcinoembryonic antigen (CEA) levels according to smoking status in Korean population. *S.M. Jeong, J.M. Yun, H.T. Kwon, J.H. Park, B. Cho.* Department of Family Medicine, Seoul National University Hospital, Seoul, South Korea.

Carcinoembryonic antigen (CEA) is known as an indicator of cancer progression and therapeutic responsiveness. Although high concentration of CEA is usually observed in people with several cancers, it is also known as being influenced by smoking through the mechanism increasing mRNA and protein expression of CEA in normal lung tissue compared with non smokers. There was a study in China that reported the significant association between four loci and plasma CEA concentration. We evaluated CEA-level associated single nucleotide polymorphisms (SNPs) by smoking status in Korean. We performed a genome-wide association study of CEA in 1,949 participants who underwent routine health check-ups at Seoul National University Hospital from December 2009 to December 2013 with 4,414,664 SNPs after genomic imputation. We analyzed the association between SNP and CEA level in current smokers as well as in total subjects. CEA level in current smoker (1.98 ± 1.00 ng/ml) was significantly higher than in non smoker (1.51 ± 0.68 ng/ml) (Mean ± SD). We identified two SNPs with significant association with CEA level after adjusting smoking status in all subjects. In accordance with previous study, rs8176749 at 9q34.2 in *ABO* gene was identified showing most significant association with CEA concentration in all subjects (p = 1.71×10^{-14}) and non smoker (p = 5.92×10^{-13}). The result suggests that ABO blood type may have different effects on CEA level. The ABO blood type is reported to be related to prognosis and survival of several malignancies. In current smokers, rs2140345 at 3q13.2 in *ZBTB20* gene was associated with high CEA level (p = 6.20×10^{-10}). *ZBTB20* gene is known to be associated with poor prognosis of hepatocellular carcinoma and cell proliferation in non-small cell lung cancer in previous reports. In conclusion, rs2140345 in current smokers was associated with increased levels of CEA in Korean adults. Our study suggests that smoking may "turn on" the effect of rs2140345 on serum CEA level by certain genetic mechanism.

2748F

Identification of novel genetic variant on *KLKP1* associated with prostate-specific antigen level: A genome-wide association study in the Korean population. *T.G. Yoo, B. Cho, J.H. Park, H.T. Kwon, J.M. Yun.* Department of Family Medicine, Seoul National University Hospital, Seoul, South Korea.

Measuring serum levels of prostate-specific antigen (PSA) is a useful screening test for prostate cancer in Korea. However, PSA levels may be influenced by several factors including genetic components, which lead to poor sensitivity and specificity for detecting prostate cancer. In this study, we aimed to conduct a genome-wide association study (GWAS) in the Korean population to identify genetic variation correlated with PSA levels. We performed a GWAS among 1936 Korean men, aged 20 or above, who had no medical history of cancer and underwent routine health examination at Seoul National University Hospital from December 2009 to December 2013. DNA samples were genotyped in Infinium HumanCore v1.0 BeadChip (Illumina, USA). After imputing genotypes, we were able to analyze a total of 4,414,664 single nucleotide polymorphisms (SNPs). Multivariable linear regression analysis was performed to investigate the association between genetic variants and log-transformed PSA levels, adjusted for age. We detected 10 SNPs on chromosome 1q32.1 (*NUCKS1*, *RAB29*, and *SLC41A1* locus) and 19q13.41 (*KLK3*, *KLK2*, and *KLKP1* locus) that showed statistically significant associations with $p \leq 1 \times 10^{-6}$. The SNPs that showed strongest signals in each region were rs823136 located on 1q32.1 ($p=4.65 \times 10^{-9}$) and rs1058205 located on 19q13.41 ($p=1.92 \times 10^{-15}$). For rs823126, PSA levels are 1.08 ± 0.02 ng/ml in major allele homozygote, 1.19 ± 0.04 ng/ml in heterozygote, and 1.92 ± 0.24 ng/ml in minor allele homozygote (mean±SE). For rs1058205, PSA levels are 1.26 ± 0.03 ng/ml in major allele homozygote, 1.09 ± 0.02 ng/ml in heterozygote, and 0.95 ± 0.04 ng/ml in minor allele homozygote (mean±SE). *KLK3*, encoding the PSA protein, has been proven to have strong association with PSA levels beyond ethnicity. *KLKP1*, which is expressed at lower levels in advanced prostate cancer compared to benign prostate hyperplasia, is a novel finding in our GWAS. We suggest that personalizing a cutoff value for PSA level, using the genetic markers identified in this study, could improve the validity of diagnosing prostate cancer.

2749W

Analysis of *cis*-eQTLs in normal and tumor-derived pancreatic tissues reveals functional insights, including for the 9q34.1/*ABO* pancreatic cancer risk locus. *L. Amundadottir[1], S. Lykke-Andersen[2], B. Zhu[3], W. Xiao[4], J.W. Hoskins[1], A. Jermusyk[1], X. Zhang[5], L. Rost[1], I. Collins[1], J. Jia[1], H. Parikh[1], T. Zhang[1], L. Song[3], B. Zhu[5], W. Zhou[5], G. Matters[7], R. Kurtz[8], M. Yeager[5], T. Jensen[2], K. Brown[1], W. Bamlet[9], S. Chanock[6], N. Chatterjee[10], B. Wolpin[11], J. Smith[12], S. Olson[13], G. Petersen[9], J. Shi[3], M. Zhang[1], TCGA Research Network.* 1) NIH/NCI/DCEG/LTG, National Institute of Health, Bethesda, MD; 2) Department of Molecular Biology and Genetics, Aarhus University, DK-8000 Aarhus, Denmark; 3) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA; 4) Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, FDA, Jefferson, Arkansas 72079, USA; 5) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland, USA; 6) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA; 7) Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, PA 17033, USA; 8) Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA; 9) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA; 10) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA; 11) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; 12) Division of Gastroenterology and Hepatology, Georgetown University Hospital, Washington, DC 20007, USA; 13) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.

Objective: To elucidate the genetic architecture of gene expression in pancreatic tissues. **Design**: We performed expression quantitative trait locus (eQTL) and allele specific expression (ASE) analyses using RNA-sequence data and 1000 Genomes (1000G) imputed GWAS genotypes from 95 fresh frozen histologically normal pancreatic tissue samples. Data from 115 pancreatic tumor-derived tissue samples from The Cancer Genome Atlas (TCGA) was included for comparison. **Results**: We identified 38,615 *cis*-eQTLs (corresponding to 484 Genes) in histologically normal tissues and 39,713 *cis*-eQTL (corresponding to 237 Genes) in tumor tissues (FDR<0.1), with the strongest effects seen near transcriptional start sites (TSS). Approximately 23% and 42% of genes with significant *cis*-eQTLs (eGenes) appeared to be specific for tumor and normal derived tissues, respectively. Significant enrichment of *cis*-eQTL variants was noted in noncoding regulatory regions marked by modified histones, DNAse hypersensitivity and bound transcription factors, in particular for pancreatic tissues (1.53-3.12 fold, $P \leq 0.0001$), indicating tissue-specific functional relevance. A common pancreatic cancer risk locus on 9q34.2 in the *ABO* gene (rs687289) was associated with *ABO* expression in histologically normal ($P=5.8 \times 10^{-8}$) and tumor-derived ($P=8.3 \times 10^{-5}$) pancreatic tissues. The high linkage disequilibrium (LD) between this variant and the O blood group generating deletion variant in exon 6 of *ABO* suggested that nonsense-mediated decay (NMD) of the "O" mRNA could explain the eQTL. However, knock-down of crucial NMD regulators did not influence decay of the *ABO* "O" mRNA, indicating that a gene regulatory element influenced by pancreatic cancer risk alleles may underlie the eQTL. **Conclusions**: We have identified *cis*-eQTLs representing potential functional regulatory variants in the pancreas and generated a rich dataset for further studies on gene expression and regulation in pancreatic tissues.

2750T

Partitioning of genetic variation for breast cancer using two genome-wide association studies containing genetically enriched cases.

O.C. Leavy[1], L. Palla[1], J. Peto[1], D.F. Easton[2], F. Dudbridge[1]. 1) Dept Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom; 2) Dept Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.

   Breast cancer is the most common type of cancer in women worldwide. Over recent years genome-wide association studies (GWAS) have identified over 70 individual genetic markers associated with breast cancer risk. However much of the heritability for this disease remains to be explained. Much of this remaining heritability may be explained by a collection of genetic markers that individually have a small effect on disease risk. Polygenic scoring has been used to examine whether a combination of genetic variants affect disease risk. We studied two GWAS datasets: the UK2 study and the British Breast Cancer Study (BBCS), both containing women who have at least two close relatives that have developed breast cancer with most of the cases in the BBCS dataset being bilateral breast cancer cases. Methods given by Dudbridge (2013) and Palla and Dudbridge (2015) can be used to estimate the genetic variance explained by an entire GWAS using polygenic scoring and information on internal training and replication datasets. We estimated the genetic variance explained to be between 16% and 21% on the liability scale for the two GWAS. Interestingly, the estimated genetic variance explained by markers related to breast cancer on the iCOGS array is 3%, with all markers on the array explaining 6%. These results indicate that much of the heritability can be explained by markers across the genome that are currently not reaching genome-wide significance. Partitioning the genetic variance by minor allele frequency (MAF), chromosome and SNP annotation can further improve our understanding of how genetic variation is spread across the genome. We use genome partitioning to show evidence of a linear relationship between the genetic variation explained by a chromosome and chromosome length. We show that, for both GWAS, a large proportion of the genetic variance (94%) is explained by common SNPs with MAFs > 0.1. The results from partitioning the genetic variance for each GWAS by SNP annotation suggest that proportionally the UK2 SNPs mapping to an intergenic region explain more genetic variance, whereas the BBCS SNPs mapping to gene regions explain more of the genetic variance. The results from this analysis give insight into the genetic architecture of breast cancer and provide evidence that, similarly to other complex diseases, a large proportion of the genetic variation of breast cancer can be explained by common markers.

2751F

Unilateral vestibular schwannoma in a patient with PIK3CA-related segmental overgrowth: Co-occurrence of mosaicism for two rare disorders.

J.R. Mills[1], A.M. Moyer[1], A.B. Poplawski[2], L. Messiaen[2], D. Babovic-Vuksanovic[1,3]. 1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Department of Genetics, University of Alabama at Birmingham Birmingham, AL; 3) Department of Clinical Genomics, Mayo Clinic.

   A 28-year-old female with a history of PIK3CA-related segmental overgrowth presented to Mayo Clinic Rochester with headaches and was found to have a 5.4 cm left-sided vestibular schwannoma, as well as three small meningiomas. At birth, she was large, had macrocephaly, and by 6-months of age she was noted to have generalized body asymmetry, with her right side being larger than her left. Her hemihypertrophy persisted and required multiple surgeries. She has no significant family history. She underwent surgical resection of the presumed unilateral vestibular schwannoma, which was pathologically confirmed.   According to the current NIH consensus diagnostic criteria for neurofibromatosis 2 (NF2), the presence of a unilateral vestibular schwannoma along with three meningiomas is sufficient for a clinical diagnosis of NF2. While prior genetic analysis of DNA extracted from blood revealed the presence of a mosaic heterozygous c.2740G>A (p.Gly914Arg) *PIK3CA* mutation present at ~10% allele frequency, confirming the diagnosis of PIK3CA-related overgrowth; however, no mutations in *NF2* were detected. Although vestibular schwannoma has not previously been reported in PIK3CA-related segmental overgrowth, meningiomas have been associated, raising the question of whether this patient's vestibular schwannoma and meningiomas represent coincidental NF2 or additional manifestations of her overgrowth syndrome. Therefore, we tested the vestibular schwannoma for *NF2* mutations.   In the vestibular schwannoma we identified a heterozygous *NF2* mutation c.784C>T (p.Arg262*) and partial loss of 22q, including *NF2*, *SMARCB1*, and *LZTR1* genes, confirming *NF2* involvement in the development of the vestibular schwannoma. No other *SMARCB1* or *LZTR1* mutations were identified. We also confirmed the presence of the *PIK3CA* (p.Gly194Arg) mutation in this tumor. This finding suggests that two hits in *NF2* are necessary for development of vestibular schwannoma, despite the functional overlap between the *NF2* tumor suppressor function and the PI3K-signaling pathway. While this patient's vestibular schwannoma most likely represents mosaic NF2, we cannot exclude the possibility that her meningiomas are a result of the *PIK3CA* mutation and the unilateral vestibular schwannoma represents a sporadic tumor. Identification of *NF2* mutations in this patient supports the clinical diagnosis of mosaic NF2 and highlights the possibility of co-occurrence of mosaicism for multiple rare disorders in a single patient. .

2752W

Pharmacogenomic prediction of anthracycline-induced cardiotoxicity.
F. Aminkeng[1], A. Bhavsar[2], S. Rassekh[3], M. Rieder[4], J. Stortz[4], F. Dionne[5], G. Groenweg[3], M. Hayden[1], B. Carleton[3], C. Ross[2], The Canadian Pharmacogenomics Network for Drug Safety Consortium. 1) Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada; 2) Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, British Columbia, Canada; 3) Pediatrics, The University of British Columbia, Vancouver, British Columbia, Canada; 4) Pediatrics, University of Western Ontario, London, ON, Canada; 5) Centre for Clinical Epidemiology and Evaluation, Vancouver, BC, Canada.

Background/Rationale Cancer is a leading cause of morbidity and mortality. Over 14.1 million people are diagnosed every year and it accounts for approximately 15% of all deaths. Anthracyclines are used to treat over 70% of childhood and 50-90% of adult breast cancer patients each year, and have significantly contributed to the improved 5-year survival rates to over 80%. Their use is limited by cardiac toxicity and congestive heart failure in 57% and 20% of treated patients, respectively. Uncovering the genetics and pathophysiology of anthracycline-induced cardiotoxicity (ACT) will inform treatment decisions and future drug development. *Methods* Patients were recruited and clinically characterized by the Canadian Pharmacogenomics Network for Drug Safety. *Results Pharmacogenomic Discovery*: We recently published the first GWAS (740 000 SNPs) of ACT and discovered a new gene (*RARG*) and a new variant (rs2229774, p.Ser427Leu) in children ($P = 5.9 \times 10^{-8}$, OR (95%CI) = 4.7 (2.7–8.3)). We showed that *RARG* transcriptionally regulates *Top2b* expression, an activity that is impaired by the rs2229774 variation, thereby providing new insights into the pathophysiology of ACT (*Aminkeng F et al. Nature Genetics. 2015 Sep; 47(9):1079-84*). We are analyzing the results of another GWAS (1.7 million SNPs) of ACT in 565 adult breast cancer patients. Preliminary results suggest a role of *RARG* rs2229774 in ACT in adult breast cancer patients (*Aminkeng F et al, Manuscript in preparation*). We have shown in a separate study that pharmacogenomic testing improves ACT prediction beyond currently implemented clinical risk factors and can inform treatment decisions (*Aminkeng F et al, Manuscript in preparation*). Implementation: We developed comprehensive, evidence-based, clinical practice guidelines for ACT pharmacogenomic screening (*Aminkeng F et al. Br J Clin Pharmacol. 2016 May 16. doi: 10.1111/bcp.13008*). We conducted a comprehensive health economic evaluation to examine the cost-effectiveness of ACT pharmacogenomic testing and showed that pharmacogenomics screening reduces morbidity and mortality and saves cost (*Dionne F, Aminkeng F et al, Manuscript in preparation*). We are currently implementing point-of-care genetic testing to enhance treatment decisions and 150 childhood cancer patients have been screened in British Columbia (*Manuscript in preparation*). *Conclusion* Pharmacogenomic prediction of anthracycline-induced cardiotoxicity informs oncology treatment decisions and is cost effective.

2753T

Novel genetic variants in carboxylesterase 1 predict early-onset capecitabine-related toxicity. *U. Amstutz[1], S. Hamzic[1], D. Kummer[1], S. Milesi[1], D. Mueller[1], S. Aebi[2], M. Joerger[3], C.R. Largiadèr[1].* 1) University Institute of Clinical Chemistry, Inselspital Bern University Hospital, University of Bern, Bern, Switzerland; 2) Division of Medical Oncology, Cantonal Hospital Lucerne, Lucerne, Switzerland; 3) Department of Medical Oncology and Hematology, Cantonal Hospital St. Gallen, St. Gallen, Switzerland.

Background: Capecitabine (Cp), an oral prodrug of 5-fluorouracil, is commonly prescribed to treat gastrointestinal and breast tumors. However, dose-limiting adverse effects occur in 20-35% of patients at standard doses, in particular hand-foot syndrome and diarrhea. The aim of this study was to evaluate, for the first time, the association of genetic variability in all enzymes of the Cp activation pathway with early-onset toxicity from Cp-based chemotherapy. **Patients & Methods:** The coding and exon-flanking regions of the cytidine deaminase gene (*CDA*) were sequenced in 144 Cp-treated patients, in whom Cp-related toxicities in the first two chemotherapy cycles were recorded. For the other investigated genes (*CES1*, *CES2*, *TYMP*, *UPP1*, and *UPP2*), sequencing of coding and exon-flanking regions was performed in a discovery subset of 48 patients (24 with severe Cp-related toxicity, 24 with no or mild toxicity), and associated candidate variants were subsequently genotyped in the full cohort. **Results:** We identified a haplotype in the carboxylesterase 1 gene (*CES1*) associated with Cp-related toxicity ($OR_{additive}$ = 2.3, 95% CI: 1.25-4.32, $P_{adjusted}$ = 0.008; $OR_{recessive}$ = 16.6, 95% CI: 2.78-98.7, $P_{adjusted}$ = 0.002). This common haplotype (frequency = 14%) encompassed five noncoding variants, including an expression quantitative trait locus (rs7187684) for *CES1*. In addition, the association of two common linked *CDA* promoter variants (c.1-451C>T: $OR_{dominant}$ = 4.29, 95% CI: 1.30-14.2, $P_{adjusted}$ = 0.017; and c.1-92A>G: $P_{adjusted}$ = 0.015; $OR_{dominant}$ = 4.40; 95% CI 1.34-14.5) with increased risk of Cp-induced diarrhea was replicated. For all other investigated genes, no significant association of genetic variants with Cp-related toxicities was detected. **Conclusions:** This is the first study to identify an association of genetic variation in *CES1* with Cp-related toxicity. Given that a variant (rs2244613) of the same *CES1* haplotype was previously associated with trough concentrations and bleeding from the CES1-metabolized anticoagulant dabigatran, this finding provides important evidence for the existence of a common regulatory *CES1* variant with possible clinical relevance for carboxylesterase-metabolized drugs..

2754F

Genetic variation in *SLC16A5* confers protection from cisplatin-induced ototoxicity in adult testicular cancer patients. *B.I. Drogemoller[1,2,13], J.G. Monzon[3,13], A.P. Bhavsar[1,2], A.E. Borrie[4,5], B. Brooks[6], G.E.B. Wright[2,7], G. Liu[8], E. Fadhel[8], D.J. Renouf[9], C.K. Kollmannsberger[9], P.L. Bedard[10], F. Aminkeng[2,7], C.A. Hildebrand[5], E.P. Gunaretnam[2,4], C. Critchley[11], Z. Chen[8], L.R. Brunham[4,12], M.R. Hayden[2,7], C.J.D. Ross[1,2,5,14], K.A. Gelmon[9,14], B.C. Carleton[4,5,14].* 1) Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 2) Child & Family Research Institute, Vancouver, BC, Canada; 3) Tom Baker Cancer Centre, Calgary, AB, Canada; 4) Department of Pediatrics, Division of Translational Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 5) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada; 6) Audiology and Speech Pathology Department, BC Children's Hospital, Vancouver, BC, Canada; 7) Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 8) Medical Oncology and Hematology, Department of Medicine, Princess Margaret Cancer Centre – University Health Network and University of Toronto, Toronto, ON, Canada; 9) BC Cancer Agency, Vancouver, BC, Canada; 10) Princess Margaret Cancer Centre and University of Toronto, Toronto, Ontario, Canada; 11) Neuro-Otology unit, Vancouver General Hospital, Vancouver, BC, Canada; 12) Department of Medicine, Centre for Heart Lung Innovation, University of British Columbia, Vancouver, British Columbia, Canada; 13) These authors contributed equally to this work; 14) These authors jointly supervised this work.

Background: Cisplatin, a chemotherapeutic agent used to treat many cancers, is highly effective in the treatment of testicular cancer. Cisplatin-induced ototoxicity (CIO), which occurs in 10-25% of adult patients, is an important complication that can affect testicular cancer survivors. Although clinical variables are associated with CIO, they do not completely explain the observed inter-individual variability and we hypothesize that pharmacogenomic variants also contribute to the occurrence of CIO.   Methods: Discovery (n=96, 23 cases and 73 controls) and replication cohorts (n=92, 14 cases and 78 controls) of adult testicular cancer patients previously treated with cisplatin were recruited and extensive clinical and demographic data were collected to facilitate case-control designation. All samples were genotyped for 7,907 variants using a custom pharmacogenomic genotyping panel. Logistic regression was performed to identify variants that were significantly associated with CIO and functional validation assays were utilized to substantiate these findings. Results: Association and fine-mapping analyses identified one significantly associated variant in *SLC16A5* (combined cohort: $P = 2.17 \times 10^{-7}$; OR = 0.06; 95% CI: 0.02-0.22) that conferred protection against CIO. Functional validation of this transporter gene revealed that *in vitro SLC16A5*-silencing significantly altered cellular responses to cisplatin treatment ($P < 0.0001$), thereby providing suggestive evidence of a role for *SLC16A5* in the development of CIO. Conclusions: This study has identified a strong association with a synonymous variant in *SLC16A5* and protection against CIO for the first time. This variant could be used to better predict which patients are at risk of CIO before therapy begins. These findings also provide insight into the molecular mechanisms of CIO in adult cancer patients with implications for potential otoprotectant strategies.

2755W

Exome sequencing of cancer related genes to identify genetic markers for sensitivity to cytotoxic anticancer drugs. *C. Udagawa[1], T. Tokino[2], Y. Sasaki[2], B. Shiotani[1], Y. Ohnishi[3], H. Suemizu[3], H. Zembutsu[1].* 1) National cancer center, Tokyo, Japan; 2) Sapporo Medical University School of Medicine, Sapporo, Japan; 3) Central Institute for Experimental Animals, Kawasaki, Japan.

Exome sequencing of 409 cancer related genes was conducted in 59 DNAs extracted from human tumor xenografts, of which sensitivity to nine cytotoxic anticancer drugs had been examined, in a nude mice model. We classified these xenografts into high (> 50%) and low (≤ 50%) variant allele frequency (VF) groups, and examined the difference in sensitivity to the anticancer drugs between the two groups using Welch's t test. The number of SNVs showing association with the anticancer drugs (P<0.05) were 211, 182, 167, 136, 126, 188, 179, 131 and 153 in MMC, CPM, ACNU, DDP, ADR, VCR, VLB, 5FU and MTX, respectively. 13 SNVs of them showed significant association (P<0.05) with five drugs or more, and one of them was located in a gene which is an essential component of NF-κB signaling. The most convincing candidate SNV was on a tumor suppressor gene, and xenografts with high VF of the SNV were more sensitive to ADR compared with xenografts with low VF ($P=8.66 \times 10^{-5}$). These data could contribute useful information for identifying biomarkers for drug sensitivity that may eventually provide "cancer precision medicine".

2756T

Estimating the joint distribution of rare variants, polygenic risk and family history to support analysis of the prospective benefits of risk-based mammographic screening. *M.C. Wolfson[1], S. Gribble[1], A. Antoniou[2], D. Easton[2], N. Pashayan[3], A. Lee[4], J. Simard[5]. 1) Epidemiology, University of Ottawa, Ottawa, Ontario, Canada; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, England; 3) Dept of Applied Health Research, University College London, London, England; 4) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Carre and Department of Oncology, University of Cambridge, Cambridge, UK; 5) Genomics Center, CHU de Québe, Université Laval Research Center, Quebec City, Quebec, Canada.*

The vast majority of breast cancer screening in Canada and other countries is based on a woman's age, e.g. starting age 50. However, women with some genotypes can be at high risk of breast cancer at earlier ages. Knowing genetic risk at an earlier age could enable organized screening programs cost-effectively to offer routine mammographic screening at earlier ages / higher frequencies to high risk women, and later ages / lower frequencies to low risk women. Multiple rare but moderate to high risk mutations for breast cancer, together with many more common single nucleotide polymorphisms (SNPs) conferring lower risks, have now been identified. The combined effects of multiple SNPs can be summarised into a polygenic risk score (PRS). These known variants explain ~45% of familial aggregation to breast cancer, so that detailed family history (FH) still provides substantial incremental information on risk.A significant health policy question is whether population-based assessment using such genetic information for risk-based rather than primarily age-based breast cancer screening would be warranted. However, no data exist for a representative population sample of the joint distribution of rare breast cancer genetic variants and PRS and FH.As part of the Genome Canada/CIHR/Genome Quebec-funded PERSPECTIVE project, a Genetic Mixing Model (GMM) has been developed to estimate this joint distribution. GMM is an interacting agent microsimulation model, simulating nuptiality, nulliparity, parity-specific fertility, and genetic inheritance for populations > 1M individuals, drawing on the underlying genetics of the BOADICEA algorithm (Lee et al. Genet Med (2016)). GMM is part of a larger simulation model to assess cost-effectiveness of risk stratified vs. current primarily age-based organized breast cancer screening programs in Canada.The GMM-simulated joint distribution of rare variants, PRS, and FH provides estimated proportions of women identified as high risk for breast cancer (e.g. >3x average for age). Of the three factors, the PRS identifies the most women at high-risk, followed by FH. In terms of clinical application for breast cancer risk stratification, these results suggest that SNP data are most important on a population basis, followed by FH. Testing for rare genetic variants like BRCA1/2, while important for younger women with a strong FH of breast cancer, are not most important in assessing genetic breast cancer risk on a realistic population basis.

2757F

Large-scale rare variants association studies in hereditary cutaneous and ocular melanoma uncover new risk gene with tumor suppressor activity. *M. Artomov[1,2,3], A.J. Stratigos[4], I. Kim[5], C.N. Njauw[6], K. Shannon[7], E.S. Gragoudas[5], A.M. Lane[5], M. Lauss[8], R. Kumar[8], B. Miao[6], G. Jonsson[8], M.J. Daly[1,2], H. Tsao[6,7]. 1) ATGU, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute, Cambridge, MA, USA; 3) Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA; 4) 1st Department of Dermatology, University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece; 5) Retina Service, Massachusetts Eye and Ear Infirmary, Boston, MA, USA; 6) Department of Dermatology, Wellman Center for Photomedicine, MGH, Boston, MA, USA; 7) Melanoma Genetics Program, MGH Cancer Center, MGH, Boston, MA, USA; 8) Department of Oncology, Clinical Sciences, Lund University, Lund, Sweden.*

Extraordinary progress has been made in our understanding of common variants in all aspects of medicine. However, the contribution of rare coding variants to inherited disease predisposition is not as well characterized. With the advent of routine whole exome sequencing, this next frontier has become more amenable to exploration. Here we performed an exome-wide, gene-based association study of familial cutaneous melanoma (CM; n=273) and ocular melanoma (OM; n=99) cases along with 7,629 ancestry-matched controls. We established a novel analytical framework for a large-scale exome-based rare variant association study in hereditary cancer cohort. Strong signals were detected for *CDKN2A* and *BAP1* in the CM and OM cohorts, respectively. In addition, multiple candidates exhibited some evidence of association (P<10[-4]); these were then subjected to replication with germline CM and OM sequences available for individuals in The Cancer Genome Atlas (TCGA) project. One of these novel secondary candidates, *EBF3*, was replicated in TCGA and then subjected to proof-of-concept functional credentialing and found to have properties of a *bona fide* tumor suppressor gene. Taken together, these results represent the largest rare variant germline association study in cancer to date, delineate the mutational landscape of hereditary cutaneous and ocular melanomas and implicate *EBF3* as a possible predisposition gene and a novel tumor suppressor in melanoma.

2758W

Linkage analyses reveal significant signals on multiple chromosomes for familial lung cancer. *J.E. Bailey-Wilson[1], A.M. Musolf[1], C.L. Simpson[1,2], M. de Andrade[3], D. Mandal[4], C. Gaba[5], P. Yang[3], M. You[6], E.Y. Kupert[6], M.W. Anderson[6], A.G. Schwartz[7], S.M. Pinney[8], C.I. Amos[9], Genetic Epidemiology of Lung Cancer Consortium.* 1) Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD; 2) University of Tennessee Health Science Center, Memphis, TN; 3) Mayo Clinic, Rochester, MN; 4) Louisiana State University Health Sciences Center, New Orleans, LA; 5) University of Toledo Dana Cancer Center, Toledo, OH; 6) Medical College of Wisconsin, Milwaukee, WI; 7) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 8) University of Cincinnati College of Medicine, Cincinnati, OH; 9) Geisel School of Medicine, Dartmouth College, Lebanon, NH.

Lung cancer has been the leading cancer killer of men since the 1950s and of women since 1987 and will kill an estimated 158,000 Americans in 2016. Lung cancer risk is known to increase with environmental exposures such as cigarette smoking; there is also a substantial genetic risk. We analyzed genotypes on about 240,000 SNPs (Illumina HumanCore-1 2v1-0 array) on 190 individuals from 26 extended families with a family history of lung cancer (Genetic Epidemiology of Lung Cancer Consortium). An affected-only model assuming autosomal dominant inheritance with 80% penetrance in carriers and 1% penetrance in non-carriers was used in three types of linkage analyses. Two-point linkage was performed using TwoPointLods; multipoint analyses were run using SIMWALK2 and regional-based linkage was run using SEQLinkage and MERLIN. SEQLinkage builds multiallelic regional haplotype-based markers (similar to microsatellites in information content) using rare variants (here MAF≤0.15) within a gene or a portion of a gene. Two-point linkage analyses were then performed on the regional markers using MERLIN. We found 7 regional markers (genes) with heterogeneity LOD (HLOD) scores that were genome-wide significant (HLOD≥3.4) on chromosomes 2, 3, 6p, 8, 16, 18, and 20. Four of the five genes with highest HLODs have been implicated in cancer, with three specifically in lung cancer. The highest HLOD (3.97) was on chromosome 18 at the PTPRM gene, which has been reported to control methylation patterns in pulmonary tumors. Both PTPRT (HLOD=3.8) and NRG1 (HLOD=3.5) have been found to be involved in lung cancer as well, and the CNTN6 gene (HLOD=3.87) has been implicated in familial prostate cancer. Two-point analyses were not genome-wide significant. Multipoint analyses showed a signal on chromosome 18 near PTPRM but no multipoint HLODs were genome-wide significant. This is most likely due to the fact that multipoint analyses require pruning of SNPs to remove LD, resulting in a very sparse map. The regional approach has been shown to control Type I error while improving power when using rare variants. Because both locus and allelic heterogeneity are quite common in familial cancers, it is reasonable to expect that different families contain different high risk variants. Here, we report significant linkage peaks to multiple chromosomes for familial lung cancer. Targeted sequencing of the regions of interest is needed to identify the actual source of a particular linkage peak.

2759T

Role of stochasticity in tumor formation in mice and humans with germline p53 mutations. *C. Chan[1,2], H. Ke[1], Y. Sun[1], A. Levine[3].* 1) Rutgers Cancer Institute of New Jersey, New Brunswick, NJ; 2) Department of Medicine, Rutgers Robert-Wood Johnson Medical School, New Brunswick, NJ; 3) Institute for Advanced Study, Princeton, NJ.

Germline p53 mutations predispose mice and humans to a high risk of getting a wide spectrum of tumors with a lifetime risk of getting cancer greater than ninety percent. Yet, genetically identical mice with the same p53 mutation will get tumors in a wide distribution of age as well as tumor types. We observe a similar disparate phenotype in a pair of human identical twins who have germline p53 mutation. We create a probabilistic model using epidemiological data to show how stochasticity can produce much of the heterogeneity in phenotypes. Moreover, the model provides some insights into the process of tumorigenesis with germline p53 mutations.

2760F

Targeted next-generation sequencing panel for detection of mutations in breast cancer. *Y. Chang[1], Y. Hsu[1], S. Su[2], H. Chen[3], Y. Huang[4], S. Yu[5], J. Chen[6], H. Chen[1], K. Li[1].* 1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, Taiwan; 2) Center of Biotechnology, National Taiwan University, Taipei, Taiwan; 3) Graduate Institute of Toxicology, College of Medicine, National Taiwan University, Taipei, Taiwan; 4) Department of Computer Science and Information Engineering, National Chung Cheng University, Taipei, Taiwan; 5) Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University, Taipei, Taiwan; 6) Institutes of Biomedical Sciences and Molecular Biology, National Chung Hsing University, Taichung, Taiwan.

The detection of mutations can improve our ability to detect high-risk patients who experience relapse and die early. There were 147 breast cancer patients sequenced by Raindance Hotspot targeted sequencing. The average age of all 147 breast cancer patients included in the study was 51.7 years, with a range of 21–77 years. Breast cancer samples categorized as ER positive, HER2 positive, PR positive and triple negative subtypes. The average sequencing depth is 5222X (range 2900x-8633x) and the coverage is around 100%. A total 235 variants in 43 genes were detected in 147 patients by high depth Illumina sequencing. 219 single nucleotide variations were found in 42 genes in from 147 patients and 16 indel mutations in 13 genes from 84 patients. After filtering the 1000 genome database and synonymous SNPs, we focused on the 54 somatic functional point mutations. The functional point mutations contained the 54 missense mutations in 22 genes. We found 37 novel functional mutations (filter by 1000 genome and dbSNP database) in 37 patients and 7 patients were with 2 novel functional mutations. Additionally, we performed the patient with the number of missense somatic mutations predicted the overall survival for the breast cancer patients (log-rank P = 0.036).

2761W

Whole-exome association study of colorectal cancer. *J. Chen[1], Y. Yu[1], F. Hu[1], H. Hu[1], E. Vilar[2], C. Huff[1].* 1) Epidemiology, UT MD Anderson Cancer Center, Houston, TX; 2) Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, TX.

Although familial studies have identified a number of high-risk colorectal cancer genes, the role of rare variants conferring an intermediate risk of colorectal cancer remains uncertain. To further evaluate the contribution of rare exonic variation to colorectal cancer susceptibility, we conducted a whole-exome case-control study among individuals of European ancestry involving 398 colorectal cancer cases from TCGA and 3507 controls from the National Database for Autism Research. We controlled for technological stratification resulting from heterogeneous sequencing platforms using the XQC toolkit and conducted gene-based association tests using VAAST2.1. Our results replicated five established colorectal cancer susceptibility genes at p <0.05: *MLH1, APC, POLD1, PMS1,* and *MLH3*. We also report effect size estimates for two classes of rare variants in each of these genes: Likely Gene-Disrupting (LGD) (nonsense, frame-shift and canonical splice-site mutations) and non-LGD variants predicted to be damaging by VAAST and PolyPhen-2. Our results yield insights into the role of rare non-LGD risk variants among established colorectal cancer susceptibility genes and provide a foundation for future exome association studies in colorectal cancer.

2762T

Search for genetic factor association with prostate specific antigen level based on genome-wide association study. *E.K. Choe[1], J.Y. Cho[2], Y. Lee[3], J. Lee[4], E.Y. Cho[4], J.Y. Lee[4], H.J. Choi[5].* 1) Surgery, Seoul National University Hospital, Healthcare System Gangnam Center, Seoul, South Korea; 2) Department of Radiology, Seoul National University Hospital, Healthcare System Gangnam Center, Seoul, Korea; 3) Seoul National University Hospital, Healthcare System Gangnam Center, Seoul, Korea; 4) DNA Link, Inc., Seoul, Korea; 5) Department of Anatomy, Seoul National University College of Medicine.

Background: Serum prostate specific antigen (PSA) level is a well-known tumor marker for prostate cancer as a screening tool. However, the genetic determinants of PSA level in healthy and prostate cancer population is not well known. We investigated the genetic markers associated with elevated serum sPSA level in these two populations and its clinical implications. Methods and Findings: Genome-wide association study (GWAS) was conducted in 4,124 healthy Korean male adults (discovery set 2841, replication set 1283) using the Affymetrix Axiom® Customized Biobank Genotyping Arrays for PSA levels with age adjustment. Subgroup analysis in population who had prostate biopsy for increased PSA level (n=65) were done. Results: We detected 11 genome-wide significant signals associated with PSA levels. The top SNP associated with log PSA levels was rs72434280 in SLC45A3 (p value, discovery set=2.97×10^{-20}, replication set = 7.31×10^{-10}). 4 SNPs (rs2735839, rs266849, rs16856139, rs10993994) that has been reported to be associated with PSA levels previously were validated in our population (P=7.57×10^{-20}, P=1.83×10^{-14}, P=1.35×10^{-11}, P=1.45×10^{-10}, respectively). With 11 Candidate SNPs associated with elevated PSA levels, case control study (50 biopsy negative vs 15 cancer proved) in PSA >3ng/ml group , whose biopsy reports were available, was performed after age adjustment. rs266849-G, rs2411332-A and rs2735839-A were associated with false negative increase in PSA levels (normal biopsy result as reference, Odds ratio (OR) =0.215, 95% Confidence interval (CI) = 0.049~0.956, P=0.04; OR=0.1252, 95% CI=0.027~0.572, P=0007; OR=0.054, 95% CI= 0.0008~0.383, P=0.003, respectively). However, the remaining 8 SNPs were not significantly associated with the biopsy result. Conclusions: We found new 7 candidate SNPs associated with elevated PSA levels in healthy males. 4 SNPs previously reported, were validated in Korean population. In males with increased PSA levels, genotyping rs266849 (19q13.33), rs2411332 (19q.33) and rs2735839 (19q.33, KLK3 gene) could be informative to decide the indication of prostate biopsy.

2763F

Comparative analysis of *ATRX* mutations in cancers and ATR-X syndrome. *A.S. Harutyunyan[1], S. Khazaei[1], K. Rossokhata[1], N. Ibrahim[1], M. Osmond[1], B. Krug[1], J. Majewski[1], N. Jabado[1,2].* 1) Department of Human Genetics, McGill University, Montreal, Quebec H3Z 2Z3, Canada; 2) Department of Pediatrics, McGill University, Montreal, Quebec H3Z 2Z3, Canada.

Mutations in the chromatin remodeler alpha-thalassemia mental retardation X-linked *(ATRX)* gene have recently been identified in gliomas and several other cancers. To better understand their contribution to cancer pathogenesis, we have analyzed *ATRX* mutational spectra in gliomas, other cancers and ATR-X syndrome using sequencing of pediatric gliomas combined with publicly available data (mainly from The Cancer Genome Atlas (TCGA) database). Comparison of somatic and germline variants identified distinct distributions of *ATRX* mutations. Mutations are most frequent in cancers of neural origin (glioma, neuroblastoma, pancreatic neuroendocrine tumors, pheochromocytoma, paraganglioma, adrenocortical carcinoma) and are mainly truncating, while other cancers and ATR-X syndrome predominantly carry missense mutations. Furthermore, the germline missense *ATRX* mutations mainly localize in two functional domains of the protein, ADD domain and ATPase/MecP2 domain, while somatic missense mutations in neural cancers cluster in the ATPase/MecP2 domain. This suggests that these domains play different functional roles in germline and somatic contexts. In gliomas, *ATRX* mutations define a molecular subtype of tumors and are strongly associated with somatic mutations in *TP53*, *IDH1/2* and *H3F3A*. In adult gliomas, the subgroup with *ATRX-TP53-IDH1/2* mutations markedly decreases in frequency with more advanced grades (46% in grade II vs. 30% in grade III vs. 4% in grade IV), while the association of the three mutations remains equally strong. In pediatric gliomas, *ATRX* mutations co-occur preferentially with *H3F3A* (encoding histone H3.3) mutations, particularly the G34R mutation, and are never present in histone H3.1 (*HIST1H3B*, *HIST1H3C*) mutant tumors. Our findings provide a better understanding of the type, location and role of *ATRX* mutations in cancers and ATR-X syndrome. The specificity of variants to particular domains of *ATRX* may provide insight into disease mechanisms. The results of the current study allow us to conclude that *ATRX* mutations have a prominent role in the pathogenesis of neural cancers, while their functional role in other tumors is questionable and the observed rare missense mutations are most likely "passenger" mutations.

2764W

Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. *C. He[1,4,7], W. Wu[1], E.K. Wagner[1], Y. Hao[2], X. Rao[2], H. Dai[1], J. Han[1,3,4], J. Chen[5], A.M.V. Storniolo[6], Y. Liu[2,7].* 1) Department of Epidemiology, Indiana Univ Richard M. Fairbanks School of Public Health, Indianapolis, IN; 2) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA; 4) Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN; 5) Spinal Cord and Brain Injury Research Group, Department of Neurosurgery, Stark Neuroscience Research Institute, Indiana University, Indianapolis, IN; 6) Susan G. Komen Tissue Bank at the Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN; 7) Center for Computational Biology and Bioinformatics, Indiana University, Indianapolis, IN.

   Background: Increasing evidence suggests that long non-coding RNAs (lncRNAs) have key regulatory functions in chromatin remodeling and gene expression, and a few dysregulated lncRNAs have been linked to breast cancer. However, inference of the biological and pathological roles of lncRNAs in breast cancer development remains a challenge. **Results:** We analyzed RNA-sequencing (RNA-seq) data in tumor and normal breast tissue samples from 18 breast cancer patients and 18 demographically matched healthy controls, respectively. We identified 598 lncRNAs and 2,980 mRNAs that were differentially expressed between tumor and normal breast tissue. Genome-wide co-expression analysis of lncRNAs and mRNAs revealed two distinctive correlation patterns associated with breast cancer, reflecting different underlying regulatory mechanisms: 1) 516 pairs of lncRNAs –mRNAs have differential co-expression patterns, in which the correlation between lncRNA and mRNA expression differs in tumor and normal breast tissue; 2) 291 pairs of lncRNAs – mRNAs have dose-response co-expression patterns, in which the correlation is similar in tumor and normal breast tissue, but the expression level of lncRNA or mRNA differs in the two tissue types. We further validated our findings in an additional dataset of 744 breast tumors and 104 adjacent normal tissues from TCGA and annotated the function of lncRNAs using TANRIC. One novel lncRNA, *AC145110.1* on the 8p12 region, was found differentially co-expressed with 127 mRNAs in tumor and normal breast tissue. This lncRNA was also found to harbor a known breast cancer risk locus rs9693444, and was highly correlated with clinical outcomes of breast cancer. Functional enrichment and pathway analyses identified distinct biological functions for different patterns of co-expression regulations. **Conclusions:** Using normal breast tissue from healthy women as a desirable baseline for the first time, we identified breast cancer-associated lncRNAs and further explored their potential roles on gene regulation in a genome-wide co-expression analysis. Our data suggested that lncRNAs might be involved in breast tumorigenesis through the modulation of gene expression in multiple pathologic pathways.

2765T

Analysis of shared heritability and functional enrichment based on genome-wide association studies for five common cancers. *X. Jiang[1], H. Finucane[1,2], F. Schumacher[3], A. Price[1], D. Hunter[1], D. Conti[3], G. Casey[3], T. Sellers[4], D. Easton[5], S. Gruber[3], M. Jerkins[6], R. Eeles[7], Z. Kote-Jarai[7], S. Benlloch[7], A. Olama[7], C. Haiman[3], P. Pharoah[5], C. Amos[8], P. Kraft[1], S. Lindstrom[9] for BCAC, CORECT/Colon CFR, TRICL, OCAC and PRACTICAL/ELLIPSE.* 1) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA; 3) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL; 5) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge; 6) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population and Global Health, University of Melbourne, Australia; 7) The Institute of Cancer Research; 8) Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH; 9) Department of Epidemiology, University of Washington.

   Breast, colorectal, lung, ovarian and prostate cancer constitute more than 50% of all annual new cancer diagnoses in the US. These cancers have a significant inherited component but the extent to which their inherited basis is shared is not known. We used GWAS summary statistics from over 560,000 individuals of European ancestry (279,000 cases, 281,000 controls) across these five cancers, generated through the OncoArray initiative. We applied LD score regression to estimate pairwise genetic correlation among all five cancers. In addition, we partitioned the heritability of each cancer by functional annotations, through: 1) a baseline model using 52 publicly available main annotations that are not specific to any cell types; 2) a cell-type specific model including 220 cell-type specific annotations for four histone marks. We identified significant genetic correlations between breast cancer and lung (rg=0.17, se=0.05), colorectal (rg=0.15, se=0.06) and prostate cancer (rg=0.08, se=0.03), but not between any other pairs of cancer types. Functional enrichment analysis across the five cancers revealed that conserved regions are highly enriched, with 2.6% of SNPs explaining an average of 34.5% of the heritability (12.7-fold enrichment, 95%CI: 9.4-16.0). In addition, coding regions, transcription start sites, transcription factor binding sites and DNase I hypersensitive sites showed strong enrichments across multiple cancer types. In contrast, repressed regions exhibited depletion for all five cancers (0.29, 95%CI: 0.01-0.56). Cell-type specific analysis using histone marks revealed similarities between breast, colorectal and prostate cancer, where we observed high enrichments in multiple tissue types including breast and colon. We did not observe any statistically significant cell type-specific results for either lung or ovarian cancer. Our results indicate that solid tumors arising from different tissues share some common germline genetic influences. In addition, these analyses highlight conserved, coding and regulatory regions as functional components of the genome that are significant contributors to cancer heritability across multiple cancer types.

2766F

Cross-cancer pleiotropic analysis of lung cancer in African Americans.

C.C. Jones[1,2], Y. Bradford[3], C.I. Amos[4], W.J. Blot[5,6], S.J. Chanock[7], C.C. Harris[4], A.G. Schwartz[8], J.K. Wiencke[9], X. Wu[10], M.C. Aldrich[1,2,5]. 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 3) Biomedical and Translational Informatics Program, Geisinger Health System, Danville, PA; 4) Department of Biomedical Data Science, Geisel School of Medicine, Lebanon, NH; 5) Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN; 6) International Epidemiology Institute, Rockville, MD; 7) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 8) Department of Oncology, Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI; 9) Department of Neurological Surgery, University of California, San Francisco, CA; 10) Department of Epidemiology, Division of OVP, Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX.

Lung cancer (LC) is one of the most commonly diagnosed cancers in the United States. Genome-wide association studies (GWAS) have successfully identified several loci associated with LC risk. However, despite higher incidence and mortality rates in African Americans, few LC GWAS have been conducted in African Americans. Given the high burden of multiple testing intrinsic to GWAS, studies have shifted toward examining variants selected based on biological plausibility. One such method of selecting variants involves the investigation of pleiotropy, which occurs when a single variant or locus affects more than one phenotype. Here, we sought to examine the pleiotropic effects of SNPs previously associated with other cancers with lung cancer risk. African American lung cancer cases (N=1,410) and controls (N=2,843) from 5 independent studies were genotyped on the Illumina 1M-Duo array and standard quality control filters were applied; imputation was performed using IMPUTE2 and the 1000G cosmopolitan reference panel (June 2014). SNPs were extracted from the National Human Genome Research Institute - European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog based on the search term "neoplasm". Manual review filtered out any study in which the outcome was not cancer risk, resulting in 984 SNPs from 220 studies. We examined 916 SNPs, of which 629 were directly genotyped and 287 were imputed (info score >0.8, minor allele frequency >0.01). For each SNP, we performed logistic regression using SNPTest, adjusting for age at diagnosis, sex, smoking status, global African ancestry, and study site. Two SNPs, rs2853677 (in *TERT* on chromosome 5p15) and rs1051730 (in *CHRNA3* on chromosome 15q25), were significantly associated with LC risk (odds ratio (OR) = 1.27, 95% confidence interval (CI): 1.13-1.41 and OR = 1.37, 95% CI: 1.18-1.60, respectively) applying a false discovery rate of 5%. Both SNPs have been previously associated with LC risk in African Americans. The current analysis examined only the variants directly reported in the NHGRI-EBI GWAS catalog. However, because linkage disequilibrium (LD) patterns are known to differ between racial/ethnic populations, it is possible that SNPs discovered in European or Asian populations do not tag the causative SNP in African Americans. Consideration of LD differences between populations will allow for the discovery of associations not captured by examining the reported SNP alone.

2767W

Integrative Bayesian network analysis of prostate tissue lincRNA-mRNA expression profiles reveals potential regulatory mechanisms of prostate cancer risk loci.

N.B. Larson, S. McDonnell, Z. Fogarty, M. Larson, S. Riska, S. Baheti, A.A. Nair, D.J. Schaid, S.N. Thibodeau. Mayo Clinic, Rochester, MN.

Large-scale GWAS have identified 107 loci associated with risk of developing prostate cancer (PRCA). However, most of these loci do not lie in close proximity to protein coding genes and are presumed to be regulatory in nature. Downstream regulation of protein coding genes related to PRCA development may be mediated by *cis*-acting transcriptional regulation of nearby long intergenic non-coding RNAs (lincRNAs), which correspond to a host of regulatory functions such as chromatin remodeling and transcriptional co-activation. LincRNAs have previously been identified as diagnostic and prognostic biomarkers for a number of cancers, although their role in cancer development and progression is poorly understood. To explore this hypothesis, we leveraged an eQTL dataset of 471 samples of normal prostate tissue from prostate/bladder cancer patients with available RNA-Seq and imputed Illumina Infinium 2.5M genotype data. We first conducted an initial *cis*-eQTL screening of 1,169 lincRNAs within 1.1 Mb of 8,073 SNPs in high linkage disequilibrium (r^2>0.5) with previously identified PRCA risk-associated variants, identifying 173 significant lincRNAs (FDR<0.05) to be putatively *cis*-regulated. We then constructed an undirected Bayesian regulatory network from the expression profiles of the cis-regulated lincRNAs along with a total of 17,154 expressed protein-coding mRNAs using semi-interleaved HITON-PC, identifying 24,006 connections between 12,930 transcripts (156 lincRNAs, 12,774 mRNAs). Examination of 158 unique lincRNA-neighboring mRNAs within the regulatory network revealed significant enrichment for Gene Ontology terms related to tissue development and morphogenesis, consistent with previous functional studies of lincRNAs. GLAD4U disease gene-set analysis also revealed enrichment for male urogenital disease genes. Previous studies have indicated potential *cis*-acting transcriptional mechanisms for lincRNAs on neighboring mRNAs; however, 65% (114/176) of identified lincRNA-mRNA connections involved transcript pairs on different chromosomes. One negatively correlated *trans* connection between lincRNA CTD-2194D22.2 and putative oncogene *CAPRIN2* [MIM: 610375] was largely explained by *cis*-associated SNP rs12655062. Notably, CTD-2194D22.2 was also connected to upstream PRCA gene *IRX4* [MIM: 606199], which has previously been shown to be *cis*-regulated by rs12655062. Our findings suggest dysregulation of lincRNA expression may play a role in genetic predisposition to PRCA.

2768T

Admixture and association mapping in African Americans identifies susceptibility loci for prostate cancer on chromosome 7. *A.M. Levin[1], G. Dyson[2], B. Pasaniuc[3], J.J. Ruterbusch[2], N. Kubinec[2], T.J. Hoffmann[4], S.K. Van Den Eeden[5], S.A. Ingles[6,7], S.S. Strom[8], B. Nemesure[9], W.B. Isaacs[10], J.L. Stanford[11,12], W. Zheng[13], J. Xu[14], S.M. Gapstur[15], A.J.M. Hennis[9,16], E.A. Klein[17], S.J. Chanock[18], W.J. Blot[13,19], D.V. Conti[6], S. Watya[20], J. Boerner[2], C.A. Haiman[6,7], J.S. Witte[4], R.A. Kittles[21], B.A. Rybicki[1], C.H. Bock[2], African Ancestry Prostate Cancer (AAPC) GWAS Consortium.* 1) Public Health Sciences, Henry Ford Hlth System, Detroit, MI; 2) Karmanos Cancer Institute, Oncology Department, Wayne State University School of Medicine, 4100 John R, Detroit, MI 48201; 3) Pathology and Laboratory Medicine and Human Genetics Departments, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California; 4) Department of Epidemiology and Biostatistics, Institute for Human Genetics, University of California, San Francisco School of Medicine; 5) Division of Research, Kaiser Permanente Northern California; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; 7) Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA; 8) Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230, USA; 9) Department of Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA; 10) James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institution, Baltimore, MD 21287, USA; 11) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; 12) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA 98195, USA; 13) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; 14) Program for Personalized Cancer Care and Department of Surgery, NorthShore University Health System, Evanston, IL 60201, USA; 15) Epidemiology Research Program, American Cancer Society, Atlanta, GA 30303, USA; 16) Chronic Disease Research Centre and Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados; 17) Glickman Urological & Kidney Institute, Cleveland Clinic, Cleveland, OH 44195, USA; 18) Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute Bethesda, MD, 20892; 19) Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; 20) School of Public Health, Makerere University College of Health Sciences, Kampala Uganda; 21) Department of Surgery, University of Arizona College of Medicine, Tucson, AZ 85724, USA.

Purpose: The racial disparity in prostate cancer incidence in the United States is well documented, with African American (AA) men at highest risk of disease. To identify germline genetic variation that could explain some of this disparity, our previous admixture mapping study identified a novel region on chromosome 7 where local ancestry was associated with risk of disease. Association mapping that combines both single nucleotide polymorphism (SNP) genotype and local ancestry information was used to further identify prostate cancer susceptibility variants in this region that explain the admixture association, and independent data sets were interrogated for validation. Methods: Genotype data from 420 AA prostate cancer cases and 244 AA controls for the two studies (Henry Ford Health System, Detroit, MI, & Howard University, Washington, DC) that were part of original admixture mapping study were obtained from the African Ancestry Prostate Cancer GWAS consortium (AAPC). Admixture linkage was confirmed using an affected-only test and association mapping within the region was conducted using the MIX-score method to identify the SNP(s) most likely to explain the admixture linkage. Validation testing was conducted in the remainder of the AAPC subjects and an independent sample of African Americans from a recent Kaiser Permanente prostate cancer GWAS (KP). Results: The admixture linkage signal spanned a 25 Mb region from 7q31.1 (112.27Mb) to 7q34 (137.90Mb). Within this region, there were four SNPs with MIX-score p-values less than the Bonferroni corrected significance threshold. The highest level of significance was attained at SNP rs13221538 (p=4.11*10-8), located in an intron of the cadherin-like and PC-esterase domain-containing protein 1 (*CPED1*, not characterized in OMIM) gene and the minor allele was associated with a lower risk. In the validation studies, rs13221538 showed suggestive evidence for validation (p=0.053) but with an association estimate in the opposite direction from the discovery sample. An imputation-based meta-analysis of the AAPC and KP studies revealed an additional low frequency variant (rs73228923; p=5.68*10-7) within the glutamate receptor, metabotropic, 8 (*GRM8* [MIM 601116]) gene, with consistent associations in both AAPC and KP studies. Conclusions: Follow-up association mapping of a novel admixture region on chromosome 7 has identified new variants associated with prostate cancer risk.

2769F

Multiple-gene research and its clinical evaluation in Chinese high-risk breast cancer. *G. Li[1,2], X. Zhang[3], X. Wang[1], K. Shao[1], J. Xie[3], B. Xia[3], Y. Zhao[4], B. Cao[1], J. Geng[5], F. Liu[3], S. Zhu[1,2], M. Dean[6], D. Pang[1,7], Y. Hou[1,2].* 1) BGI-Research, BGI-Shenzhen, China; 2) Department of Biology, University of Copenhagen, Denmark; 3) Department of Breast Surgery, Harbin Medical University Cancer Hospital, China;; 4) Department of Epidemiology, Public Health College of Harbin Medical University, China; 5) 5 Department of Pathology, Harbin Medical University Cancer Hospital, China; 6) National Cancer Institute, USA;; 7) Translational Medicine Research and Cooperation Center of Northern China, Heilongjiang Academy of Medical Sciences, China.

Background: Currently, dozens of susceptibility genes have been defined for breast cancer. Though several researches have proved the clinical utility of multiple-gene sequencing in assessment of people with hereditary cancer risk, little clinical approach is available. Especially, there is limited such research in China, which country counts for approximately one fifth of world population. Here, we use a multiple-gene sequencing assessment to identify the variant spectrum and evaluate the clinical utility of this assessment in Chinese with high risk of breast cancer. **Method:** Breast cancer patients, who meet the enrollment criteria of NCCN guidelines of breast and/or ovarian cancer genetic assessment were informed and involved in this research. DNA was extracted from patient's peripheral blood. The coding regions of 115 hereditary cancer susceptibility genes were captured and sequenced using next generation sequencing platform. Single Nucleotide Polymorphism and small Insertion/deletion variants were defined and analyzed. Variants were classified into pathogenic, variants of uncertain significance (VUS) and benign according to ACMG recommendation for interpretation of sequence variants. VUS which locate in BRCA1 RING and BRCT domain had functional analysis. **Result:** Totally, 586 breast cancer patients participated in the research. From which, 121(20.6%) patients carried pathogenic variants, including 61(10.4%) with *BRCA1/2* pathogenic variants only, 53(9.0%) with other genes pathogenic variants only and 7(1.2%) with *BRCA1/2* and other gene pathogenic variants simultaneously, meanwhile, there were 29.6% and 73.1% novel spots in 27 BRCA1 and 33 *BRCA2* pathogenic variant spots respectively. 1163 VUS spots from 115 hereditary cancer susceptibility genes were identified in 586 patients. There were 70(11.9%) patients carried *BRCA1/2* VUS, 5 BRCA1 VUS had done functional analysis, and 1 was identified as pathogenic. Additionally, other gene pathogenic variant carriers in 11 families accepted clinical screening such as breast ultrasound, gastroscopy, fibercolonoscopy and serum tumor marker detection. Positive gastroenteric precancerous lesions have been observed in 3 families with *ATM*, *PMS2*, *ATM* and *PALB2* pathogenic variants respectively. **Conclusion**: Overall, based on this research, we have revealed a comprehensive susceptibility variations spectrum and firstly evaluated clinical utility of a multiple-gene sequencing assessment in Chinese with high risk of breast cancer.

2770W

Assessing hereditary risk of breast cancer - Effectiveness of current clinical guidelines. *L.A. Littlejohn[1,3], J. Gibb[2], D. Goudie[3], J. Berg[1,3].* 1) Clinical Genetics, University of Dundee, Dundee, Angus, United Kingdom; 2) Department of Pathology, Ninewells Hospital and Medical School, Dundee, Angus, United Kingdom; 3) Department of Clinical Genetics, Ninewells Hospital and Medical School, Dundee, Angus, United Kingdom.

Family history of breast cancer is one of the most common indications for referral to Clinical Genetic Services. Currently, UK National Institute of Health and Care Excellence (NICE) guidelines use family history to stratify by absolute risk for developing breast cancer between age 40 and 50. Patients are divided into low/population risk (LR, 10yr risk <3%), medium risk (MR, 10yr risk 3-8%) and high risk (HR, 10yr risk >8%). Risk stratification is separate from *BRCA1 & 2* mutation status. Women at increased cancer risk are offered screening at or prior to age 40. Accurate stratification of risk is essential for effective patient management. Family history data was obtained for all unaffected women with a family history of breast cancer aged ≤50, referred to Tayside clinical genetics between 2000 and 2010. All patients were risk stratified *de novo* according to NICE criteria. Patients who developed breast cancer were identified. Known *BRCA1 & 2* mutation carriers were excluded from initial analysis. 1410 women had 15,414 patient-years of follow up. 30 invasive breast cancers developed in the group, 13 in MR and 13 in HR women. Excluding *BRCA* carriers, more cancers developed in MR and HR groups. LR vs MR: *p*=0.048 (Fisher's exact test) and LR vs HR: *p*=0.029. Inclusion of *BRCA* carriers led to a greater degree of significance, (*p*=0.003 and 0.007 respectively). Excluding *BRCA* carriers, Kaplan-Meier analysis demonstrated the highest rate of breast cancer between ages 40-49 in the HR group followed by MR; log-rank scores: LR vs MR *p*=0.025, LR vs HR *p*=0.021. There was no significant difference in the rate of cancers between MR and HR groups (Log rank *p*=0.840). Combining the MR+HR groups for invasive disease gave a sensitivity of 86.7% and specificity of 63.6%, with a positive predictive value of 2.88%. The 10 year incidences of invasive cancer between 40-49 in MR and HR patients were 2.4% (0.04-0.44%) and 3.6% (0.7-6.5%) repectively. NICE guidelines appear to identify patients at increased risk of breast cancer aged 40-49 with reasonable sensitivity. Discrimination between moderate risk and high risk groups was poor. The 10-year incidence of breast cancer observed in this group did not meet the values suggested by NICE for clinical benefit and cost effectiveness. Improving discrimination between those at MR and HR will help to better target screening resources. Family history criteria required to meet these thresholds may need to be revised.

2771T

Common and rare variants in the exons and regulatory regions of known loci are associated with lung cancer risk. *Y. Liu[1], C. Lusk[2], M. Cho[3], R. Zhang[4], M. Scheurer[1], F. Kheradmand[5], D. Qiao[3], K. Covington[6], D. Wheeler[6], S. Tsavachidis[1], G. Armstrong[1], C. Chow[7], C. Behrens[8], C. Pikeleny[9], I. Wistuba[7], C. Amos[9], D. Christiani[4,10], E. Silverman[3], A. Schwartz[2], M. Spitz[1], The Genetic Epidemiology of Lung Cancer Consortium, The COPDGene Investigators.* 1) BAYLOR COLLEGE OF MEDICINE, HOUSTON, TX., USA; 2) Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and women's Hospital and Harvard Medical School, Boston, MA , USA; 4) Harvard T.H. Chan School of Public Health, Harvard Medical School, Boston, MA, USA; 5) Department of Medicine, Michael E. DeBakey Veterans Affairs Medical Center, Baylor College of Medicine, Houston, TX, USA; 6) Human Genome Sequence Center, Baylor College of Medicine, Houston, TX , USA; 7) Department of Translational Molecular Pathology, UT MD Anderson Cancer Center, Houston, TX, USA; 8) Department of Thoracic/Head and Neck Medical Center, MD Anderson Cancer Center, Houston, TX, USA; 9) Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA; 10) Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

Background: Genome-wide association studies (GWAS) are widely used to map genomic regions contributing to lung cancer (LC) susceptibility but they typically do not identify the precise disease-causing genes/variants. To identify causative LC variants, we focused on common and rare functional variants in exon and regulatory regions of 109 GWAS loci associated with LC, chronic obstructive pulmonary disease (COPD), lung function and smoking behavior. **Methods:** Using an extreme-phenotype approach, we selected: resistant-smokers with normal lung function defined as post-bronchodilator FEV1 ≥ 80% predicted, FEV1/FVC ≥ 0.7, 15+ pack-years, n = 318; and susceptible-smokers with LC, 10+ pack-years, n = 247. Normal DNA was sequenced utilizing VCRome 2.1 (depth 200X over exonic regions). Promising candidates were validated in an independent set of 510 susceptible-smokers with LC and 155 resistant-smokers with normal lung function, genotyped using the MEGA panel, a high-density custom exome-array. **Results:** We discovered 30 common regulatory polymorphisms ($P \leq 1.0 \times 10^{-6}$) and 19 rare non-synonymous substitutions (occurring in 3+ susceptible-smokers and zero resistant-smokers) associated with LC risk. The top deleterious rare variants were 10q25 *CCDC147* p.Arg696Cys (occurring in eight susceptible-smokers in validation), 10q23.33 *IFIT3* p.Leu390Arg and 11p14.1 *BDNF* p.Thr2Ile (occurring in one susceptible-smoker in validation). These mutations were very rare (MAFs < 0.005). Gene-burden tests revealed strong association of common regulatory variants in 10q23.33 and LC risk ($P \leq 1.0 \times 10^{-5}$). **Conclusion:** Our results support the risk-conferring role of *CCDC147* and *IFIT3* in LC and demonstrate that exome sequencing analysis may pinpoint disease-causing functional variants in genomic regions initially identified by GWAS.

2772F

Genome-wide association study identifies new susceptibility regions associated with Ewing sarcoma risk. *M.J. Machiela[1], E. Karlins[1], W. Zhou[1], J. Khan[2], S. Ballet[3], E. Lapouble[3], V. Laurence[3], J. Michon[3], G. Pierron[3], L. Mirabello[1], L. Morton[1], G. Armstrong[4], L.L. Robison[4], S. Bhatia[5], R. Hoover[1], M.A. Tucker[1], F. Tirode[3], O. Delattre[3], S.J. Chanock[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 2) Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA; 3) Institut Curie, Paris, France; 4) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA; 5) Institute for Cancer Outcomes and Survivorship, University of Alabama at Birmingham, Birmingham, AL, USA.

Ewing sarcoma is an aggressive pediatric bone tumor predominantly observed in individuals of European ancestry. A defining characteristic of Ewing sarcoma is a fusion of a member of the ETS transcription factor family, usually the *FLI1* gene, with *EWSR1* resulting in an aberrant transcription factor that promotes cell transformation. A previous genome-wide association study identified candidate risk loci at 1p36.22, 10q21 and 15q15. A functional study of the 10q21 region has since localized the association signal to variation in a GGAA repeat motif that when bound by *EWSR1-FLI1* functions as an active regulatory element of *EGR2*. We performed an expanded genome-wide association study of 700 Ewing sarcoma cases (162 within the Childhood Cancer Survivor Study) and 1,400 ancestry matched controls. Our analysis strengthened previously published association signals at 1p36.22, 10q21 and 15q15 as well as identified two novel candidate loci: a locus on chromosome 6 and a locus on chromosome 20. The chromosome 6 locus is tagged by rs7742053 (OR=1.83, 95% CI=1.50-2.24, P-value=3.88×10^{-9}) while the chromosome 20 locus is tagged by rs12106193 (OR=1.98, 95% CI=1.65-2.37, P-value=1.33×10^{-13}). Notably, the estimated effect sizes of these associations are of greater magnitude than that observed in adult cancers. Extended haplotypes in these associated regions suggest a set of plausible candidate genes for future functional analyses. An additional promising Ewing sarcoma associated locus tagged by rs10627880 was identified on chromosome 3 (P-value=7.79×10^{-8}), but the locus did not achieve the genome-wide significance threshold and thus requires further replication and investigation. Collectively, our findings replicate prior Ewing sarcoma genetic associations and provide evidence for novel Ewing sarcoma susceptibility regions.

2773W

Identification of novel epithelial ovarian cancer loci in women of African ancestry from the Ovarian Cancer Association Consortium. *A. Manichaikul[1,2], L.C. Peres[2], X-Q. Wang[2], K. Lawrenson[3], S. Abbott[2], A.G. Schwartz[4], A.H. Wu[5], E. Peters[6], P.G. Moorman[7], M.L. Cote[4], M. Bonds[8], L.E. Kelemen[9,10], E.L. Goode[11], J. Barnholtz-Sloan[12], S.A. Gayther[13], A. Berchuck[14], J.A. Doherty[15], P. Pharoah[16], J.M. Schildkraut[2], the African American Cancer Epidemiology Study (AACES) and the Ovarian Cancer Association Consortium.* 1) Center for Public Health Genomics, University of Virginia, Charlottesville, US; 2) Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; 3) 3Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 4) Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI, USA; 5) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA; 6) Epidemiology Program, Louisiana State University Health Sciences Center School of Public Health, New Orleans, Louisiana, USA; 7) Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA; 8) Cancer Prevention and Population Sciences Program, Baylor College of Medicine, Houston, TX, USA; 9) Department of Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, USA; 10) Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, USA; 11) Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA; 12) Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA; 13) The Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 14) Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA; 15) Departments of Epidemiology, and Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; 16) Strangeways Research Laboratory, University of Cambridge, 2 Worts' Causeway, Cambridge CB1 8RN, UK.

Epithelial ovarian cancer (EOC) is a rare but deadly disease for which there is notably poorer survival in women of African Ancestry (AA) compared to women of European Ancestry (EA). Previous EA-based genome-wide association studies (GWAS) have identified 30 common, low penetrant EOC susceptibility alleles. Using the custom-designed 533,631 SNP array, the Illumina OncoArray and imputation to ~8 million genetic variants in the 1000 Genomes Phase 3, we conducted a GWAS in 767 AA EOC cases, including 213 confirmed high-grade serous ovarian cancer (HGSOC) cases, and 1,305 AA controls from the Ovarian Cancer Association Consortium (OCAC). We identified novel susceptibility loci at 4q13.3 for all EOC (lead SNP rs4286604, chr 4 bp 70,442,165, 15 kb 3' of *UGT2A1*, P=2.1x10^{-8}, MAF=0.27) and 4p16.1 for HGSOC (lead SNP rs149394398, chr 4 bp 8,878,796, 5 kb 5' of *HMX1*, P=5.0x10^{-9}, MAF=0.09). These SNPs were not genome-wide significant in the larger OCAC EA GWAS including up to 23,543 EOC cases and 29,444 controls, although an alternate SNP near *UGT2A1* showed regional significance in EA (rs7667921, chr 4 bp 69,665,106, P=1.5x10^{-5}). The *UGT2A1* region SNP identified in EA women in not in linkage disequilibrium (LD) with the lead SNP identified in AA women (R-squared<0.02 in1000G EUR and AFR). *UGT2A1* belongs to the UDP-glycosyltransferase family and plays a role in regulation of the Vitamin D receptor (*VDR*) gene, representing an important pathway previously identified in relation to risk of EOC in AA. Of the 30 SNPs identified in previous EA EOC GWAS, we observed nominally significant associations (P<0.05) with consistent direction of effect for six SNPs in AA women. Additional analysis of these loci revealed regionally significant association at 8q21.13 in AA (rs1451991, chr 8 bp 82,862,754, 100 kb 5' of *SNX16*, P=2.6x10^{-6}, MAF=0.382) at a SNP distinct from that identified in EA (rs11782652, R-squared<0.01 in 1000G EUR and AFR). *SNX16* promotes degradation of EGFR after EGF signaling, further linking it to the *VDR* pathway. Overall, our investigation presents evidence of (1) variants for EOC shared among EA and AA women, (2) susceptibility loci representing distinct underlying variants in EA and AA, and (3) novel EOC risk loci of importance for AA women, possibly representing genes related to the Vitamin D Receptor (*VDR*). Our findings underscore the importance of further genetic studies to identify variants and pathways associated with pathogenesis of EOC in AA women.

2774T

Novel pleiotropic risk loci for melanoma and nevus count. *N.G. Martin[1], G. Zhu[1], D.L. Duffy[1], G.W. Montomery[1,9], N.K. Hayward[1], S. MacGregor[1], M. Law[1], D. Glass[2], M. Falchi[2], V. Bataille[2], T.D. Spector[2], M. Iles[6], T. Bishop[6], J. Barrett[6], J. Taylor[6], J. Newton-Bishop[6], G. Davey Smith[3], D.M. Evans[3,10], S. Yazae[7], A. Hewitt[7], J. Mountain[8], D. Mackey[7], L. Jacobs[4], T.E. Nijsten[4], F. Lui[4], M.H. Kayser[4], X. Li[5], J. Han[5], S. Chanock[5], D. Hunter[5].* 1) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland, Australia; 2) Department of Twin Research & Genetic Epidemiology, St Thomas Hospital Campus, Kings College, London, UK; 3) MRC Integrative Epidemiology Unit and MRC Centre for Causal Analyses in Translational Epidemiology,; 4) Erasmus MC, University Medical Centre, Rotterdam, the Netherlands; 5) Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA; 6) Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK; 7) Centre for Ophthalmology and Visual Science (incorporating Lions Eye Institute), The University of Western Australia Perth, Australia; 8) Child Health Research, University of Western Australia, Perth, Australia; 9) Queensland Brain Institute, Brisbane, Australia; 10) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia.

The main clinical interest in total number of acquired melanocytic nevi on the skin is its strong correlation with melanoma risk. We have previously described genomewide association scans (GWAS) in Australian and British samples that identified loci influencing both nevus count and melanoma. Here we report a meta-analysis of 10 nevus GWAS from Australia, the Netherlands, United Kingdom and the United States, comprising a total of 23,570 phenotyped individuals. We confirmed known loci, including *MTAP*, *PLA2G6*, and *IRF4*. We detected two novel SNPs are at a genome-wide level of significance in their own right, near *LMX1B* on chromosome 8, and intergenic near *PTBP2* on chromosome 1. On combining the nevus results with those from a recent melanoma meta-analysis, SNPs near *GPRC5A*, *CYP1B1*, and *SYNE2* reached genome-wide significance. Several other loci, such as *PPARGC1B*, previously implicated in melanoma risk, reached a suggestive level of significance when evidence from nevi and melanoma was combined. While some loci act on nevus count and melanoma risk via the "nevus pathway", others are pure nevus or pure melanoma loci.

2775F

Whole-exome sequencing identifies a high frequency of germline deleterious variants in cancer predisposition genes in individuals with osteosarcoma. *L. Mirabello[1], R. Koster[1], B. Zhu[1], M. Yeager[1,2], M. Dean[1], M. Gianferante[1], L. Song[1], J.N. Sampson[1], J. Gastier-Foster[3], R. Gorlick[4], S.R. Caminada de Toledo[5], A.S. Petrilli[5], A. Patiño-Garcia[6], F. Lecanda[6], M. Serra[7], P. Picci[7], A.M. Flanagan[8,9], M. Fernanda Amary[9], N. Kurucu[10], I. Ergurhan Il-han[10], D.A. Barkauskas[11], M.L. Ballinger[12], D.M. Thomas[12], B.D. Hicks[2], B. Zhu[2], N.E. Caporaso[1], M. Tucker[1], R.N. Hoover[1], S.J. Chanock[1,13], S.A. Savage[1,13].*
1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD 20877, USA; 3) Nationwide Children's Hospital, and The Ohio State University Department of Pathology and Pediatrics, 700 Children's Dr., C0988A, Columbus, OH 43205, USA; 4) Albert Einstein College of Medicine, The Children's Hospital at Montefiore, 3415 Bainbridge Avenue, Rosenthal Room 300, Bronx, New York 10467, USA; 5) Pediatric Oncology Institute, GRAACC/UNIFESP, Rua Botucatu, 743 8º andar Laboratorio de Genética, CEP 04023-061 São Paulo SP Brazil; 6) Department Of Pediatrics, University Clinic of Navarra, Universidad de Navarra, Pío XII 36, 31080 Pamplona, Spain; 7) Laboratory of Experimental Oncology, Orthopaedic Rizzoli Institute, Bologna, Italy; 8) UCL Cancer Institute, Huntley Street, London WC1E 6BT, UK; 9) Royal National Orthopaedic Hospital NHS Trust, Stanmore, Middlesex HA7 4LP, UK; 10) A.Y. Ankara Oncology Training and Research Hospital, Department of Pediatric Oncology, Yenimahalle- Ankara, Turkey; 11) Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA, 90032, USA; 12) The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; 13) contributed equally.

   Osteosarcoma (OS), the most common malignant bone tumor in children and adolescents, occurs at a higher than expected frequency in individuals with certain cancer predisposition syndromes. Sporadic OS is significantly more common and its genetic etiology is poorly understood. We conducted an evaluation of rare exonic variants in 552 unselected OS cases compared with 1072 cancer-free controls using whole-exome sequencing (WES) to estimate the prevalence and burden of rare deleterious germline variants. WES of OS cases (extracted from either blood or buccal cells) included 347 European (EUR), 51 African (AFR), 61 admixed, 43 Hispanic, and 49 Brazilian ancestry patients. We evaluated rare (MAF <0.01) non-silent genetic variants that passed quality control filters. We performed burden tests comparing 347 EUR cases with 1001 EUR controls sequenced at the same time; initial analyses evaluated potentially pathogenic rare variants in 126 established cancer predisposing genes (CPG). A total of 12.5% of cases had a predicted pathogenic or high impact variant in an autosomal dominant (AD) CPG (29 of 92 AD CPGs); a significantly higher rare variant burden was present in the EUR cases compared to controls (P_{burden}=3.7x10^{-05}). *TP53* had the highest prevalence of pathogenic mutations (4.9% of patients, excluding variants of uncertain significance); the highest frequency was in individuals of AFR (7.8% of AFR cases) and EUR ancestry (5.7% of EUR cases; P_{burden}=1.05x10^{-09}). *COL7A1* and *BRCA2* had the next highest frequencies of rare AD CPG variants in cases but were not significantly higher than the controls. *RB1*, *MEN1* and *CDKN2A* had a significantly higher rare variant burden in the EUR cases (P=0.01). 6.5% of cases had a predicted pathogenic mutation in an autosomal recessive (AR) CPG (19 of 28 AR CPGs; P_{burden}=5.4x10^{-03}). Variants in these AR genes were heterozygous, suggesting they could be risk factors for OS. Two male patients had a mutation in an X-linked CPG (*DKC1* and *GPC3*; 2 of 6 X/Y-linked CPGs). Overall, 107 OS cases (19.4%) had a predicted pathogenic or high impact mutation in a CPG, which was significantly higher than the controls (P_{burden}=2.4x10^{-07}; even after excluding *TP53*, P_{burden}=6.2x10^{-04}). Our data suggest that additional CPGs not previously associated with OS could be important and merit further follow-up. .

2776W

A polygenic risk score analysis of body mass index and the risk of non-Hodgkin lymphoma. *A. Moore[1], E. Kane[2], L.R. Teras[3], A. Monnereau[4], C.F. Skibola[5], N. Wong Doo[6,7], Z. Wang[8], M. Machiela[1], O.A. Panagiotou[1], B.M. Birmann[9], A.R. Brooks-Wilson[10], R. Vermeulen[11], J. McKay[12], P. Cocco[13], G.G. Giles[6], C.M. Vajdic[14], N. Camp[15], L.M. Morton[1], K.E. Smedby[16], D. Albanes[1], B.K. Link[17], A. Zeleniuch-Jacquotte[18], Y. Zhang[19], J. Vijai[20], S. Chanock[1], N. Rothman[1], S.L. Slager[21], J.R. Cerhan[21], S.I. Berndt[1], NHL GWAS Initiative.*
1) Division of Cancer Epidemiology & Genetics, National Cancer Institute, Rockville, Maryland, USA; 2) Epidemiology and Genetics Unit, Department of Health Sciences, University of York, York, UK; 3) Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA; 4) Epidemiology of Childhood and Adolescent Cancers Group, Inserm, Center of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), Paris, France; 5) Department of Epidemiology, School of Public Health and Comprehensive Cancer Center, Birmingham, Alabama, USA; 6) Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria, Australia; 7) Concord General & Repatriation Hospital, NSW, Australia; 8) St. Jude Children's Research Hospital, Memphis, Tennessee, USA; 9) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 10) Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 11) Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands; 12) Genetic Cancer susceptibility Group, Section of Genetics, International Agency for Research on Cancer, Lyon, France; 13) Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Monserrato, Cagliari, Italy; 14) Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia; 15) Division of Hematology and Hematologic Malignancies, Department of Internal Medicine, Huntsman Cancer Institute and University of Utah School of Medicine, Salt Lake City, Utah, USA; 16) Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 17) Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa, USA; 18) Department of Population Health, New York University School of Medicine, New York, New York, USA; 19) Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA; 20) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA; 21) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA.

   Background: Currently, some studies report an association between increasing body mass index (BMI) and the risk of non-Hodgkin lymphoma (NHL). Some studies also report differences in the association by sex, timing of BMI measurement, and NHL subtype. **Methods**: To investigate the association between BMI and NHL, we used a polygenic risk scores (PRS) to estimate the association between the genetically inferred BMI and risk of NHL. Using data from eight NHL GWAS, including 10,629 cases and 9,505 controls of European ancestry, we constructed both unweighted and weighted PRS using the 97 SNPs previously-identified with genome-wide significance in association with BMI in the GIANT consortium. The unweighted PRS was the sum of the imputed dosage of BMI-increasing alleles. The weighted PRS was the sum of the weighted dosage of BMI-increasing alleles, where weights were the previously reported regression coefficients for each SNP by the GIANT consortium. We used logistic regression to evaluate PRS associations with four NHL subtypes and then combined individual study results in a fixed-effects meta-analysis for each subtype: diffuse large B-cell lymphoma (DLBCL, n=3,857), follicular lymphoma (FL, n=2,847), chronic lymphocytic leukemia (CLL, n=3,100), and marginal zone lymphoma (MZL, n=825). We adjusted for sex, age at NHL diagnosis, and population stratification. **Results**: We did not find evidence of an association between either the unweighted or the weighted BMI PRS and the risk of individual NHL subtypes (p>0.05). When stratified by sex, there was evidence for an association between the continuous weighted PRS and DLBCL risk among males (OR$_{fixed}$ per one-unit increase in PRS=1.50, 95% CI=1.07-2.11, p=0.019). No association was observed among females for DLBCL (OR$_{fixed}$= 0.93, 95% CI=0.64-1.37, p=0.727). Additional analyses will explore the risk association with PRS for waist-to-hip ratio and other obesity-related traits. **Conclusions**: Although the established BMI loci only explain a small fraction of the variance in BMI, this study complements previous epidemiologic studies and supports an association between increased genetic BMI and the risk of DLBCL among males. This analysis also supports the importance of studying NHL subtypes individually and the possibility of sex differences in etiology of NHL.

2777T

Familial lung cancer is significantly linked to six genes on 6q24-25. *A.M. Musolf[1], C.L. Simpson[1], M. de Andrade[2], D. Mandal[3], C. Gaba[4], P. Yang[2], M. You[5], E.Y. Kupert[5], M.W. Anderson[5], A.G. Schwartz[6], S.M. Pinney[7], C.I. Amos[8], J.E. Bailey-Wilson[1].* 1) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Mayo Clinic, Rochester, MN; 3) Louisiana State University Health Sciences Center, New Orleans, LA; 4) University of Toledo Dana Cancer Center, Toledo, OH; 5) Medical College of Wisconsin, Milwaukee, WI; 6) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 7) University of Cincinnati College of Medicine, Cincinnati, OH; 8) Geisel School of Medicine, Dartmouth College, Lebanon, NH.

   Lung cancer is the leading cancer-related cause of death in the United States, accounting for about 27% of all cancer deaths. Though lung cancer risk has long been known to be highly affected by the environment (particularly tobacco smoking), there is evidence for a genetic component as well, supported by familial aggregation and segregation studies. We had previously performed multipoint linkage analysis on families with a strong family history of lung cancer from the Genetic Epidemiology of Lung Cancer Consortium and found significant linkage to the 6q24-25 region. We performed targeted sequencing on 75 individuals in the 9 most strongly linked families. Two-point linkage analysis was calculated via paramlink. We also performed regional linkage analysis using SEQLinkage, which built haplotypes to create regional multiallelic markers corresponding to a gene or piece of a gene. These markers were then analyzed for two-point linkage by MERLIN.   We report 7 significant regional HLOD scores (HLOD> 3) across 6 genes: *ZBTB2*, *PARK2*, *CNKSR3*, *EZR-AS1*, *PHACTR2*, and *ADGB*. Both *PHACTR2* and *PARK2* have been implicated directly in lung cancer, and *ZBTB2*, *CNKSR3*, and *ADGB* have been implicated in other cancers. The antisense RNA gene *EZR-AS1* has not been implicated in cancer; though its target has been and noncoding RNAs in general have been found to be involved in various cancers. We did not find any genome-wide significant two-point LOD scores, but we did find several suggestive (HLOD>1) signals. Interestingly, many of these signals were located in known regulatory regions, enhancers, or TF binding sites of the top genes in the regional analyses.   It is not particularly surprising that we found multiple significant signals within 6q24-25. This region is rich with good candidate genes (including those found to be significant) and these families are linked to the region. Also allelic and locus heterogeneity is quite common in hereditary cancers so it is not unreasonable that different families might be segregating different high-risk variants in this region.   Thus, we report that the significant linkage signal previously discovered at 6q24-25 is at least partially explained by the narrowed region encompassing the 6 genome-wide significant genes reported here. Functional analysis of candidate variants in this region is needed to confirm which variants/genes (in conjunction with smoking) substantially increase risk of lung cancer in these high-risk families. .

2778F

Field synopsis of genetic variations and colorectal cancer: Unified 2016 update. *C. Nyiraneza[1], Z. Montazeri[1], E. Theodoratou[2], W. Zheng[3], H. Campbell[2], J. Little[1].* 1) University of Ottawa, Ottawa, Ontario, Canada; 2) University of Edinburgh, Edinburgh, United Kingdom; 3) Vanderbilt University School of Medicine, Nashville, TN, United States.

Background: According to current predictions of global demographic changes, the WHO estimates a 77% increase in colorectal cancer (CRC) incidence and 80% increase in CRC-related deaths by 2030 (Bray etal., 2012). Identifying genetic variants that influence susceptibility to disease potentially can inform the development of approaches for primary and secondary prevention. Two groups have published field synopses on genetic variants associated with CRC, in JNCI in 2012 and Gut in 2013. These groups are now working together on an update. **Methods:** We have searched all published (and some unpublished) genetic association data –including candidate gene and GWAS- for CRC to the end of 2015. Furthermore, we have access to 4 GWAS datasets. We are conducting meta-analyses. We have reached a consensus on the operationalization of the Venice guidelines on the credibility of genetic association, which had differed between the two previous field synopses. **Results:** More than 200 genes in about 770 SNPs were identified as putatively associated with CRC. For 450 SNPs, there was a single study only; for 90 SNPs, two studies; and for 230 SNPs, 3 or more studies. Meta-analysis will be carried out for SNPs with 3 or more studies, and credibility of association assessed according to the Venice criteria. **Conclusions:** The identification of genetic variants with influence on CRC risk may reflect an importance of genes involved in CRC risk. Our data should help results of genetic associations studies to be placed in context and interpreted appropriately and should help direct future research effort.

2779W

Genetic characterization of somatic alterations in radiation-associated chronic lymphocytic leukemia. *J. Ojha[1], A.J. de Smith[1], S. Gonseth[1], M. Zhou[1], H.M Hansen[1], .J. Nakamura[1], P.M. Bracci[1], K.M. Walsh[1], S.C. Finch[6], R.F. Reiss[2,3], M. Little[5], I. Dyagil[4], S.N. Gudzenko[5], D. Bazyka[4], J.L. Wiemles[1], L.B Zablotska[1].* 1) Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Department of Pathology and Cell Biology; 3) Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, USA; 4) National Research Center for Radiation Medicine, Kyiv, Ukraine; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; 6) Robert Wood Johnson Medical School, Camden, New Jersey, USA.

Chronic lymphocytic leukemia (CLL) was the predominant leukemia in the recent study of cleanup workers of the Chernobyl nuclear accident in Ukraine. Significantly elevated radiation risks were reported for CLL, with CLL cases characterized by early age of onset, advanced Rai staging at diagnosis and shorter survival with higher doses. In this study we analyzed the effects of radiation exposure on leukemogenesis in Chernobyl cleanup workers. We performed targeted deep sequencing of 538 cancer-associated genes in tumor DNA from 17 radiation-exposed cleanup workers (Ex) and 28 unexposed general population Ukrainian CLL cases (UnEx), and compared the results with 100 western CLL cases (W-CLL) (publically available data from previous sequencing studies, matched on age males only). *POT1* was the most frequently mutated gene in Ex cases (25%) followed by *NOTCH1* (16%)*, RB1 (*16%), and *ATM, APC, MED12, SF3B1, KMT2C* (2% each). Total number of point mutations was comparable in Ex (range 3-12, median 9), UnEx (range 2-12, median 8) and W-CLL (range 2-11, median 8) cases. Copy number alterations (CNAs) were screened via CNVkit. The most common CNAs in exposed cases were -13q14, -11q (12% each), -17p and +12q (6% each). Total number of lesions (mutations & CNAs) among exposed cases was strongly associated with type of work performed in the Chernobyl zone (p=0.013), number of doctor visits prior to diagnosis, and several time-dependent variables (combined R-square= 0.96). Additionally, progressively stronger association of total lesions with radiation dose of increasing latency periods (p=0.11 for lag=15 years) was observed. We estimated telomere length (TL) using Tel-Seq algorithm and identified significantly longer TL in Ex compared to UnEx cases (p=0.009, adjusted for age). In addition, lifestyle risk factors such as alcohol consumption and smoking, type of cleanup work performed was also associated with differences in TL. The higher prevalence of *POT1* (25%, 0% and 3%) and *ATM* (11%, 3% and 8%) mutations in exposed Chernobyl workers compared to UnEx and W-CLL cases, respectively, suggests that mutations in telomere-related genes may be critical to radiation-associated leukemogenesis. As *POT1* mutation was also associated with poorer patient survival, the relationship between telomere maintenance, radiation exposure, and CLL prognosis merits further investigation.

2780T

A genetic analysis of longitudinal prostate-specific antigen concentrations. *S. Rashkin[1], T. Hoffmann[1,2], L. Sakoda[3], E. Jorgenson[3], L. Habel[3], J. Shan[3], D. Ranatunga[3], C. Chao[4], N. Ghai[4], D. Aaronson[5], J. Presti[5], C. Schaefer[3], N. Risch[1,2,3], S. Van Den Eeden[3,6], J. Witte[1,2,6].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA; 5) Department of Urology, Kaiser Oakland Medical Center, Oakland, CA; 6) Department of Urology, University of California San Francisco, San Francisco, CA.

Prostate-specific antigen (PSA) measures have been used to screen for prostate cancer as increased PSA concentrations are indicative of prostate cancer. However, PSA concentration can vary between individuals due to various factors, leading to only moderate accuracy of PSA testing with a universal threshold. It has been proposed that PSA screening for prostate cancer can be improved by incorporating information from genetic markers influencing PSA concentrations independent of prostate cancer to determine individual-specific screening thresholds.   Studies have been conducted examining the relationship between PSA concentrations and genetic variation, but most use PSA levels measured at a single point in time or simplify longitudinal measurements to a single summary statistic. However, knowledge of how genetics affects PSA concentrations over time may better aid in determining a personalized screening threshold for an individual by incorporating how PSA is likely to change for an individual over time independent of prostate cancer to determine if an increase in PSA over time falls within normal bounds for that individual.   Here, we explore the genetic basis of changes in PSA concentrations over time in Kaiser Permanente health plan members. A previous genome-wide association study in this dataset (28,503 non-Hispanic white Kaiser Permanente health plan members and 18,615 men from replication cohorts) compared each SNP to the median PSA level over time (at least nine measures per man) and found multiple SNPs associated with PSA levels, and many of the novel SNPs were not associated with prostate cancer. We now improve on this analysis with a study looking at the longitudinal measures and trends over time by using longitudinal regression analysis to model PSA concentration.

2781F

Genetic variation in the eicosanoid pathway is associated with non-small-cell lung cancer (NSCLC) survival. *L.N. Sausville[1,2], C.C Jones[3,4], M.C. Aldrich[3,4], W.J. Blot[5,6], A. Pozzi[7,8], S.M. Williams[1,2].* 1) Genetics, Dartmouth College, Hanover, NH., United States; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH, United States; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, United States; 4) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN, United States; 5) Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, TN, United States; 6) International Epidemiology Institute, Rockville, MD, United States; 7) Department of Biochemistry, Vanderbilt University Medical Center, Nashville, TN, United States; 8) Department of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, TN, United States.

Globally lung cancer results in more deaths worldwide than any other cancer, indicating a need for better treatments. Members of the eicosanoid metabolism pathway represent promising therapeutic targets, as several enzymes involved in the generation of these lipids are dysregulated in many cancers and their inhibition reduces lung cancer growth in mouse models. However, genetic variations of enzymes involved in eicosanoid metabolism have not been adequately examined for association with lung cancer. The goal of this study was to determine whether genetic variations altering eicosanoid producing enzyme function and/or expression are associated with differences in lung cancer survival.   We examined the association of genetic variation within eicosanoid metabolism genes and lung cancer mortality in 395 non-small-cell lung cancer (NSCLC) cases from the Southern Community Cohort Study (SCCS). A total of 103 SNPs in 19 genes, both common and rare SNPs, were examined for association with NSCLC mortality. Analyses revealed no single common genetic variant or collapsed rare genetic variants within one gene associated with NSCLC survival across the entire study population. However, significant subgroup effects were observed. Rare variants in *ALOX15B* (arachidonate 15-lipoxygenase, type B) were associated with worse survival in women (HR = 2.630, p-value = 0.002), and the common variant rs12529 in *AKR1C3* (prostaglandin F synthase) was associated with improved survival in African Americans (HR = 0.726, p-value = 0.009). The lack of genetic associations among all NSCLC cases and the association among women only for rare variants in *ALOX15B* may, in part, explain the better NSCLC survival observed among women. These results raise the possibility that that some subgroups within the NSCLC population may benefit from drugs targeting eicosanoid metabolism.

2782W

Fanconi anemia gene variants predispose to hereditary pancreatic cancer. *T.P. Slavin, S. Neuhausen, M. Niell-Swiller, C. Rybak, I. Solomon, K. Blazer, A. Adamson, Y. Yate-Ching, K. Yang, S. Sand, D. Castillo, J. Herzog, X. Wu, S. Tao, S. Raja, T. Chavez, V. Chung, G. Singh, J.N. Weitzel.* City of Hope, Duarte, CA.

Approximately 5-10% of all pancreatic cancers have a single gene hereditary component. Hereditary pancreatic cancer susceptibility has been associated with familial atypical multiple mole and melanoma syndrome, hereditary breast and ovarian cancer syndrome, Lynch syndrome, Peutz-Jeghers syndrome, von Hippel-Lindau syndrome, hereditary pancreatitis, and *ATM* and *PALB2*-related cancer. Adults with pancreatic cancer of any histology and a germline DNA sample (n= 54), were selected from the Clinical Cancer Genomics Cancer Research Network (CCGCRN) registry. Germline next generation sequencing was completed using a targeted 706 gene capture kit with full exon coverage for candidate (~600 genes), and previously characterized, hereditary cancer susceptibility genes. The majority (70%) of samples sequenced had previously uninformative clinical testing. Five individuals were previously identified to have a pathogenic mutation through clinical testing. Forty-nine research participants were sequenced and analyzed. Results showed 32 American College of Medical Genetics and Genomics pathogenic (P), likely pathogenic (LP), or protein truncating variants of unknown significance in 27 participants. Seventeen out of 54 participants were found to be carriers of a P or LP variant (31% of the sample set). Participants with P or LP variants had a younger mean age of pancreatic cancer diagnosis (53.9 vs 57.2 years, T-test p-value <0.0001). Seventy percent of those with a P or LP variant had a first, second, or third degree family member with cancer (Chi-square p-value <0.0001). Eight out of 54 participants (15%) had pathogenic or likely pathogenic variants in the hereditary breast cancer signaling pathway (*ATM*, *ATR*, *BRCA2*, *FANCA*, *FANCF*, *FANCM*, and *RAD50*; p-value 5.55E-15); all of these participants had adenocarcinomas or presumed adenocarcinomas. This study describes the spectrum of identified pancreatic cancer susceptibility in the CCGCRN registry. Further exploration into germline hereditary pancreatic cancer susceptibility should include evaluation of the hereditary breast cancer signaling pathway, and particularly the Fanconi anemia genes.

2783T

Association of high-evidence gastric cancer susceptibility loci and somatic gene expression levels with survival. *H. Sung[1], H. Nan[2], C. Giffin[3], L. Song[1], H. Su[1], C. Wang[1], T. Ding[4], D. Parisi[3], A. Goldstein[1], P. Taylor[1], P. Hyland[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD; 2) High-dimension Data Analysis Group, Basic Research Laboratory, Center for Cancer Research, NCI, NIH, Bethesda, MD; 3) Information Management Services, Inc, Calverton, MD; 4) Shanxi Cancer Hospital, Taiyuan, Shanxi, People's Republic China.

Objective: Eleven high-evidence SNPs at nine loci for gastric cancer (GC) risk have been reported, but their association with survival remains unknown. **Design**: SNP associations and hazard ratios for GC survival overall and by anatomic location and histology were examined among 1,147 incident cases from the Shanxi Upper Gastrointestinal Genetics Project. We further examined if SNPs were expression quantitative trait loci (eQTL) in normal and tumor gastric tissues, and whether tumor *versus* normal somatic mRNA differences in126 cases were associated with survival. **Results**: No SNPs were associated with GC survival overall. However, specific associations with survival for: gastric cardia adenocarcinomas (GCA), *MUC1/MTX1/*1q22 rs2070803 (HR$_{\text{AA vs. GA+GG}}$=2.16; 95% CI=1.24-3.78; P_{FDR}=8.4E-02) and *LTA/TNF/*6p21.33 rs1799724 (HR$_{\text{TT+CT vs. CC}}$=1.30; 95% CI=1.07-1.57; P_{FDR}=8.4E-02), and diffuse-type GC, *PSCA/JRK/*8q24.3 rs2294008 (HR$_{\text{TT vs. CT+CC}}$=1.99; 95% CI=1.33-2.97; P_{FDR}=1.3E-02) were observed. Rs2294008T was an eQTL for *PSCA*, upregulating mRNA in normal gastric (beta=0.60; P=5.7E-21) and GC (beta=0.30; P=8.9E-03) tissues. Also, diffuse-type GC cases with the highest *PSCA* mRNA had much shorter survival than cases with the lowest mRNA (median survival of 0.47 years in the highest quartile *versus* 3.73 years in the lowest quartile; HR=9.70; 95% CI=2.46-38.4; P=1.2E-03). Similar, but less striking survival effects were observed for *MTX1* in GCA and for *JRK* in diffuse GC tissues. **Conclusion:** Our results suggest that three high-evidence GC risk loci also have prognostic importance in GC subtypes. Since PSCA protein is also a target for immunotherapy in other cancers, future clinical studies are needed to validate PSCA as a therapeutic target for GC.

2784F

Heritability of aggressive prostate cancer. *C. Tai[1], T.J. Hoffmann[1,2], L.C. Sakoda[3], E. Jorgenson[3], L.A. Habel[3], J. Shan[3], D.K. Ranatunga[3], C.R. Chao[4], N.R. Ghai[4], D. Aaronson[5], J. Presti[5], C. Schaefer[3], N.J. Risch[1,2,3], S.K. Van Den Eeden[3,6], J.S. Witte[1,2,6].* 1) Department of Epidemiology & Biostatistics, UCSF, San Francisco, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA; 3) Division of Research, Kaiser Permanente Northern Califoria, Oakland, CA; 4) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena CA; 5) Department of Urology, Kaiser Oakland Medical Center, Oakland, CA; 6) Department of Urology, UCSF, San Francisco, CA.

Prostate cancer has been documented to be one of the most heritable cancer phenotypes with reported estimates ranging from 0.42 to 0.58. However many known prostate cancer loci are not strongly associated with aggressive prostate cancer phenotypes such as biochemical recurrence or high-grade cancer. Previous efforts have been largely underpowered or not replicated. Therefore, it is unclear whether more aggressive prostate cancer phenotypes are as heritable as overall prostate cancer and whether we can identify SNPs specific for aggressive prostate cancer. Using the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, this study aimed to determine how much of the phenotypic variance could be explained with a GWAS panel with 650,000 SNPs in 4,720 cases Non-Hispanic white prostate cancer with Gleason scores, of which recurrence could be determined in 3,583 cases. We evaluated two aggressive phenotype definitions, high-grade disease (Gleason score ≥ 7 or ≥ 8 compared to <7 or <8) and biochemical recurrence (defined separately for radiation and surgery treated cases, and compared to non-recurrent cases). Analyses were adjusted for age, BMI, and ancestry principal components. Heritability was estimated in unrelated individuals based on a pairwise relatedness cutoff of 0.025. We found the heritability for high-grade disease to be 0.09 (SE=0.27) and 0.67 (SE=0.48) for Gleason ≥ 7 and ≥ 8, respectively. The heritability for biochemical recurrence among cases with radiation treatment was 0.17 (SE=0.54). These estimates provide a starting point to identifying SNPs specific for aggressive disease and suggest that aggressive prostate cancer is heritable.

2785W

Role of *GSTM1* and *GSTT1* copy number variations in differentiated thyroid cancer risk in Europeans from metropolitan France and Melanesians from New Caledonia. *C. Tcheandjieu, E. Cordina-Duverger, M. Sanchez, P. Guenel, T. Truong.* CESP-Centre for research in Cancer Epidemiology and Population Health, Cancer Epidemiology: genes and environment, Inserm U1018, Universite Paris Sud, Universite Paris Saclay, Villejuif, France.

GSTM1 and *GSTT1* encode for enzymes that are involved in the detoxification of various compounds including carcinogens. A dose-effect relationship has been reported between copy number variations (CNVs) of these genes and their enzyme activity. In this study, we analyze whether they influenced the risk of differentiated thyroid carcinoma (DTC) and whether they modify the association between DTC risk and lifestyle risk factors including body mass index (BMI), tobacco smoking, alcohol intake and reproductive factors such as age at menarche, oral contraceptive and parity. We analyzed *GSTM1* and *GSTT1* CNVs in 505 cases and 622 controls of European descent and in 156 cases and 114 controls of Melanesian descent from 2 population-based case-control studies conducted in Metropolitan France and in New-Caledonia. Overall, *GSTM1* and *GSTT1* CNVs were not associated with DTC risk. However, the association between DTC risk and BMI and alcohol drinking in both populations were modulated by the presence or the absence of these genes. The OR contrasting BMI≥ 30 and BMI<25 kg/m² was stronger among *GSTT1*-null subjects (OR= 2.17 (95%CI: 0.86–5.47) in Europeans and OR= 7.63 (95%CI: 2.21–27.64) in Melanesians than among those with non-null genotype ((OR 1.25 (95%CI: 0.87–1.81) and 1.30 (95%CI: 0.48-3.54), respectively). This association was even more striking in *GSTM1*-null and *GSTT1*-null subjects with an OR and 95%CI of 4.09 (0.98-17.13) in Europeans and 13.28 (2.79-63.17) in Melanesians. We reported an inverse association between alcohol drinking and DTC risk (OR (95%CI)= 0.71 (0.53–0.94) in Europeans and 0.58 (0.31–1.11) in Melanesians), which was stronger among subjects with *GSTM1*-null, *GSTT1*-null or *GSTM1*-null/*GSTT1*-null genotypes. The ORs and 95%CI comparing ever vs never drinking alcohol in Europeans were 0.50 (0.32–0.76) in *GSTM1*-null subjects; 0.61 (0.29–1.28) in *GSTT1*-null subjects; and 0.28 (0.08–1.02) in *GSTM1/GSTT1*-null subjects; the corresponding ORs in Melanesians were 0.70 (0.37–1.32); 0.36 (0.13–1.04) and 0.24 (0.07–0.81). In Europeans the risk of DTC associated with age at menarche >13 years was stronger in *GSTT1* non-null than in *GSTT1* null genotype (p-interaction=0.02), and the risk of DTC associated with parity was stronger in *GSTT1*-null than in non-null genotype (p-interaction=0.02). Our results suggest a possible modification effect of *GSTM1* and *GSTT1* genotypes in the association between DTC risk and BMI, alcohol consumption and some reproductive factors.

2786T

Prediction of years of life after diagnosis of breast cancer using multi-layer omic and clinically relevant covariates: Two studies using METABRIC and TCGA. *A. Vazquez[1], A. Gonzalez[1], Y. Veturi[2], G. de los Campos[1,3].* 1) Epidemiology and Biostatistics, Michigan State University, Lansing, MI; 2) Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 3) Department of Statistics and Probability, Michigan State University, East Lansing, MI.

High-throughput technologies provide large volumes of data from multiple 'omics', including the genome, epigenome, transcriptome, proteome, metabolome and other omics. These data hold valuable information for prediction of disease outcomes. However, integrating high-dimensional multi-layer omic data into risk assessment models can be statistically and computationally challenging: inputs are multi-layer and high-dimensional and the underlying biological processes may involve complex interaction patterns between omics (e.g., gene expression by treatment). We describe a statistical framework (BGAM=Bayesian Generalized Additive Model) to incorporate multi-layer omics into disease risk assessment. BGAM integrates into a unified framework, ideas from Generalized Additive Models, Bayesian shrinkage and variable selection methods, and Reproducing Kernel Hilbert Spaces. We used BGAM in two independent datasets, the Molecular Taxonomy of Breast Cancer Consortium (METABRIC, n=1977) and The Cancer Genome Atlas (TCGA, n=285), to predict (three-year) survival after diagnosis of breast cancer. The information used includes clinical covariates (COV) and genome-wide multi-layer omic profiles from the primary tumor, including gene expression (GE), copy number variants (CNV), and for TCGA methylation (MET). GE and MET alone offered more predictive power than any of the commonly used COV including stage and cancer subtype. The best performing models were those combining COV+GE (CV-AUC for these models were 0.797 and 0.721 for METABRIC and TCGA, respectively). COV+MET also achieved high predictive accuracy (0.672 in TCGA). Use of whole-omic profiles gave significantly higher CV-AUC than prediction based on well-known GE of oncogene signatures alone. The improvement in prediction with CNV was low but positive in both data sets. We also found that GE is most predictive for patients in early stage and without nodal metastasis than for late cancers, achieving 5 (to 7) points of increment in CV-AUC beyond COV. We demonstrate in two independent data sets that the use of whole-genome GE and MET profiles can yield substantial gains in prediction accuracy relative to predictions based on commonly used COV or GE from oncogenes. In summary, we believe that for patients in early stage and without nodal metastasis there is great potential and value for the use of whole-genome GE and MET with COV for prediction of breast cancer outcomes.

2787F

Variants associated with telomere length give rise to multiple cancer associations. *J.S. Witte, R.E. Graff, J.D. Hoffman.* Epidemiology & Biostatistics, Univ California,San Francisco, San Francisco, CA.

Telomeres are repeating DNA-protein structures that mark the end of chromosomes and extremes of their length have been implicated in cancer risk. Recent genome-wide association studies have identified multiple genetic variants associated with telomere length. Our goal was to evaluate these genetic variants with respect to multiple cancers in the UK Biobank cohort. We combined genome-wide significant variants associated with telomere length into a polygenic risk score (PRS) by weighting each variant by the beta coefficient for its association with telomere length and summing across all variants. We then tested the telomere length PRS against the risk of cancer overall and of 14 individual cancers. Telomere PRS was significantly associated with the diagnosis of one cancer relative to controls with no diagnoses (p = 1.31e-9; beta = 0.324). Telomere PRS was also significantly associated with diagnosis of multiple cancers versus controls with no cancers (p = 2.82e-3; beta = 0.498). Examination of the PRS with respect to individual cancers identified significant associations with kidney cancer (p= 1.68e-3; beta = 1.18), melanoma (p = 2.73e-4; beta 0.615), testicular cancer (p = 0.028; beta = -0.875), and breast cancer (p = 0.043;beta = 0.217). Overall, our results implicate SNPs associated with longer telomere length in the risk of cancer. That the observed beta for the risk of multiple cancers was larger than that for the risk of a single cancer suggests that telomere length may have an impact on susceptibility to developing multiple different cancers. It should also be noted that we found shorter rather than longer telomere length to be associated with testicular cancer risk, highlighting potential differences in the role telomeres may play in the carcinogenesis of specific cancers.

2788W

Identifying functional differentially methylated regions in colorectal cancer. *H. Xu, S. Ghosh, V. George.* Dept Biostatistics & Epidemiology, Augusta University, Augusta, GA.

DNA methylation is critical in the regulation of gene expression and abnormality in DNA methylation has been involved in my human diseases, especially cancers. Substantial efforts have been put into identifying differentially methylated regions (DMRs) between disease and normal samples. However, due to the limited sample size, the large number of CpG sites, and the complexity in gene regulation, many of the resulting DMRs may not be functional in terms of regulating gene expression. As more data with both gene expression and DNA methylation are available, we propose an integrated approach to identify functional differentially methylated regions (fDMRs) based on meta-analysis that combines evidence from both types of genomic data. We evaluated our method through simulation studies. Results from simulation study showed that our integrated approach had well control of type-I error rate. It had improved power compared with the sequential analysis approach. We applied our data to the DNA methylation and gene expression data on colorectal cancer from the Cancer Genome Atlas and identified several fDMRs.

2789T

Integration of multi-omic layers for estimation and prediction of genetic variation in glioblastoma multiforme. *Y.L. Bernal Rubio[1], K.H. Wu[1], C. Griguer[2], J.P. Steibel[3,4], A. Gonzalez Reymundez[1], G. de los Campos[1,5], A.I. Vazquez[1].* 1) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Department of Neurosurgery, University of Alabama at Birmingham, Birmingham, AL; 3) Department of Animal Science, Michigan State University, East Lansing, MI; 4) Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI; 5) Department of Statistics, Michigan State University, East Lansing, MI.

Glioblastoma multiforme (GBM) is the highest grade and the most malignant form of an astrocytoma tumor (grade IV), which is characterized for being the most common and lethal primary brain tumor. Although standard treatments for GBM patients have improved, the targeting of treatments remains a challenge and patient survival is still very low. The availability of high throughput multi-omic data has opened new research opportunities that can lead to a better understanding of the genetic architecture of GBM and the factors that determine survival of GBM patients. This information could be used to identify genetic signatures that are predictive of survival and response to treatment. In this study we used survival data, clinical information and multi-omic profiles from GBM patients provided by TCGA consortium (n = 476). Censoring rate was 24.5% and, among uncensored observations, 53.5% of the records had time to death lower or equal than one year. Clinical covariates (CLIN) included: age at diagnosis, race, gender, pathological diagnosis method, Karnofsky performance score (KPS) and treatments, which were coded as chemotherapy Yes/No and using dummy variables that indicated the use of 18 different drugs (TRT). Omic information included gene expression profiles (GE, assessed at the tumor using RNA-seq), SNP genotypes (SNP) and methylation β-values (METH). We used Bayesian mixed models to estimate the proportion of variance explained by CLIN and omic profiles. Prediction accuracy of each of the models was assessed using cross-validations (CV) area under the ROC curve (AUC). AGE, KPS and CHEMO had significant ($p<0.05$) effects on survival, with AGE having the lowest p-value (4.1e-10) and largest effect among the CLIN covariates used. Among the omics, METH was the one that explained the largest fraction of variance in survival (23.4%, when considered simultaneously with CLIN, TRT, GE and SNP in a bayesian mixed model). Although TRT did not explain a very large fraction of variance of survival time, in CV TRT improved prediction of short term survival. For long term survival, METH had higher AUC than GE and SNP. A principal component analysis showed a differentiation in METH according to age at diagnosis, suggesting a possible effect of the age on methylation β-values and a concomitant effect of METH on survival of GBM patients.

2790F

Network analysis facilitates the prediction of ovarian cancer progression. *U. Ozbek[1], C. Conley[3], J. Peng[3], P. Wang[2].* 1) Department of Population Health Sciences and Policy, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Statistics, University of California, Davis, CA.

Ovarian cancer is the fifth leading cause of cancer death. It's estimated that there will be 22,280 new cases and 14,420 deaths of ovarian cancer in the US in 2016. To improve the treatment of ovarian cancer patients, it is of great importance to accurately predict patients' disease progression. In this study, our goal is to use network analysis to derive genomic signatures useful to predict the progression of ovarian cancer. We hypothesize that hub genes in gene regulatory network play important roles in shaping the cell phenotypes and are particularly useful for predicting patients' disease progression. To test this hypothesis, we first construct a regulatory network among DNA copy number alterations (CNA) and RNA expressions based on 293 ovarian cancer genomic profiles from The Cancer Genome Atlas using our newly developed novel statistical tool, SpaceMap. This new tool derives conditional dependence relationships among DNA copy number alterations and the RNA expression levels using a set of multivariate linear regressions with novel penalty terms to incorporate the network structures. Based on the inferred regulatory network, we identify a collection of hub genes that are influenced directly by CNA and are interacting with a large number of other genes. We then evaluate the prediction power of these hub genes for patient survival outcome and find these hub genes are in general more information than other non-hub genes. We expect such a strategy could lead to better prediction models for disease outcome in clinical practice. The candidate hub genes could also lead to novel drug targets to improve the patient disease outcome.

2791W

Methods for fine-mapping with chromatin and expression population data. *M. Roytman[1], G. Kichaev[1], B. Pasaniuc[1,2,3].* 1) Bioinformatics IDP, UCLA, Los Angeles California, USA; 2) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles California, USA; 3) Department of Human Genetics, Geffen School of Medicine, UCLA, Los Angeles California, USA.

Discerning the genetic basis of complex traits is a fundamental problem in biology. Genome-wide association studies (GWAS) have revealed that the majority of variants associated with disease lie in noncoding regulatory sequences. However, the primary mechanisms by which these variants act are poorly understood. One proposed mechanism is that regulatory variants may affect histone modifications in local regions which in turn have downstream effects on transcription. Recent studies have identified thousands of QTLs associated with local and distal histone modifications. In addition, chromatin states defined by histone marks are known to associate with functional elements and changes in gene regulation. However, this proposed chain of causality has yet to be established. In this project we present a statistical framework designed to detect such interactions in the genome. Given a locus centered around a gene, containing potentially thousands of SNPs and hundreds of chromatin peaks, our method uses a joint likelihood model to assign posterior probabilities to each possible causal path through this architecture. In simulations we show that our integrative method outperforms the naïve approach, whereby the effects of SNPs on chromatin and of chromatin on expression are fine-mapped independently and posteriors are multiplied together, reducing the size of the 90% credible set from 52 paths to 11 paths. Thus, by integrating three levels of data - genomic, chromatin, and expression - into one probabilistic model, our method improves our ability to detect the genomic variants influencing gene expression as well as the chromatin peaks through which they act.

2792T

Identification of breast cancer subtype-specific DNA methylation signatures using both mean and variance signals. *Y. Wang, S. Wang.* Department of Biostatistics, Mailman School of Public Health Columbia University, New York, New York, USA.

DNA methylation plays an important role in gene expression and cancer progression. Studies have found that DNA methylation levels between different cancer subgroups could differ not only in means but also in variances. We recently proposed a semiparametric test based on exponential tilt model (ETM) to detect both mean and variance differences in case-control designs. Elastic-net is a widely used variable selection method in the analysis of high dimensional DNA methylation data where neighboring CpG sites are highly correlated. Here, we extend on the ETM method that considers both mean and variance signals and add the elastic-net regularization that encourages a grouping effect of highly correlated methylation sites. We analyzed four subtypes of the breast invasive carcinoma (BRCA) Illumina HM450K DNA methylation data of 187 tumor samples and 59 normal samples from The Cancer Genome Atlas (TCGA) project, and identified subtype-specific DNA methylation loci for Luminal-A (LumA), Luminal-B (LumB), basal-like (Basal) and human epidermal growth factor receptor 2 (HER2) subtypes, respectively. The combined signal approach identified additional CpG sites that were missed when using mean signal only. Among the candidate genes identified by variance signal only in LumA, LumB, Basal and HER2 subtypes, many of them have also been reported to be differentially methylated in other cancers. Hence, the inclusion of variance signal in addition to the mean signal helps to achieve a better discrimination in breast cancer subtypes.

2793F

A novel region-based bayesian approach for genetic association with Next Generation Sequencing (NGS) data. *J. Xu[1,2], W. Xu[1,3], H. Massam[4], M. Escobar[1], L. Briollais[1,2].* 1) Dalla Lana School of public Health, University of Toronto, Ontario, Canada; 2) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada; 3) Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada; 4) Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada.

The discovery of rare variants through sequencing studies is becoming a very challenging issue in the field of human genetics and could help deciphering the genetic basis of many complex human diseases. Because rare variants occur too infrequently in the general population, single-variant association tests lack power in NGS analyses. We proposed here a novel region-based statistic based on a Bayes Factor (BF) approach to assess the evidence of association between a set of rare variants located in same chromosomal region and a disease outcome. Marginal likelihood functions are computed under the null and alternative hypotheses assuming a binomial distribution for the rare variants count in the region and either a beta prior distribution or a mixture of Dirac and beta distributions for the probability of observing a rare variant at a specific locus. The hyper-parameters for the prior specification are determined empirically from the data. A permutation test is used to assess the distribution of the BF under the null hypothesis of no association. Our simulations studies showed that the new BF statistic outperforms popular methods such as the Burden test and SKAT (Sequence Kernel Association Test) under most situations. Our real application to a study of lung cancer from Toronto including 258 cases and 257 matched controls suggests interesting genes such as CHEK2 associated with the cancer outcome. Sensitivity analyses and further developments of the BF approach to multi-region NGS analysis will also be discussed.

2794W

Comparison and easy implementation of methods for cosegregation analysis. *J.M.O. Ranola[1], E.A. Rosenthal[2], Q. Liu[3], B.H. Shirts[1]*. 1) Laboratory Medicine , University of Washington, Seattle, WA; 2) Medical Genetics, Medicine, University of Washington, Seattle, WA; 3) Engineering, University of Washington, Seattle, WA.

Cosegregation analysis provides crucial information in evaluating the pathogenicity of variants of unknown significance in known pathogenic genes. However, proper cosegregation analysis has required training in the correct use of available software. Genetics professionals without such training often get around this hurdle by using less robust methods, which usually consider only affected carriers.   We compared two robust methods, full-likelihood Bayes factors (FLB)(Thompson et al 2003) and co-segregation likelihood ratios (CSLR) (Mohammadi et al 2009), with the simpler method of counting meioses. The two robust methods rely on similar underlying statistical models, but allow somewhat different assumptions of variant frequency, use different methods to define penetrance, and handle age of onset in different ways. We compared all three methods using 2742 simulated 3-generation pedigrees with pathogenic or benign variants in *BRCA1* or *MLH1*. We show that CLSR and FLB methods generated median likelihood ratios (LR) of 1.31 and 1.18 for pathogenic *BRCA1* variants, respectively, with ~1% having relatively strong evidence for pathogenicity (LR > 1.3) for both methods. CSLR and FLB methods generated median LR of 0.93 and 0.88 for benign *BRCA1* variants respectively, with ~2% having strong evidence against pathogenicity (LR < -1.3) for both methods. Meiosis counting generated a median Bayes factor of 1 for both pathogenic and benign variants, with Bayes factor >1 for 20% of pathogenic and 1% of benign *BRCA1* variants. CSLR generates results that are generally 10% higher than FLB results for all *BRCA1* variants. We found similar results for simulated pedigrees with pathogenic and benign *MLH1* variants . In classifying benign variants, only CSLR and FLB methods are able to quantify the evidence against pathogenicity. Both CSLR and FLB methods perform better with larger families for both pathogenic and benign variants. Meiosis counting performs worse with larger family size for benign variants.   To facilitate the use of more robust cosegregation analysis methods we present a web tool (analyze.myvariant.org) for the calculation of FLB using FASTLINK backend calculations, which can be used for large pedigrees (N<600). We also present the CoSeg R package, available at R-forge, which can be used to simulate pedigrees using demographic parameters, implement the CSLR method, and count meioses in small to medium sized pedigrees (N<50).

2795T

Single nucleotide polymorphisms of multidrug resistance gene 1 *(MDR1)* **and risk of chronic myeloid leukemia.** *Y. Kassogue[1,2], H. Dehbi[2], M. Quachouh[3], A. Quessar[3], S. Benchekroun[3], S. Nadifi[3]*. 1) University of Sciences, techniques and Technologiques of Bamako, Bamako, Mali; 2) Genetics and Molecular Pathology Laboratory, Medical school of Casablanca, University Hassan II, Morocco; 3) Department of Onco-Hematology, Ibn Rochd University Hospital, Casablanca, Morocco.

Abstract Multi-Drug résistance gene 1 (*MDR1*) is known for its involvement in the detoxification through the active transport of toxic compounds from diverse origins outside the cells. These compounds could cause injury to cell DNA, which might lead in cancer like chronic myeloid leukemia (CML). Individual inherited genetic differences related to polymorphism in detoxification enzymes could be an important factor not only in carcinogen metabolism but also in susceptibility of cancer. The present study aimed to investigate the association of three single nucleotide polymorphisms (SNPs) of the *MDR1* gene in the susceptibility of CML. We successively have determined the genotype profiles of 1236C→T (exon 12); 2677G→T (exon 21) and 3435C→T (exon 26) SNPs by PCR-RFLP in 88 patients and 99 unrelated healthy controls. Logistic regression was used to assess the effect of each SNP on the development of CML. Interestingly, in exon 12 the 1236TT was significantly associated with the susceptibility of CML when compared to the wild type 1236CC (OR: 2.7; 95% CI: 1-7.32, p = 0.041). Additionally, the recessive model 1236TT vs. 1236CC/CT showed a risk of 3.3 fold (p = 0.011) with CML. In exon 26, the 3435CT genotype was associated with a reduced risk of CML (OR: 0.5; 95% CI: 0.3-1, p = 0.042). In exon 21, the 2677GT genotype seems to have a protective effect (OR: 0.6; 95% CI: 0.32-1.1, p = 0.074). Diplotype analysis has demonstrated no effect in susceptibility of CML, but 1236CT/3435CC and 1236CC/2677GT were associated with a protective effect. The haplotypes analysis showed no particular trend (global association p = 0.33). Our findings demonstrate that 1236TT in exon 12 might contribute in the susceptibility of CML, while the 3435CT in exon 26 as well as 1236CT/3435CC and 1236CC/2677GT combinations might be protective factors. **Keywords:** *MDR1* gene; Polymorphism; Haplotype, CML.

2796F

Genome-wide gene x smoking interaction analysis in lung adenocarcinoma. *Y. Li, X. Xiao, Y. Han, D. Qian, C. Amos.* Biomedical Data Science Department, Dartmouth College, Hanover, NH.

Purpose: Genome-wide association studies (GWAS) have successfully identified multiple genetic susceptibility genes associated with lung cancer but they only explain limited proportion of the heritability of lung cancer. Smoking is an important environmental risk factor associated with lung cancer and the interactions between smoking and genes have been suggested to play an important role in lung carcinogenesis. Our study is a genome-wide interaction analysis aiming to reveal smoking-gene interactions not identified in GWAS. Methods: The discovery data on 409398 SNPs came from four subset data-sets with a total of 20891 individuals with Caucasian ancestry including 13910 controls and 6981 patients with lung adenocarcinoma. A logistic regression model was used to test the interaction between additive SNP genotype and binary smoking status adjusting for the first three principal components. Potential signals were validated using independent genotype data from 3160 individuals including 2223 controls and 937 cases. Interaction analysis using imputed genotype at the candidate region was used as additional validation. Results: six significant SNPs with p value less than 0.1 from each subset analysis and less than 3.16×10^{-5} from analysis using all the data were detected at gene ZC3H11A/ZBED6 on chromosome 1 in discovery study. And the most significant SNP rs7540041 had a p value of 4.74×10^{-6} in analysis using all the discovery data (p values for subset 1-4: 2.69×10^{-2}, 4.05×10^{-3}, 5.63×10^{-2} and 1.83×10^{-2}, Odds Ratio varied from 1.47 to 1.97). The validation study produced p value of 0.071(OR=1.57) and 0.059 (OR=1.89) at top SNP rs7552670 and rs7540041, respectively. Imputation analysis further identified additional 28 significant SNPs at ZC3H11A/ZBED6 gene. When we combined both discovery and replication data sets, the p values at these two SNPs were 8.46×10^{-7} and 2.44×10^{-7}, respectively; the estimated interaction ORs were 1.49 (95% CI: (1.27, 1.74)) and 1.59 (95% CI: (1.33, 1.89)) at rs7552670 and rs7540041, respectively. The association analysis stratified by smoking status showed that rs7552670 and rs7540041 were protective factor (OR=0.78 and 0.73, respectively) for lung cancer in nonsmokers and risk factor (OD=1.16 for both of the two SNPs) in smokers. All the results in our study suggest the interaction between ZC3H11A/ZBED6 gene and smoking status in lung adenocarcinoma development.

2797W

Genetic predictors of gene expression associated with risk of colorectal cancer. *S.A. Bien[1], C. Qu[1], T.A. Harrison[1], Y. Su[1], D. Conti[2], S. Chen[3], H.M. Kang[3], C.M. Connolly[1], P.L. Auer[4], J. Huyghe[1], S.I. Berndt[5], S. Bézieau[6], A.T. Chan[7,8], M. Hoffmeister[9], T. Hudson[10,11], M.A. Jenkins[12], P. Newcomb[1,13], M.L. Slattery[14], J.D. Smith[15], E. White[1,13], L. Le Marchand[16], C.S. Carlson[1,13], A. Kundaje[17], G. Casey[2], L. Hsu[1,18], G.R. Abeçasis[3], D.A. Nickerson[15], U. Peters[1,13], GECCO and CCFR.* 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Preventive Medicine,University of Southern California, Keck School of Medicine, Los Angeles, CA; 3) Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 4) Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee, WI; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 6) Service de Génétique Médicale, CHU Nantes, Nantes, France; 7) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 8) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 9) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; 10) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 11) Department Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 12) Melbourne School of Population Health, The University of Melbourne, The University of Melbourne, Parkville, VIC, Australia; 13) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA; 14) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; 15) Department of Genome Sciences, University of Washington, Seattle, WA; 16) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 17) Genetics and of Computer Science, Stanford University, Stanford, CA; 18) Department of Biostatistics, University of Washington School of Public Health, Seattle, WA.

Genetic factors play an important role in the etiology of Colorectal Cancer (CRC), with an estimated 12-35% of CRC risk being explained by heritable factors. Recently, genome-wide association studies (GWAS) have successfully discovered over 50 common variant loci with weak to moderate effects on CRC risk. These genetic factors in aggregate explain only a small fraction of familial risk of CRC, suggesting that a large fraction of risk loci have not yet been identified. Given that most of the reported CRC risk loci are located in noncoding regions we hypothesized that additional common variant risk loci are likely to exert weak effects on CRC risk through cis-regulation of gene expression. We performed a gene-level association test using the recently published method PrediXcan, which uses transcriptome regulation data to characterize tissue-specific variant function. Prediction models were developed for mRNA levels in 169 transverse and 124 sigmoid colon transcriptomes using data from the GTEx project. Predicted expression in 15,506 cases and 13,079 controls from GECCO-CCFR were tested for association with CRC in up to 11,996 genes. In total, 20 genes, representing 12 unique loci, showed statistically significant CRC association at the false discovery rate = 0.2. We identified six novel loci at 2q35, 7q22.1, 12q24.13, 16q24.3, 17p11.2, 19p13.11. These results implicate both known genes/pathways (Wnt signaling, DNA repair), as well as new genes in biologically relevant pathways (e.g. chemokine receptors and MTOR signaling). If replicated, these novel findings provide new biological insight into CRC pathogenicity.

2798T

HERVnGoSeq: Identifying and mapping unfixed HERV elements in the genome and application in cancer genomics. *S.S. Francis[1,2], A.D. Wallace[2], J.A. Wendt[3], L.F. Barcellos[3], J.F. Costello[2], M.R. Wrench[2], J.L. Wiemels[2].* 1) Epidemiology , University of California, San Francisco, San Francisco , CA; 2) University of Nevada, Reno; 3) University of California, Berkeley.

   Objectives: Approximately 8% of the human genome is comprised of viral DNA. Human endogenous retroviruses (HERV) are expressed in a variety of cancers including glioblastoma (GBM). The totality of the genomic effects of HERV insertions is not fully understood. However, human and non-human studies show that ERVs have exerted strong evolutionary pressure. We seek to understand the diversity of HERV insertions in humans and to assess their role in complex diseases including cancer. **Methods:** We initiated a large-scale bioinformatics screen of HERVs in the complete 1000 Genomes dataset of 2557 whole genome (WGS) samples from individuals worldwide as baseline and The Cancer Genome Atlas (TCGA) GBM subjects n=52tumor/52 normal. We developed a custom bioinformatic pipeline for detecting HERV insertions, HERVnGoSeq, which uses a 2-stage approach, wherein the first stage nominates HERV LTR chimeric reads that mark the integration point into the genome and maps locations. The second phase of creates synthetic sequences of all insertion points (novel and those in hg19) for re-screening of raw reads. HERVnGoSeq was performed on Amazon's Elastic Computing Cloud. To map additional putative HERV integrations sites a novel GWAS based mapping approach was conducted to identify genomic positions. **Results:** The analysis was designed to detect all HERVs, yet only HERV-K was variable between humans. We identified 1788 putative HERV-K insertion sites across 2557 genomes, including 251 that are not represented in the reference genome. Many of these novel insertions were not mappable by sequence homology yet we successfully mapped >67 previously unreported sites to genomic coordinates using a novel GWAS based approach to identify SNPs associated with given insertions. We observed no HERV-K differences between TCGA GBM germ-line and tumor samples. Hierarchical clustering of TCGA GBM subjects revealed two distinct groups based solely on HERV insertions; the two groups differed significantly by age at diagnosis (mean age= 55.4, 64.2 years p=0.001) independent of ancestry. **Conclusion:** We discovered heterogeneous HERV-K insertions within and between diverse human populations where we mapped >67 previously unreported HERV-K insertions to genomic coordinates. In the TCGA GBM data, we show that two groups of GBM cases based on variable HERV-K insertions had significantly different ages at diagnosis. Further investigation of HERV-K as a germ-line risk factor for GBM is warranted.

2799F

The *R* package genMOSS for Genome-wide Association Studies (GWAS). *L. Briollais.* Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada.

   The analysis of GWAS data has long been restricted to simple models that cannot fully capture the genetic architecture of complex human diseases. As a shift from standard approaches, we present here the *R* package genMOSS that implements a Bayesian graphical model approach and stochastic search algorithm for multi-SNP analysis from GWAS data. This approach is applicable to a wide range of genetic association problems, including GWAS, fine-mapping studies and SNP-SNP interaction analyses. More specifically, genMOSS is able to: 1) Assess the joint effect of multiple SNPs that can be linked or unlinked and interact or not; 2) Explore the multi-SNP model space efficiently using the Mode Oriented Stochastic Search (MOSS) algorithm and determine the best models. 3) Include prior information on SNPs to enhance the detection of genetic associations. We illustrate the *R* package through several applications to cancer genetic studies including GWAS, fine mapping studies and gene network analysis.

2800W

Identifying potentially significant causative chromosomal regions from GWAS data. *R.T. Relator, J. Sese.* Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan.

The goal of genome-wide association studies is to find chromosomal positions associated with common, complex diseases. Conventional association tests include single locus analysis, focusing on identifying important mutations related to the disease by assessing them individually. However, especially for common diseases, many non-significant small peaks often appear, and no causative mutations could be found. Here we focus on identifying potential causative "regions" in diseases, which would be overlooked when variants are considered individually. We propose a chromosomal sweeping approach for using the Limitless Arity Multiple-testing Procedure (LAMP). We divide chromosomes into sub-regions while taking into consideration linkage disequilibrium information, and focus on each region for analysis. By doing this, we can exploit all SNP combinations for assessment without sacrificing computational cost or statistical value since LAMP can handle any combinations of SNPs and is known to be less conservative in controlling the family-wise error rate than the usual multiple testing procedures such as Bonferroni correction. We applied our method to postmenopausal breast cancer GWAS data [Nat Genet. 2007 Jul; 39(7):870-4], including 2,287 subjects and around 550,000 SNPs. While the paper identified the importance of the mutations around the FGFR2 gene in chromosome 10 for the disease, our method also detected a wide area in the HLA region in chromosome 6 to have a high relationship to this cancer type. Our technique may contribute to identifying new regions that are possibly relevant to diseases by re-analysis of existing GWAS datasets.

2801T

Assessing clinically relevant variation in 52,000 individuals from over 100 global populations within the Population Architecture using Genomics and Epidemiology (PAGE) study. *E.P. Sorokin[1], G.L. Wojcik[1], N. Abdul-Husn[2], S. Bien[3], N. Zubair[3], P.J. Norman[4], G. Belbin[2], G. Nadharni[5], C. Hodonsky[6], J. Odgis[2], C. Avery[6], S. Buyske[7], T. Matise[8], J. Kocarnik[3], L. Hindorff[9], R. James[10], K.E. North[6], R. Loos[2], C. Haiman[11], C. Kooperberg[3], C. Carlson[3], C.D. Bustamante[1], C.R. Gignoux[1], E.E. Kenny[2].* 1) Genetics, Stanford University, Stanford, CA, CA; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York NY; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA; 4) Department of Structural Biology, Stanford University School of Medicine, Stanford CA; 5) Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York NY; 6) Department of Epidemiology, University of North Carolina, Chapel Hill NC; 7) Department of Statistics & Biostatistics, Rutgers University, New Brunswick NJ; 8) Department of Genetics, Rutgers University, New Brunswick NJ; 9) Division of Genomic Medicine, NHGRI, NIH, Bethesda MD; 10) Division of Clinical Research & Data Management, NIMHD, NIH, Bethesda MD; 11) Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles CA.

During the past decade, major advances in genomics have resulted in the identification of more than 6,000 genetic disorders. However, implementing these findings into clinical care poses a difficult challenge. Vast quantities of rare variants in human populations make questions of causality, penetrance, and expressivity for any particular variant difficult to ascertain. This problem is compounded by theory and observations from population genetics that rare variants are typically population-restricted. Therefore, the characterization of clinically relevant variants (CRVs) in multi-ethnic cohorts is a critical step in advancing global precision medicine efforts. To address this gap in knowledge, the PAGE Study was formed as a multi-center collaboration, to characterize trait-associated genetic variation in ancestrally diverse populations within the United States. Through PAGE-II, we genotyped over 50,000 individuals with ancestry from more than 100 global populations, linked to deep phenotypic data, including diagnostic and procedural codes, radiologic, labs, epidemiologic and disease data. Using the newly-designed Illumina Infinium Multi-Ethnic Genotyping Array (MEGA) including tens of thousands of CRVs, we report population-stratified frequencies for pathogenic variants of ACMG- and pharmacogenomics-prioritized genes, including *BRCA1/2, TP53, PTEN*, and *CFTR*. These include 1,736 ClinVar-annotated pathogenic variants observed more than once in PAGE, and either not found, or monomorphic, in the Exome Aggregation Consortium (ExAC) browser. We observe 27 variants in *BRCA1*, 40 variants in *BRCA2*, and 67 variants in *CFTR*, all of which are annotated as pathogenic in ClinVar and segregating in the PAGE-II populations above a singleton threshold. Additionally, we identify 167 ClinVar-pathogenic alleles at a frequency above 5% in at least one population. We expect the outcome of this work to enable genomic interpretation, assess resources for defining pathogenicity, including ClinVar, ACMG, PharmGKB, and influence current practices around incidental findings and return of genetic findings to study participants. Population-specific allele frequencies of CRVs will ultimately be incorporated into the NHGRI-sponsored PAGE, ClinVar and ClinGen initiatives. Screens of CRVs such as in PAGE provide needed insight into the segregation of variants contributing to clinical conditions worldwide.

2802F

Simultaneous genetic regulatory networks inference and genomic hub identification through conditional graphical models. *C.J. Conley[1], P. Wang[2], J. Peng[1].* 1) Statistics, University of California, Davis, CA; 2) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Models for integrative genomics applications face diverse challenges such as data heterogeneity and limits on scaling dimensionality, computation, and biological interpretation. All these challenges are present while learning genetic regulatory networks from RNA expression data or protein expression data and simultaneously identifying which DNA copy number alterations perturb these networks. For this purpose we developed a sparse conditional graphical model (CG model), called Spacemap, which integrates a response data type (e.g. RNA/protein expression) and a predictor data type (e.g. DNA copy number alterations) through a penalized multivariate regression framework. Specifically, Spacemap infers an undirected network among response nodes in tandem with a directed graph encoding how predictor nodes perturb the responses. Thus, Spacemap predicts the cis- and trans- regulatory structure of genomic perturbations on a response network and points to genomic hubs driving the responses. Spacemap also scales in dimensionality to infer networks with a few thousand of nodes. In a simulation study, Spacemap is shown to improve power in network inference over models where the two types of nodes are not differentiated as well as a likelihood-based CG model. We apply Spacemap to characterize gene regulatory patterns among RNAs and proteins conditioned on DNA copy number alterations using RNAseq and deep protein profiling data of 77 breast cancer tumors from the TCGA-CPTAC studies . In both the inferred mRNA and protein co-expression networks, many known cancer genes involved in cell cycle were found to be directly perturbed by DNA copy number alterations. Specifically, the copy number alterations of a genome region in 5q34 was found to influence a large number of RNAs and proteins in these tumors. Results from this analysis cast new light on the gene regulatory mechanisms in breast cancer.

2803W

SlopeCCA — Sparse CCA with false discovery rate control. *A. Gossmann[1], Y.-P. Wang[2].* 1) Mathematics, Tulane University, New Orleans, LA; 2) Biomedical Engineering, Tulane University, New Orleans, LA.

Canonical correlation analysis (CCA) is a classical statistical technique, which is used to make sense of the cross-correlation of two sets of measurements collected on the same set of samples (e.g., two genomic assays for the same set of cancer cells, or fMRI imaging and DNA sequencing data for the same mental illness patients). More precisely, given two sets of random variables, CCA identifies linear combinations of each, which have maximum correlation with each other. Like many classical statistical techniques, CCA fails in high-dimensional settings, when the number of variables in either of the two cross-correlated datasets strictly exceeds the number of samples. To counteract this issue, several methods for sparse canonical correlation analysis (sparse CCA) have been proposed, where sparse linear combinations of variables from the two datasets are identified to maximize the correlation. However, commonly used sparse CCA methods determine the level of sparsity of the canonical vectors either based on criteria of model fit (danger of overfitting), or according to a preconceived sparsity level (questionable validity). Motivated by Sorted L1 Penalized Estimation (SLOPE), a sparse regression method that offers false discovery rate (FDR) control, we propose a definition of FDR for the canonical vectors, and a sorted l1 penalized CCA (slopeCCA) method. We prove that under certain conditions our method keeps the FDR of the canonical vectors below a user-specified target level. Additionally, FDR control is shown in other less restrictive settings via a simulation study. To our knowledge, this is the first attempt to provably control FDR in a sparse CCA model. As a demonstration of slopeCCA on real genomic data, we apply the method to data from The Cancer Genome Atlas (TCGA), in order to identify significant relationships between gene expressions and miRNA levels on Glioblastoma Multiforme tumor samples.

2804T

Genetic ancestry affects the predictive power of PRS in Latinas. *S.D. Sawyer[1], L. Fejerman[1], E. John[2], L. Kushi[3], G. Torres-Majia[4], D. Hu[1], S. Huntsman[1], C. Haiman[5], E. Ziv[1].* 1) Medicine, University of California, San Francisco, CA., United States; 2) Cancer Prevention Institute of California, Fremont, CA, United States; 3) Kaiser Permanente Northern California, Oakland, CA, United States; 4) Instituto Nacional de Salud Pública, Cuernavaca, Mexico; 5) University of Southern California, Los Angeles, CA, United States.

   Introduction: Genome-wide association studies (GWAS) have identified >90 single nucleotide polymorphisms (SNPs). Individually, these SNPs are weak predictors of breast cancer, but when combined in a polygenic risk score (PRS) they may be useful to identify high risk women whom would benefit from more aggressive screening and prevention. GWAS for breast cancer were mostly conducted in Europeans and PRS based on these results may not perform as well in other populations. We sought to examine and optimize the predictive performance of PRS for breast cancer in Latinas. **METHODS:** We used data from 1679 Latina breast cancer cases and 5743 controls from four studies: San Francisco Bay Area Cancer Study, Northern California Breast Cancer Family Registry, Kaiser Permanente Research Program on Genes, Environment and Heath, Mexico Breast Cancer Study and Health Multiethnic Cohort. Samples were typed on genome-wide SNP arrays and imputed. We included 103 SNPs previously associated with breast cancer ($p < 5 \times 10^{-8}$) in at least one population. The PRS was calculated using a log-additive model. We developed PRS's using both published odds ratios (ORs) and refitting ORs in our own data. Discrimination for each PRS was measured by receiver operating characteristic area under the curve (AUC). **RESULTS:** The PRS based on 75 SNPs (PRS75) using published ORs was significantly associated with breast cancer in Latinas ($p = 2.2 \times 10^{-26}$) and had an AUC of 0.584 (95% CI 0.569-0.600), which was lower than a published AUC of 0.615 in a European study (Mavaddat et al JNCI 2015). An expanded PRS of 103 SNPs (PRS103) using published ORs did not improve the AUC=0.593 (95% CI 0.578-0.608). PRS75 performed better (p=0.035) among Latinas in the high quartile of European Ancestry (AUC=0.611, 95% CI 0.532-0.592) compared to Latinas in the lowest quartile of European (and higher Indigenous American) ancestry (AUC=0.562, 95% CI 0.532-0.592). Next we developed a PRS with the same SNPs, refitting the ORs for Latinas. We calculated the ORs in a training subset (2/3) of our data, and tested the PRS using those ORs in a validation subset (1/3). The discrimination of PRS75 using Latina ORs was AUC=0.585 (95% CI 0.559-0.612) and extended PRS103 was AUC=0.595 (95% CI 0.568-0.621). **CONCLUSION:** PRS using published OR's performs better in Latinas with higher European ancestry. Fine mapping may boost the predictive power of the PRS in Latinas. Results will be used in a trial of breast cancer screening in California.

2805F

A spectrum of common and rare germ-line variation in a group of Puerto Rican women with double and triple negative tumors. *H. Diaz-Zabala[1], N. Arroyo[1], W. Luo[2], M. Dean[2], J. Matta[1], J. Dutil[1].* 1) Ponce Health Sciences University - School of Medicine, Ponce, PR; 2) Laboratory of Translational Genomics, National Cancer Institute, Frederick, MD, USA.

   Introduction: Germ-line variants are known to predispose to sporadic and inherited forms of breast cancer. Rare and common germ-line variants can modify breast cancer risk with different effect sizes and may be the basis to explain susceptibility to molecular subtypes of breast cancer. **Objective:** This study aims to identify the proportion of rare and common germ-line variants that are present in a group of breast cancer cases with triple and double negative tumors. **Methods:** Targeted sequencing with an Ion AmpliSeqä Cancer Hotspot Panel was used to genotype germ-line variants in a group of 52 Puerto Rican breast cancer cases with confirmed triple and double negative pathology reports. Germ-line variant discovery from amplicons was performed using GATK Haplotype Caller. Variant annotation and functional prediction in protein sequence was performed using Golden Helix SNP & Variation Suite (SVS) package. **Results:** A total of 668 germ-line variants was detected in 52 triple and double negative breast cancer cases. Most identified variants were intronic, synonymous and rare with allelic frequencies <1.5% in the general population. Rare variants with damaging effects in protein sequence were identified in *ATM*, *CDKN2A*, *MET*, *EGFR* and *APC*. **Conclusion:** In this study, we identified a spectrum of rare and common germ-line variants in a group of Puerto Rican breast cancer cases with triple and double negative tumors. Predisposition effects of these germ-line mutations need to be addressed in future studies.

2806W

Sequencing for all: An economical workflow for *BRCA* genotyping of breast and ovarian cancer in underserved women from Latin America.
J. Herzog[1], D. Castillo[1], T. Slavin[1], S. Sand[1], R. Mejia[1], C. Villarreal-Garza[2,3], R.M. Alvarez-Gómez[4], T. Wegman-Ostrosky[5], J. Abugattas[6], P. Mora[6], A. Del Toro-Valero[7], Y. Rodriguez[8], M. Cruz-Correa[9], I. Komenaka[10], P. Ashton-Prolla[11], B. Alemar[11], J.N. Weitzel[1]. 1) Clinical Cancer Genetics - Laboratory, City of Hope, Beckman Research Institute, Duarte, California, USA; 2) Departamento de Investigacion y de Tumores Mamarios del Instituto Nacional de Cancerologia, México City, México; 3) Centro de Cancer de Mama del Hospital Zambrano Hellion, Tecnologico de Monterrey, San Pedro Garza Garcia, Nuevo Leon, México; 4) Unidad de Genómica y Secuenciación Masiva, Subdirección de Investigación Básica, Instituto Nacional de Cancerología, México City, México; 5) Instituto Nacional de Cancerología, División de Investigación, México Universidad de Guadalajara, Instituto de Génetica Humana, México; 6) Departamento de Mamas, Partes Blandas y Piel, Instituto Nacional de Enfermedades Neoplásicas, Lima, Peru; 7) Departamento de Fisiología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Jalisco, México; 8) Centro de Oncologia, Clínica del Country, Bogotá, Colombia; 9) Division of Gastrointestinal Oncology, University of Puerto Rico Comprehensive Cancer Center, San Juan, Puerto Rico; 10) Department of Surgery, Maricopa Medical Center, Phoenix, Arizona; Arizona Cancer Center, University of Arizona, Tucson, Arizona, USA; 11) Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

    Germline mutations in the *BRCA1* or *BRCA2* (*BRCA*) gene predispose to an increased lifetime risk of breast and ovarian cancer. In Latin America the prevalence and contribution of *BRCA* mutations to the overall cancer burden is not well understood. This is primarily due to limited access to genetic cancer risk assessment (GCRA) and testing. To address this disparity we have implemented an efficient, cost-effective testing strategy for at risk individuals. Utilizing a combination of Sequenom MassARRAY technology, Ion Torrent semiconductor sequencing, and PCR-based methodologies for copy number variation (CNV) detection, 1,489 individuals have been screened by our laboratory to date. Patients were prospectively accrued from Mexico (Mexico City and Guadalajara), Peru (Lima), Colombia (Bogota), Puerto Rico, and Arizona (Maricopa). A retrospective cohort (n = 66) from Brazil (Porto Alegre) was also included. The overall mutation rate was ~13%. Primary screening with a high-frequency Hispanic mutation Panel (*HISPANEL*) on the Sequenom MassARRAY platform, combined with a 3-primer PCR assay (3-PA) for the recurrent *BRCA1* exon9-12del CNV, yielded ~55% of overall positives (~61% in Mexico) at an estimated cost of ~$20 per case. *HISPANEL* negative cases were fully sequenced for *BRCA* on the Ion Torrent PGM platform, using a commercially available Community Panel pilot tested by our laboratory for the AmpliSeq *BRCA* Global Consortium. *BRCA* sequencing yielded ~35% of all positives at an additional cost of ~$80 per case (materials only). Positive cases were confirmed by Sanger sequencing. Sequence negative cases were further evaluated by Multiplex Ligation-dependent Probe Amplification (MLPA) for *BRCA1* CNVs, at a cost of ~$25 per case. The *BRCA2* gene was not evaluated by MLPA due to the rarity of CNVs in this gene relative to *BRCA1*. CNVs represented ~25% of all positives, with the Mexican founder mutation, *BRCA1* exon9-12del (detected by 3-PA), accounting for ~59% of all CNVs. The remaining ~41% of CNVs were detected by MLPA alone. This is consistent with previous reports of *BRCA* CNVs in Hispanic populations. Data from this ongoing study are being used to construct a population specific variant database delineating mutation frequency, spectrum, and variants of uncertain significance (VUS) by region. This information will be leveraged to customize the *HISPANEL* into a rapid, inexpensive, high-throughput screening tool with high sensitivity for populations beyond Mexico and the U.S.

2807T

Genetic association between ND3 gene polymorphisms and the risk of gastric cancer.
E. Jin[1], J. Hong[1,2]. 1) Clinical Trials Center, Chungnam National University Hospital, Daejeon, South Korea; 2) Department of Pharmacology, College of Medicine, Chungnam National University, Daejeon, South Korea.

    Recently, several studies have been reported that mitochondrial DNA alterations are related with cancer development and progression. In this study, we performed case-control association studies to evaluate the relevance between mitochondrial NADH-dehydrogenase subunit 3 (ND3) gene polymorphisms and the risk of gastric cancer (GC) in the Korean population. A case-control study was conducted with 377 GC patients and 363 controls. We genotyped 6 SNPs such as rs28358278, rs2853826, rs201397417, rs41467651, and rs28358275 in ND3 gene using direct sequencing. χ^2 tests were used to estimate the Hardy-Weinberg equilibrium of each SNP and the association between the GC and control. Binary logistic regression was conducted to estimate the GC risk by odds ratios (OR) and 95% confidence intervals (CI). Variant, rs41467651 was significantly associated with an increased gastric cancer risk of GC and diffuse-type GC (OR = 2.08, P = 0.006 and OR = 2.64, P = 0.002, respectively). Our data suggests that rs41467651 polymorphism of ND3 gene may influence ND3-mediated signaling pathway, particularly in the development of diffuse-type GC in the Korean population.

2808F

Acquisition of clonal driver mutations in VHL syndrome clear-cell renal cell carcinomas. *A.D. Mitchell[1], S.S. Fei[1], C.D. Vocke[2], C. Boniface[1], W.M. Linehan[2], P.T. Spellman[1].* 1) Department of Molecular & Medical Genetics, Oregon Health & Science University, Mail Code: CL6S, 2730 SW Moody St, Portland, Oregon 97201, USA; 2) Urologic Oncology Branch, Center for Cancer Research, National Cancer Institute, Building 10 Room 1-5940, Bethesda, Maryland 20892, USA.

VHL Syndrome is a rare genetic disorder with a dominant mode of inheritance, which predisposes one to benign and malignant neoplasms of varying origin, including renal cell carcinomas. Individuals with VHL Syndrome, who develop renal cancer, commonly present with multiple, malignant clear-cell renal carcinomas (ccRCC). These are synchronous tumors, meaning they initiate, mature and evolve, at the same time and within the same germline environment. Synchronous tumors are an extremely rare occurrence across all types of cancer. Thus, VHL Syndrome ccRCC provides one of the few presentations of cancer to study the somatic genetic features of tumors while controlling for time and environment. We previously observed modest differences in the number of somatic genetic driver events within Fuhrman grade 2 and grade 3 VHL syndrome ccRCC tumors from whole-genome sequencing. The Fuhrman grading system is assigned by pathologist based on the shape and size of the nucleus of tumor cells, as well as the prominence of nucleoli. Tumors are classified into one of four possible grades, or more generally into two grade classes: low and high. High Fuhrman grade ccRCC tumors have an increased ability for metastasis and lower survival rate. Thus, high Fuhrman grade ccRCC tumors are generally more aggressive and more evolved. We have performed resequencing of exomes at ~200X coverage on five grade 3 and five grade 2 ccRCC tumors from three VHL syndrome patients. Resequencing these tumor exomes has provided sufficient power to call variants of low allele frequency across all driver genes. We applied methods to assess clonality within these tumors to determine which SNVs are more likely to be true drivers of tumor evolution with respect to Fuhrman grade. We show that high Fuhrman grade tumors contain more nonsynonymous driver gene variants than low Fuhrman grade tumors. Clonal driver SNVs were only identified in high Fuhrman grade tumors, which suggests driver SNVs may influence the progression of Fuhrman grade in VHL Syndrome ccRCC tumors. .

2809W

Association of variants rs822396, and rs1501299 of *ADIPOQ*, rs1342387 of *ADIPOR1*, rs1862513, of *RETN* and rs35749351of *CAP1* genes with breast cancer risk in Mexican obesity women: Preliminary results. *A. Muñoz-Palomeque[1], M.A Guerrero-Ramírez[2], R.C Rosales Gómez[3], M.G López-Cardona[1,3], J.C Catón-Romero[4], H. Montoya Fuentes[3], T.A García-Cobian[1], S.A Gutierrez-Rubio[1].* 1) Departamento de Fisiología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Colonia Independencia, Guadalajara, Jalisco, Mexico; 2) Dr Valentín Gómez Farías, ISSSTE Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, Zapopan, Jalisco, Mexico; 3) UMAE Hospital de Especialidades del Centro Médico Nacional de Occidente, IMSS Hospital de Especialidades, Instituto Mexicano del Seguro Social, Guadalajara, Mexico; 4) CIBO centro de Investigación Biomédica del Occidente, IMSS Instituto Mexicano del Seguro Social, Guadalajara, Mexico4Unidad de Medicina Genómica y Genética, Hospital.

Clinical studies suggest that breast cancer (BC) and obesity are related, the adiponectin and resistin has been associated with BC, and however, the result has been inconclusive. Various studies suggest that adiponectin has anti-tumour effects but in obesity this protein is decreased, on the other side resistin may act directly by signalling through TLR4 and CAP1, which activates the JAK / STAT and MAPK pathway, proteins directly involved in the initiation, promotion, and progression of tumors (Codoñer-Franch & Alonso-Iglesias 2015; Lee, et al., 2014). In the present study the SNPs rs1862513 of *RETN*, rs822396, and rs1501299 of *ADIPOQ*, rs1342387 of *ADIPOR1* gene and rs35749351 of *CAP1*. Were selected with the aim to determine the association between these variants with BC in Mexican women with obesity. DNA samples from 56 women with breast cancer and 106 healthy women were genotyped. To measure adiponectin, resistin, TNF-α, and IL-6 levels in serum were determined using the following enzyme-linked immunosorbent assay and the reference group. The allelic frequency of the rs35749351 SNP of *CAP1* in the general population were of G allele frequency it was of 399 to 13 for A allele. Furthermore, in reference group, the allele frequency was 185 G alleles and 9 A allele. On another hand, all alleles were G (n=3) in population breast cancer women. In the rs1862513 SNP of *RETN* to general population group, the allele frequency was similar to 154 of the G allele vs. 146 C allele. Were observed allelic frequencies were equal to the frequencies found in the general population frequency of 51% for the G allele and 49% for the C allele. Frequencies allelic in the rs822396 SNP of *ADIPOQ* of 168 individuals from the general population were for A allele 259 and for 77 of the G allele. Also, the allelic frequencies of the reference group that shows the most abundant of 152 for A allele vs. 36 of G allele. Finally, allelic frequencies of the variant for 6 individuals in women with breast were observed 1:1. An assay of ELISA was performed to measure the concentration of TNF-α in serum. Twenty-nine sera from reference and 2 sera from women with breast cancer were analyzed, where differences between the two groups and higher levels of this cytokine in the serum of women with breast cancer were found.   .

2810T

Development of functional assays for identifying *BRCA1* **and** *BRCA2*
mutation carriers. *A. Salmon[1], T. Zehavi-Harel[1], G. Yahav[2], D. Fixler[2], Y.
Shimshon[3], M. Divon-Salmon[3].* 1) Dept Oncology, Hadassah Med Ctr, Jeru-
salem, Israel; 2) 3Bar-Ilan University, Faculty of Engineering and Institute of
Nanotechnology and Advanced Materials; 3) Department of Molecular Biology,
Ariel University, Ariel, Israel,.

   Worldwide, more than one million women are diagnosed with breast cancer
every year, making it the most common female malignancy in the developed
world. Germline mutations in BRCA1 and BRCA2 genes account for 4-6% of
all breast cancer cases and mutation carriers have a lifetime risk of up to 80%
for developing breast cancer and 40% for developing ovarian cancer. From a
clinical point of view, identification of BRCA1 and BRCA2 mutation carriers of-
fers an opportunity to early identify or prevent the development of malignancy;
therefore the ability to determine which women are more likely to carry BRCA1
or BRCA2 mutations is of great importance. The available diagnostic tests
for mutation analysis of BRCA1 and BRCA2 are time- and labor-intensive,
expensive, and do not allow the identification of all the functional mutations.
We utilized various simple, inexpensive and sensitive methods for identifying
BRCA1 and BRCA2 healthy mutation carriers, by using easily obtainable
lymphocyte cells from peripheral blood. First, we used the Fluorescent life
time imaging microscopy (FLIM) method which allows recognizing different
cell populations, in order to distinguish between lymphocytes of BRCA1 and
BRCA2 mutation heterozygous carriers and non-carrier women. Secondly,
using the RNA-seq technology, we performed expression profiling analysis of
normal and BRCA1/2 mutation heterozygous lymphocytes. Our results demon-
strate that cells originated from *BRCA2*-mutation carriers have significantly
lower FLT values compared with *BRCA1* and control cells. Using RNA-seq
analysis we were able to detect distinguished gene expression pattern in
BRCA1/2 mutation heterozygous lymphocytes compared to control. Together,
these two techniques may be utilized in the future to detect BRCA1/2 carriers,
particularly those bearing unknown functional mutations.

2811F

Polymorphism in the *XPD* **gene affects the risk of bladder cancer:**
A case-control study a meta-analysis and TSA. *M. Sankhwar[1], S.N.
Sankhwar[1], S. Bansal[2], S. Rajender[2].* 1) Urology, King George Medical Univer-
sity, Lucknow, Uttar Pradesh, India; 2) Division of Endocrinology, Central Drug
Research Institute, Lucknow, India.

   Purpose: DNA repair of bulky adducts is essential for a normal life. Among
the repair proteins, the XPD protein is interesting because it is a major player
in the nucleotide excision repair pathway and is also involved in transcription
initiation and in the control of the cell cycle and apoptosis. Lys751Gln and
Asp312Asn of XPD gene have been extensively studied in relation to cancer.
We have analyzed these polymorphisms in Indian bladder cancer patients and
conducted a meta-analysis to generate a pooled estimate about the effect of
these variations on bladder cancer. **Material and methods:** The polymor-
phisms was analysed using polymerase chain reaction-restriction fragment
length polymorphism (PCR-RFLP) in 234 case and 258 control samples, fol-
lowed by Chi square test for statistical analysis. The meta-analysis included a
total of 33 studies (twenty one for lys751gln and twelve for asp312asn) studies
comprising of 11,115 cases and 13,800 controls for. Forest plot and funnel plot
were used to interpret statistical significance of association and publication
bias, respectively. Egger's test was performed to confirm funnel plot symmetry
statistically. TSA was performed to validate the findings of meta- analysis.
Results: The results of our case control study showed significant association
between D312N polymorphism and bladder cancer risk (OR-2.85, CI-1.91-
4.25, P<.0001). Smoking and tobacco chewing were found to be a risk factor
for bladder cancer for D312N. XPD gene were not found to be in significant LD
(D'=0.157, LOD=2.66, r^2=0.021) with each other. Haplotypes CG and AA were
found to be significantly higher in the cases (35.7% and 42.1%, respectively)
Meta-analysis results of XPD for the both the polymorphisms also showed
strong association in all the models for D312N dominant (GA+AA vs GG: OR-
1.35, CI-1.15-1.57, P-0.000) and K751Q dominant (AC+CC vs AA: OR=1.12,
CI=1.00-1.25, P=0.03) in overall population as well as Americans and Indian
at higher risk for D312N on ethnicity basis which is further confirmed by TSA.
Conclusion: XPD substitutions increase bladder cancer risk in Indo-European
populations of Uttar Pradesh and also in other populations across the world.

2812W

Frequency of mutations in multi-gene panel testing of 3,600 individuals for hereditary breast and ovarian cancer risk. *A.Y. Zhou[1], J. van den Akker[1], K.A. Smith[1], R. O'Connor[1], S. Kim[1], D.E. DeSloover[1], T. Walsh[2], E. Gil[1], T. Sittler[1].* 1) Color Genomics, Burlingame, CA; 2) University of Washington, Seattle, WA.

Current published guidelines recommend genetic testing for breast and ovarian cancer risk to individuals with a strong personal or family history of these cancers. However, little is known about the genetics of the individuals that fall outside of these restrictive recommendation criteria. Here, we describe the demographics and characteristics of 3,595 individuals who received the Color Test by physician order. In this cohort, 282 pathogenic or likely pathogenic mutations were identified and reported. The proportion of individuals who carried a pathogenic or likely pathogenic mutation (mutation carrier rate) was 7.7% (276/3595; 95% CI: 6.9%-8.6%), similar to reports from other laboratories. Among 1,421 individuals who would not have been recommended for testing under current NCCN guidelines, the mutation carrier rate was 3.9% (95% CI: 3.0%-5.0%). Among 561 Medicare-eligible individuals, 455 would not have met current recommendations for testing under Medicare guidelines. This cohort presented with a mutation carrier rate of 6.2% (95% CI: 4.3%-8.8%). Among 434 Ashkenazi Jewish individuals, 42 pathogenic or likely pathogenic mutations were identified and almost 50% of these mutations (20 of 42) were outside the three known Ashkenazi Jewish founder mutations in BRCA1 and BRCA2. The mutation carrier rates, reported here, in cohorts of individuals who would not have qualified for current published guidelines for testing is much higher than the presently estimated rate in the general United States population. Additionally, the number of pathogenic mutations identified outside of the three Ashkenazi Jewish founder alleles in the Ashkenazi Jewish cohort indicates that testing for just these three alleles is insufficient in this population. Taken together, the results of this study demonstrate the importance of broadening the offering of genetic testing for all populations.

2813T

Germline *CDH1* mutations are a dominant cause of diffuse gastric cancer in the New Zealand Māori population. *C. Hakkaart[1], S. Chen[2], R. Day[1], L. Ellison-Loschmann[2], M. Gray[2], P. Harawira[3], J. Koea[4], N. Pearce[5], A. Sporle[6], T. Whaanga[2], P. Guilford[1].* 1) Biochemsitry Department, University of Otago, Dunedin, New Zealand; 2) Centre for Public Health Research, Massey University, Wellington, New Zealand; 3) Kimihauora Health Centre, Tauranga, New Zealand; 4) Waitemata District Health Board, Auckland, New Zealand; 5) London School of Hygiene and Tropical Medicine, London, England; 6) Department of Statistics, University of Auckland, Auckland, New Zealand.

Gastric cancer is a multifactorial disease with a complex interplay of both inherited and environmental factors contributing to its emergence. The two main histological subtypes of gastric cancer are: intestinal, which is more common and is closely associated with environmental factors; and diffuse, which is less common and is though to have a strong genetic component. In New Zealand, Māori have a notably elevated rate of gastric cancer compared to non-Māori and are one of the few populations worldwide with an overall higher incidence of the diffuse type of the disease. Genetic predisposition to gastric cancer has been well documented in New Zealand with multiple families being diagnosed with the inherited cancer syndrome Hereditary Diffuse Gastric Cancer (HDGC), which is caused by germline mutations in the tumour suppressor gene *CDH1*. *CDH1* mutation carriers have 70% lifetime risk of developing diffuse gastric cancer and females have an additional 40% risk of developing lobular breast cancer. As part of a study examining the known risk factors of gastric cancer in Māori we are using next generation sequencing and MLPA to search for germline mutations in germline *CDH1*. Our aim is to determine the prevalence of pathogenic *CDH1* mutations in the Māori gastric cancer population and understand the impact this is having on the rate of gastric cancer in New Zealand. We have sequenced blood DNA collected from an unselected series of 94 Māori gastric cancer patients diagnosed between 2009 and 2013 and more than 204 healthy matched controls. 18% of all patients and 34% of patients with the diffuse subtype carried pathogenic germline mutations in *CDH1*. This incidence increased to 66% for diffuse gastric cancer patients aged less than 45 years. These results show genetic predisposition to gastric cancer is a dominant cause of early onset diffuse gastric cancer in New Zealand Māori and that this is a major contributor to the high incidence relative to New Zealand Europeans.

2814F

A computational framework for genotyping and early detection of lung cancer. *E. Adetiba[1,2], O.O. Olugbara[1,2].* 1) Department of Information Technology, Durban University of Technology, P.O. Box 1334, Durban, 4001, South Africa; 2) ICT and Society Research Group, Durban University of Technology, P.O. Box 1334, Durban, 4001, South Africa.

A wide ranging and collaborative efforts of scientists and researchers from around the world have successfully unravelled the human genome sequences. The completed human genome sequences are about three billion characters in length. In addition, the sequences of other organisms such as virus, bacteria, archaea, fungi, mouse and several plants have been deciphered. These efforts have generated huge biological data which keep growing at an astronomical rate. Meanwhile, the analysis and interpretation of these genomic data are major challenges facing researchers and scientists in fields as diverse as information engineering, computer science, physics, mathematics, molecular biology, medicine, pharmacy, agriculture and genetics. This is consequent on the big prospects that such interpretation holds for human development in areas such as creation of new diagnostic tools, innovation of therapeutic procedures, design of novel drugs and production of genetically modified crops. We report and discuss the preliminary results of the prototype of a computational framework we developed for genotyping, screening and early detection of lung cancer mutations. To develop the framework, we extracted lung cancer biomarker genes from the COSMIC database. The extracted genes have been implicated to carry various mutations (such as deletion and substitution) that are responsible for the initiation and advancement of lung cancer. Some of these genes are tumor suppressor p53, kirsten rat sarcoma viral oncogene homolog, epidermal growth factor receptor, cyclin-dependent kinase inhibitor 2A, lysine-specific methyltransferase 2C and host of others. The computational components of the framework are Z-curve affine transform, histogram of oriented gradient and Gaussian radial basis function neural network. The neural network was trained and validated using the extracted features from the biomarker genes. A graphical user interface was developed for ease of use and the experimental results show a high classification accuracy of lung cancer genomes into the appropriate biomarker genes and mutation types. The computational framework can be interfaced with next generation sequencing platforms like Illumina MiSeq to detect lung cancer mutations in at risk persons and those with an early onset of the disease. This will assist oncologists to make timely recommendations for targeted molecular therapies and ultimately reduce the rate of mortality due to lung cancer.

2815W

Recurrent somatic copy number variation analysis identifies risk genes that modulate the survival of young women with breast cancer. *C. Chi[1], R. Ajwad[1,2], Q. Kuang[1], L.C. Murphy[1,3], P. Hu[1,4].* 1) Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada; 2) Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada; 3) Research Institute in Oncology and Hematology, Winnipeg, MB, Canada; 4) Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada.

Breast cancer (BC) diagnosis in young women (<45 years old) has come forth as an independent factor with higher recurrence risk and death than their older counterparts, and it has been suggested that it may exhibit its own unique biology. Copy number variations (CNVs) have led many to consider them as an alternate paradigm for the genetic basis of human diseases, as these large variations may encompass key genes that contribute to carcinogenesis and disease progression. Although many complex diseases have been linked to CNVs in the genomic DNA, prior studies have yet to document age-related changes in somatic CNVs for young women with BC. We hypothesize that recurrent somatic CNV regions uniquely found in young women with BC will harbor cancer-susceptibility genes that modulate the survival of these young women. We aim to find recurrent somatic CNV regions that are identified from BC microarray data and associate the CNV status of the encompassed genes to the survival of young women with BC. We have developed a new interval graph-based algorithm for identifying recurrent somatic CNVs in cancer using a maximal clique detection technique. The algorithm guarantees that the identified CNV regions are the most frequent and identifies the delineated minimal regions. By applying the algorithm to the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) CNV data consisting of 2000 breast tumour samples (equally divided into a Discovery set with 130 young women and a Validation set with 125 young women), we have identified 5 validated recurrent CNV regions (three amplifications and two deletions) that are uniquely found in young women with BC. There regions encompass 6 corresponding candidate genes (amplification regions: *AMY1A*, *KIAA1267*, *SIRPB1* and deletion regions: *OR4F5*, *GSTM1*, *GSTM2*). Interestingly, the frequency at which these CNVs occur ranges from 8.5% (*AMY1A*) to 30.8% (*SIRPB1*) in the Discovery set and from 10.4% (*AMY1A*) to 46.4% (*SIRPB1*) in the Validation set. Overall, the frequency of these genes between the young group and the old group is significantly different (p value < 5e[-10]). The amplification status of gene signature (*AMY1A*, *SIRPB1*) in young women with BC is associated with disease-specific survival in the Discovery set (hazard ratio=1.78; p value=0.058). Together, identification of the deletion and amplification events that may be prognostic in young women with BC can be used in genomic-guided treatment.

2816T

Low memory, fast, specific, sensitive, multi-reference sequence classification using Bloom filter maps. *J. Chu[1,2], H. Mohamadi[1,2], S. Yeo[2], B. Vandervalk[1,2], G. Jahesh[1,2], C. Yang[1,2], S. Jackman[1,2], R. Warren[2], I. Birol[1,2,3].* 1) The University of British Columbia, Vancouver, BC, Canada; 2) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada; 3) Simon Fraser University, Burnaby, BC, Canada.

Background Sequence classification is traditionally performed by alignments of sequencing reads onto a reference sequence set. Although alignment methodologies have the potential to map the location of these reads precisely, this information is not a prerequisite for classification and thus perform more computation than is needed. To address these shortcomings, we previously proposed an efficient classification method, BioBloom Tools (BBT) (Chu et al. 2014), that uses a low memory, probabilistic set membership query data structure called a Bloom filter. Using a Bloom filter, sequences are queried to determine whether they are or are not members of a set of sequences composed of a list of references. The accuracy and speed of BBT has proved useful towards mass screening of whole genome shotgun data from human tumors and was instrumental studying large cohorts, such as the Cancer Genome Atlas (TCGA) project on genomic characterization of papillary thyroid carcinoma (Agrawal et al. 2014). Though effective when using a small number of filters, determining the specific reference of a queried sequence requires inefficient usage of multiple Bloom filters. Here we present a new data structure we call a Bloom map, which can efficiently associate queries to a specific reference. Methods A Bloom filter works by implicitly storing sequences by obtaining a set of numbers derived from a nucleotide sequence (via a process called "hashing"), rather than storing the sequence itself. These numbers are used to set bit locations in an array from 0 to 1. To determine if a sequence is part of a filter we hash the sequence and lookup these locations to see if they are set. Conceptually, a Bloom map is a Bloom filter that contains not just presence of a set bit but an index identifier for each bit set, allowing us to look up a corresponding array that stores which genome this sequence originally came from. Results We benchmarked the performance of our tool on a simulated metagenomic dataset mimicking 1 million 2x150bp Illumina reads from all the complete genomes within the NCBI bacterial database. We correctly classified the genus of 97.3% reads utilizing 61 GB of RAM in under 2 minutes using 8 CPU cores. Conclusions Here we present an efficient low memory multi-reference sequence classification tool with broad applications contamination screening, pathogen detection, metagenomics, and preprocessing for targeted assembly from shotgun sequence data.

2817F

Resolution of complex genomic structures arising from integrated duplication, inversion, and translocation events with barcoded linked-read sequencing. *S. Greer[1], B. Lau[1], J. Chen[2], C. Xia[2], H. Ji[1,2].* 1) Stanford Genome Technology Center, Stanford University, Stanford, CA; 2) Dept of Med/Oncology, Stanford University, Stanford, CA.

Structural variants (SVs) represent an important source of somatic variation in cancer genomes, often affecting critical oncogenes and tumor suppressors to become tumorigenic. Despite the large-scale availability of whole genome sequence (WGS) data, accurate detection of SVs remains a challenge. Moreover, the resolution of complex SV structures is still near impossible. We recently demonstrated the use of linked-read sequencing (LRS) to resolve haplotypes of both normal and cancer genomes (GX Zheng, BT Lau, et al. Nature Biotechnol. 2016). This phasing technology uses droplet-partitioned barcodes to tag high molecular weight DNA molecules over 150 kb in size, enabling the resolution of haplotypes spanning multiple megabases. In this study, we used LRS to resolve the structure of complex SVs that are undetectable with traditional WGS. With LRS, SVs that bring together non-adjacent genomic elements will share the same droplet partition barcode and thus can be directly detected and robustly quantified. We performed LRS on two ovarian (left and right) metastases from a patient with diffuse gastric cancer. Detection of SVs in these samples from traditional WGS data yielded many events with limited accuracy and unknown linkage. In contrast, LRS data provided a means to resolve the structures of oncogenic SVs. In the right metastasis, we were able to temporally link multiple SV events into a single resolved structure in the *FGFR2* locus. This rearrangement involved a 30 Mb deletion, followed by inversion-translocation of the deletion-flanking regions, and subsequent tandem duplication of the remaining 1 Mb region encompassing the previous events. We performed this analysis on the left metastasis and also resolved in the *FGFR2* locus an intrachromosomal translocation event coupled with a tandem duplication, but at a location distinct from that of the right metastasis. The unique SV structures between metastases was clear with LRS data but would have been impossible to resolve using traditional WGS, highlighting the utility of LRS as a revolutionary method for SV analysis.

2818W

Improved detection of genetic variants through a non-aligning k-mer approach. *L.L. Hansen, A. Fire.* Pathology, Stanford, Stanford, CA.

Next generation sequencing (NGS) has impacted virtually all aspects of molecular biology. One prominent application of NGS technology is the sequencing of whole genomes or targeted genomic regions (exomes), with the objective of identifying genetic variants. The flood of sequencing data being applied to variant discovery has created a corresponding need for computational tools that can accurately detect variants of all types and sizes. In order to obtain a complete picture of variation, current approaches rely on complex computational pipelines combining different methods, each restricted to identifying a particular type or size of variant. Even so, some types of variants remain difficult to call, too large for small indel callers and too small for structural variant callers. In addition, the majority of existing techniques require the alignment of reads to a reference genome prior to calling variants, adding to computational expense and introducing potential alignment biases. To address these issues, we have developed a method to call variants that allows for direct comparison of sequence library(ies) with a reference genome or to each other without requiring the alignment of reads to a reference genome. Our approach allows us to detect in one method a large subset of variants, including small single nucleotide variants (SNVs), small indels, and large structural variants regardless of size. This includes medium sized structural variants (15-100 bp), which traditionally have been especially difficult to detect using currently available approaches. We are also able to identify both germline and somatic variants. Briefly, our method is based on comparative indexing of DNA words unique to one read library compared to another and/or to a reference genome. This allows us to identify reads that share the same variant because they share the same unique word(s). We then assemble reads containing the same variant into a contig, and align the contig to the reference genome, allowing us to accurately call the variant type and position. Our method will work on any variant that generates unique sequence(s), independent of variant size or type. Performance tests on gold standard datasets indicate our method outperforms popular existing small indel and SNV callers and in addition will call all sizes and types of structural variants.

2819T

Clinically actionable genetic variants in cancer-predisposing genes: A survey of 300 patients with whole-genome sequencing and lifetime electronic health records. *M.M. He[1,2,3], K.Y. He[4], Q. Li[5], K. Wang[5,6].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; 2) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; 3) Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; 5) Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA 90089, USA; 6) Department of Psychiatry, University of Southern California, Los Angeles, CA 90089, USA.

Statement of purpose: Discovery of clinically actionable genetic variants across cancer-predisposition genes using whole-genome sequencing (WGS) data and electronic health records (EHRs) may improve our understanding of carcinogenesis, direct personalized patient care, and facilitate genetic counseling for patients and families. Our objective is to identify clinically actionable germline mutations in cancer-predisposing genes using patients' WGS data and comprehensive clinical data derived from their EHRs. Methods used: We analyzed genetic mutations in 60 autosomal dominant cancer-predisposition genes in 300 deceased patients with nearly complete long-term (over 30 years) medical records. Germline mutations in these 60 genes are associated with autosomal dominant cancer-predisposition syndromes [1]. To infer biological insights from massive amounts of WGS data and comprehensive clinical data in a short period of time, we developed an in-house analysis pipeline within a software framework called SeqHBase [2] to quickly classify genetic pathogenicity based on the latest ACMG guidelines [3]. The clinical data of the carriers of pathogenic and/or likely pathogenic variants were reviewed to assess their clinical conditions using their lifetime EHRs. Summary of results: Among the 300 participants, 7 (2.3%) carried a presumed pathogenic or likely pathogenic variants (including 3 novel loss-of-function variants) in one of the 60 cancer-predisposing genes. When assessing extensive clinical data, each of the 7 patients had one or more different cancers, which are fully consistent with their genetic profiles. This work demonstrates that genetic mutations in autosomal dominant cancer-predisposing genes could be potentially used in clinical diagnosis, prevention, and personalized treatments, showcasing the power of combining WGS and EHR to improve healthcare. Reference: 1. Zhang et al: Germline Mutations in Predisposition Genes in Pediatric Cancer. The New England journal of medicine 2015, 373(24):2336-2346. 2. He et al: SeqHBase: a big data toolset for family based sequencing data analysis. Journal of medical genetics 2015, 52(4):282-288. 3. Richards et al: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics 2015, 17(5):405-424.

2820F

Inferring expressed genes by whole-genome sequencing of plasma DNA.
E. Heitzer[1], P. Ulz[1], G.G. Thallinger[2,3], M. Auer[1], K. Kashofer[4], S.W. Jahn[4], L. Abete[4], G. Pristauz[5], E. Petru[5], J.B. Geigl[1], M.R. Speicher[1]. 1) Institute of Human Genetics, Medical Unversity of Graz, Graz, Austria; 2) Institute of Molecular Biotechnology, Graz University of Technology, Graz, Austria; 3) Bio-TechMed OMICS Center Graz, Graz, Austria; 4) Institute of Pathology, Medical University of Graz, Graz, Austria; 5) Department of Obstetrics and Gynecology, Medical University of Graz, Graz, Austria.

The analysis of cell-free DNA (cfDNA) in plasma represents a rapidly advancing field in medicine. cfDNA fragments have been associated with the release of DNA from apoptotic cells after enzymatic processing since the distribution of their lengths has a mode near 166 bp in most analyses, a size which corresponds approximately to the DNA wrapped around a nucleosome (~147 bp) plus a linker fragment (~20 bp). Recently evidence that cfDNA reflects nucleosome footprints was reported. Therefore, we investigated whether plasma DNA is able to reflect expression-specific nucleosome occupancy at promoters. Furthermore, we assessed if plasma DNA possesses the sensitivity and accuracy to predict whether genes are expressed or not, and furthermore, we determined if blood samples from patients with cancer are informative for expressed cancer driver genes. We performed whole-genome sequencing (WGS) of plasma DNA and identified two discrete regions at transcription start sites (TSS) where the nucleosome occupancy results in different read-depth coverage patterns in expressed and silent genes. By employing machine learning for gene classification, we found that the plasma DNA read depth patterns from healthy donors reflected the expression signature of hematopoietic cells. In cancer patients with metastatic disease, we were able to classify expressed cancer driver genes in regions with somatic copy number gains with high accuracy. We could even determine the expressed isoform of genes with several TSSs as confirmed by RNA-Seq of the matching primary tumor. Taken together our data suggest that read depth analyses of plasma DNA can reveal functional data such as the expression status of genes due to the nucleosome occupancy pattern at promoter regions and, furthermore, that even the expression status of cancer-related genes can be deduced from the blood of patients with cancer. Due to the increasing application of molecularly driven therapeutics, which rely on accurate and timely measurements of critical biomarkers, our results are of utmost significance, as they provide a new view on the genomes of the cells which release their DNA into the circulation. Furthermore, our approach is applicable during a disease stage at which most clinical studies in oncology are conducted. This significantly expands upon the currently existing options for cfDNA analysis. .

2821W

Efficient human genome sequence assembly with ABySS v2. *S.D. Jackman, B.P. Vandervalk, H. Mohamadi, J. Chu, R.L. Warren, S. Yeo, L. Coombe, G. Jahesh, A. Hammond, I. Birol.* BC Cancer Agency Genome Sciences Centre, Vancouver, BC, Canada.

Assembling genome sequencing reads into larger contiguous sequences before comparison to a reference genome improves variant calling in highly polymorphic loci, such as HLA, and in identifying variants that are difficult to infer directly from aligned short-read sequencing data, such as large indels and structural variants. ABySS was the first genome assembler to scale to assembling the human genome using Illumina sequencing. Since then read lengths on the platform have quintupled from 50 bp to 250 bp. Paired-end sequencing of a typical library of ~500 bp DNA fragments with two 250 bp reads fully describes the sequence of the fragments, and so one can derive 500 bp pseudo-reads from the Illumina platform. Assembly software must continue to evolve to fully benefit from these improvements in sequencing technology. ABySS v1 required a terabyte of memory to assemble a human genome and employed the message passing interface (MPI) to aggregate memory across multiple machines. In the version we are reporting here, ABySS v2, we have implemented algorithms that employ a Bloom filter, a probabilistic data structure, to represent the sequence content of experimental data to be assembled. This data structure reduces the memory footprint of ABySS v2 by ten fold, making it possible to assemble a human genome in 100 GB of RAM, which is often available on a single server. As read lengths increase, the optimal value of the fundamental assembly parameter k, which describes the minimum required overlap between reads for assembly, also increases. Therefore, the memory footprint of assembly algorithms that use typical data structures, such as a hash table, increases linearly with k. The memory use of a Bloom filter, on the other hand, is independent of k, allowing ABySS v2 to scale with increasing read lengths with no increase in required memory. Here, we demonstrate the assembly of a human genome using 100 GB of RAM with a scaffold NGA50 of 4.4 Mbp using a single library of each Illumina paired-end and mate-pair sequencing. We also investigate the use of 10x Genomics Chromium data to further improve the scaffold contiguity of this assembly. We note that routine assembly of highly contiguous human genomes using high-throughput sequencing data with moderate computing resources presents an enabling technology for personalized medicine applications. These include analyzing structural variations in highly rearranged cancer genomes at a resolution not afforded by alternative methods.

2822T

Inferring clonal sequences and phylogenies from personal tumor genome profiles. *S. Kumar, K. Gomez, O. Murrillo, S. Miura.* Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA.

   Tumor progression involves evolution of clone cells that arise through cell division, mutate and spread in tumors. Genetic profiling has revealed the presence of extensive variation in tumor samples from individual patients; however, existing methods have difficulty in estimating clone genotypes and clone frequencies from heterogeneous tumor samples. We show that the clone cell genotypes present in each tumor sample can be accurately inferred from the analysis of multiple tumor sample profiles with our new CloneFinder method. CloneFinder uses molecular evolutionary principles to deduce clone genotypes at a single-nucleotide resolution along with the frequency of each clone in each sample. It performs better than existing methods and is scalable for the analysis of a large number of samples. Application of CloneFinder to more than 60 empirical datasets revealed that a vast majority of tumor samples contain at least one ancestral clone, that ancestral clones occur at high frequencies, and that many tumors are a mixture of very early and recently evolved clones. Therefore, our new method provides new insights into the clonal structure of tumors and their evolution within a personal life time.

2823F

Prediction of prostate cancer recurrence by voting feature intervals. *M. Leclercq[1,2], A. Bergeron[1], Y. Fradet[1], A. Droit[2].* 1) Centre de recherche du CHU de Québec-Université Laval; 2) Centre de recherche du CHU de Québec - Université Laval, Médecine moléculaire, Faculté de médecine, Université Laval.

   Introduction: The emergence of high-throughput genomic technologies in the recent years has allowed a better understanding of the genomic landscape of prostate cancer (PCa). The analyses of hundreds of prostate tumors showed important genomic heterogeneity. However despite these extensive studies, the prognostication of prostate tumors has not yet been greatly improved. There are accumulating evidences that a detailed analysis of the intra-tumoral immune response could help to the stratification and prognostication of tumors. The objective of this study was to apply a machine learning approach to identify an immune signature that could help predict PCa progression. **Methods**: Tumor mRNAseq and clinical dataset of 489 PCa patients from The Cancer Genome Atlas project were used for these analyses. The tumor infiltration by immune cells was assessed using the CIBERSORT approach. The infiltration of immune cells along with the expression profiles of 1418 immune response-related genes were analyzed with clinical data by machine learning approaches to identify features able to predict biochemical recurrence (BR) after surgery. **Results**: Using correlation-based feature subset selection we identified 5 genes (*PPBP, KEL, HRH4, CEACAM6* and *CXCL11*), two clinical features (Gleason grade and pathologic T) and the abundance of a subset of immune cells (gamma delta T cells), that have an individual predictive ability with low inter-correlation to predict BR. Supervised learning by voting feature intervals classifier on these features predicted BR at 70% accuracy (AUC: 75.8%, 10 fold cross validation) with a 71.8% specificity and 69.4% sensitivity. While some other classifiers provided better accuracy, none provided a sensitivity higher than 40%. **Conclusion**: We believe that these results will help to better understand the spontaneous immune response against PCa, critical for the design of new emerging immunotherapies based on highly specific biomarker combination.

2824W

Variant EXplOreR : Integrative environment for functional understanding of complex trait susceptibility loci. Application in breast cancer fine-mapping analysis. *A. Lemaçon[1], J. Allen[2], P. Soucy[1], J. Beesley[3], P. Kraft[4,5], G. Bader[6], A. Dunning[7], K. Michailidou[2,7], G. Chenevix-Trench[3], R. Milne[9,10], K.B. Kuchenbaecker[2,11], A.C. Antoniou[2], D. Easton[2,8], J. Simard[1], A. Droit[1], BCAC, CIMBA.* 1) Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Lav, Quebec, Quebec, Canada; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 4) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 5) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 6) The Donnelly Centre, University of Toronto, Toronto, Canada; 7) Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 8) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 9) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global health, The University of Melbourne, Melbourne, Australia; 10) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia; 11) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

Recent pan-genomic studies of common genetic variations conferring susceptibility to breast cancer have revealed over 180 independent loci, the vast majority located in the non-coding genome. For the purpose of determining the biological relevance of those risk loci in breast cancer risk prediction, we want to finely characterize each locus. In the context of our collaborations within the Breast Cancer Association Consortium (BCAC) and the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA), approximately 570K have been genotyped using the OncoArray (a custom Illumina chip which includes 260K "backbone" SNPs providing a genome-wide coverage of most common variants, together with SNPs of interest), leading to the imputation of 22 million genetic variants. These association results have been merged in a single file also including, for each variant, epidemiological data and the overlap status with more than 130 bio-features (introns, exons, TFBS...) used as potential genomic causality predictors. The huge dimension of the generated matrix (more than 3 billion entries), renders it impossible to manually handle all these valuable data. Therefore, we developed a bioinformatic tool named VEXOR, for Variant EXplOreR, which is designed to address this big data challenge. VEXOR is a platform-independent browser-based integrative environment for functional annotation in R, based on the Shiny package. This interface provides a comprehensive analytical framework to characterize the role of variants driving susceptibility signals in regions defined by GWAS. The matrix is connected to 1) publicly available functional annotations through UCSC Browser, 2) genetic diversity information (1000 Genomes third phase MAF, Linkage Disequilibrium) 3) effect prediction tools (VeP, HaploReg), 4) functional scoring tools (CADD, RegulomeDB) 5) prioritization algorithm (PAINTOR), 6) pathway information (GeneMANIA) 7) regulation data (Enhancers, Super-enhancers, IM-PET) and finally 8) relevant experimental data such as chromosomal interactions (Hi-C, 3C, 5C, ChIA-PET) and expression data. Thus, the user is provided with a proximal and distal context for each variant to assist in the prediction of putative functional effects. VEXOR is a versatile and scalable tool designed to help the understanding of the functional context in fine-mapping analyses of complex traits. It is now available to the scientific community on the Compute Canada Portal (http://romix.genome.ulaval.ca/vexor).

2825T

Identification of regulatory non-coding cancer drivers based on functional impact. *E. Liu, A. Fundichely, P. Dhingra, E. Khurana.* Weill Cornell Medical College, New York, NY.

The majority of variants identified in cancer whole-genome sequencing (WGS) data are located in non-coding regions. Mutation recurrence is a strong indicator of functional cancer driver and highly recurrent mutations have been identified in non-coding regions, such as the TERT promoter and the long noncoding RNA, MALAT1. However, the lack of large sample sizes and inaccurate background mutation rate estimation decrease the statistical power for recurrence-based methods to differentiate cancer driver mutations from passenger mutations in non-coding regions. We have developed a computational method to integrate the signals from variants with high functional impact with the recurrence of variants across multiple tumor samples while controlling for multiple possible covariates, such as, replication timing, histone modification marks, and open chromatin regions, to identify candidate non-coding drivers. The functional impact score is a weighted scheme that uses the properties of ENCODE elements (including nucleotide sequence conservation, transcription-factor (TF) motif disruption or gain, and network properties). By analyzing mutations in 14 different tumor types encompassing 1161 tumor WGS samples, we not only confirm well-known functional non-coding hotspots, but also detect several novel non-coding candidate elements that may target similar oncogenic pathways in the tumor samples without known functional cancer mutations. This analysis may help match patients to suitable clinical trials as well as provide crucial biological insights necessary for the translation of noncoding variants to the clinic.

2826F

Pan-cancer analysis using public research datasets can yield new clinical insights into the care of pediatric cancer patients. *O. Morozova[1], A. Tai-Shah[2], Y. Newton[1], S. Salama[1,3], I. Bjork[1], T. Goldstein[1], A. Resnick[4], S. Mueller[4], E.A. Sweet-Cordero[2], D. Haussler[1,3].* 1) Genomics Institute, University of California Santa Cruz, Santa Cruz, CA; 2) Division of Hematology and Oncology, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; 3) Howard Hughes Medical Institute, Chevy Chase, MD; 4) Pacific Pediatric Neuro-Oncology Consortium (PNOC).

Precision medicine programs are entering clinical care for adult cancer patients, but their application in pediatrics has been more limited. To address the need for innovative approaches to precision medicine in pediatric oncology, we launched California Kids Cancer Comparison (CKCC), a demonstration project for the California Initiative to Advance Precision Medicine. CKCC evaluates the efficacy and potential clinical utility of real-time pan-cancer genomics analysis, as applied to the most difficult-to-treat pediatric cancer patients, treated prospectively on clinical genomics trials at California hospitals. In our pan-cancer approach, each patient's RNA-sequencing profile and/or mutational profile is compared to over 10,000 tumor profiles from public research datasets, such as The Cancer Genome Atlas (TCGA) and Therapeutically Application Research to Generate Effective Treatments (TARGET) projects. The comparisons are conducted using our in-house Tumor Map tool, which examines nearest neighbors of each individual tumor to reveal tumor-specific pathway perturbations. We also implemented a gene expression outlier analysis that can identify pathways and genes with transcripts that are significantly over or under represented in a given patient, as compared to either the whole pan-cancer reference cohort or tumors that are most closely related to the patient's cancer based on the Tumor Map. Together, our analysis approaches are used to identify new therapeutic directions that were not previously apparent. These therapeutic directions can be further investigated by the medical teams, and confirmed using CLIA tests. Since April 2016, we have analyzed data from 12 patients, treated at University of California San Francisco and Stanford. In 12 out of 12 cases, our pan-cancer analyses found potential additional therapeutic directions. These therapeutic directions included existing clinical trials or FDA-approved drugs that could be administered off-label at the discretion of the physician. In 3 of these cases, all at Stanford, the pan-cancer analyses, together with CLIA certified validation tests, have provided key insights that have guided therapeutic decisions. Our study provides a framework of how very large, public genomic datasets collected from adult and pediatric patients in research and clinical settings can be used to provide new therapeutic opportunities for individual pediatric patients prospectively.

2827W

A *de novo* assembly pipeline to detect structural variants from RNA-seq data in clinical genomics. *K.M. Nip[1,2], R. Chiu[1], J. Chu[1,2], H. Mohamadi[1,2], G. Jahesh[1,2], I. Birol[1,3].* 1) Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada; 2) Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

High-throughput RNA sequencing (RNA-seq) is being vigorously explored as a clinical genomics tool. It is widely used for gene fusion detection and gene expression analysis. Circumventing the short read length of RNA-seq data, de novo assembly allows reliable detection of long-range rearrangements and facilitates capturing the exact sequence of the rearranged transcriptomic product. We present a bioinformatics pipeline, TAP, that performs assembly-based detection of transcript-level rearrangements. Since most clinical applications involve a known gene set that are actionable or confer knowledge about a disease of interest, TAP can be utilized to operate on a user-defined gene list to ensure a quick turnaround. TAP consists of three stages: (1) de novo assembly, (2) alignment to reference, and (3) variant detection. First, read pairs are binned with BioBloomTools for each gene of interest. Then, these read sets are assembled in parallel with Trans-ABySS. The assembled transcripts are then aligned to the reference genome with GMAP. Finally, PAVFinder is used to identify variants such as gene fusions, indels, internal and partial tandem duplications (ITDs and PTDs), and alternative splicing such as novel exon skipping. TAP's performance was assessed with the Illumina RNA-seq data of AML patients from the LEUCEGENE project. In our first test set of 4 samples, TAP was able to detect all *NUP98-NSD1* fusions and *FLT3*-ITDs of various sizes. In our second test set of 2 samples, TAP was able to detect *RUNX1-RUNX1T1* and *CBFB-MYH11* fusions and *FLT3*-ITDs. Some cancer genes involved in gene fusions are known to be promiscuous with different partners. One of the features of TAP is its ability to detect fusion events as long as one gene partner is present in the target list. We tested this feature by running TAP using a single *KMT2A* Bloom filter on 9 representative samples, each carrying a *KMT2A* fusion with a different gene partner. TAP was able to detect all the gene fusions. The RNA-seq data in our test set contains over 100 million read pairs. Our target gene list consists of 595 genes from the COSMIC Cancer Gene Census. The runtime for the TAP pipeline on the test sets is around 2 hours on average using a machine with 24 CPUs and 48GB of RAM running CentOS 6. We conclude that TAP's runtime and accuracy are suitable for potential clinical genomics applications.

2828T

ChARMDiff: **Discovery of combinatorial chromatin state difference in multiple cell types and conditions.** *S. Park, S. Kim.* Department of Bioinformatics and Life Science, Soongsil University, Seoul, R.O.Korea.

Various chromatin modifications, identified in large-scale epigenomic analyses, are associated with distinct phenotypes of different cells and disease phases. To improve our understanding of these variations, many computational methods have been developed to discover novel sites and cell-specific chromatin modifications. However, the discovery of combinatorial patterns of differential chromatin modifications across tissues, cell types, and disease phases, is a non-trivial task and remained unaddressed. In this regard, we report *ChARMDiff*, a new computational approach based on association rule mining, which is pattern discovery of *de novo* differential chromatin modifications and characterize globally occurred patterns of combinatorial chromatin state difference between multiple cell types and conditions. By applying *ChARMDiff* to two pairs of epigenomes in normal and cancer cells i.e., GM12878 and K562 from ENCODE and from normal and hepatocellular carcinoma tissues of hepatitis B virus X -transgenic mice. We find that genes in distinctive chromatin modification pattern are enriched in the significant canonical pathways accounted for the pathogenesis of leukemia and human hepatocellular carcinoma and are differentially expressed. *ChARMDiff* provides a scalable framework that can easily be applied to find various levels of combination patterns, which should reflect a range of globally common to locally rare chromatin modifications. Our approach provides new insights into the resolution of the histone code hypothesis to characterize epigenetic variations in distinct phenotypes of different cells, disease phase and experimental conditions.

2829F

Analysis of genetic variation data and its relationship to protein sequence and 3D structures. *A. Prlic[1], T. Kalro[1], R. Bhattacharya[2], C. Christie[1], P.W. Rose[1], S.K. Burley[3,4].* 1) RCSB Protein Data Bank, San Diego Supercomputer Center and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego,, La Jolla, CA 92093, USA; 2) Bioinformatics and Medical Informatics, San Diego State University, San Diego, CA 92182, USA; 3) RCSB Protein Data Bank, Department of Chemistry and Chemical Biology, Center for Integrative Proteomics Research and Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA; 4) Rutgers Cancer Institute of New Jersey, New Brunswick, NJ 08903, USA.

   Statement of purpose: Knowledge of the 3D structure of a protein at the atomic-level, provides deep insights into the molecular mechanisms underpinning normal cellular or biochemical function and potential roles in disease. To enable analyses of changes in protein sequence and structure resulting from genetic variation, we have developed novel tools that facilitate mapping of any genetic location onto corresponding protein sequence isoforms and 3D protein structures available from the Protein Data Bank archive. Our tools permit annotation of large sets of genetic variation data by correlating genomic location and sequence variation data with more than 60 external resources and protein 3D structural data housed within the data-warehouse that supports the RCSB PDB website (rcsb.org). Novel visualisation tools provide graphical depictions of the mapping of genetic variation to protein sequence and 3D structural features.**Methods used:**We developed a mapping for every nucleotide in a known human gene (as available from the HGNC; genenames.org) at a specific position in the reference genome to the mRNA and the relevant codon together with the corresponding residue of the protein isoform, plus 3D atomic coordinates. A user interface is available for manual analysis (http://www.rcsb.org/pdb/chromosome.do). The data can also be used for batch-analysis of a large number of genetic locations (e.g., for SNV analysis), using Apache Spark via 'DataFrames'.**Summary of Results:**Annotations for ~19,000 human genes are freely available from public sources. Using our analysis tools, any genomic location can be mapped to UniProt and the PDB archive of 3D structures (if available), together with a collection of additional protein-level annotations. At present, ~5200 distinct human genes can be mapped onto ~8100 experimentally-determined structures housed within the PDB archive. Extending this approach to computed homology models of human gene products will allow us to map up to 50% of all human genes on 3D structural models, when an experimental structure of the human protein is not yet available.

2830W

Landscape of somatic mutations in the normal-appearing airway cancerization field of early-stage non-small cell lung cancer. *P. Scheet[1,2], S. Sivakumar[1,2], C.L. McDowell[3], Y.A. Jakubek[1], W. Lang[3], L. Xu[3], C. Chow[4], J. Wang[5], J. Fujimoto[3], J. Fowler[1], G.E. Davies[6], I.I. Wistuba[3], E. Ehli[6], H. Kadara[3,7].* 1) Department of Epidemiology, MD Anderson Cancer Center, Houston, TX; 2) The University of Texas Graduate School of Biomedical Sciences, Houston, TX; 3) Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Department of Thoracic/Head & Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX; 5) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 6) Avera Institute for Human Genetics, Sioux Falls, SD; 7) The American University of Beirut, Beirut, Lebanon.

   Lung cancer is the second most prevalent form of cancer among men and women in the United States. It is also the leading cause of cancer mortality resulting in 1 of every 4 cancer deaths. Non-small cell lung cancer, a predominant form of lung cancer, is primarily composed of adenocarcinomas (LUAD) and squamous cell carcinomas (LUSC). "Field cancerization" theory predicts that normal appearing tissues surrounding a tumor will exhibit a "field effect" -- typically, though not absolutely, induced by smoking -- with the closer tissues showing more alterations than those farther away from the tumor. This is thought to extend to, and include, overlap with alterations in the tumor, as well. However, due to the histological complexity of LUADs, previous studies provide evidence for both linear and nonlinear progression of these tumors. Yet, the precise mechanism remains unknown. To investigate this phenomenon, we performed a genome-wide assessment of mutation profiles from targeted sequencing of 48 cases with 500 multi-region samples of the field, specifically comprising the lung tumor (LUAD and LUSC), small and large airways, nasal epithelium, as well as adjacent normal lung and blood. For 28 cases, core needle biopsy samples (CNBs) from different regions of the tumor were also profiled. We aim to compare the field cancerization of the two subtypes, assess the impact of smoking, and test for intratumor heterogeneity among the CNBs. After rigorous filtering, we obtained 267 samples with variant calls, mostly in CNBs and tumor tissues, as expected. The tumor samples showed a higher proportion of C:G>A:T variants, consistent with smoking status. We also observed concordance of mutation calls across multiple within-patient tumor/CNB samples as well as airway samples. Our initial findings identified six cases with variants in small airways comprising lung cancer drivers such as KRAS, STK11, TP53 and KEAP1; four cases showed overlapping genes with their matched tumor samples as well. We also identified variants in the normal lung (RB1, RET, IDH2), nasal epithelium (AKT1) and large airway (TP53, TSHR, SETD2, CDKN2A). These novel findings in normal-appearing tissues of the respiratory epithelium offer insights into the earliest mutational events in their progression to tumors or in local/ regional metastasis, especially in LUADs. Our results offer insights into early progression events and their potential utility as biomarkers or targets for chemoprevention of NSCLCs.

2831T

GFusion: A novel algorithm to identify fusion genes from cancer RNA-seq data. *X. Song, Q. Chen.* Nanjing University of Aseronautics and Astronautics, Nanjing, Jiangsu, China.

   Background: Cancer related fusion genes have attracted the attention of many scientists. With the development of next generation sequencing, computationally identifying fusion genes from RNA-seq data is a primary step in studying fusion genes in cancer. However, it remains a challenge to identify fusion genes with higher sensitivity and specificity. The current tools only used unmapped reads or discordant pairs, and neglected the mate reads information of the unmapped reads spanning fusion boundaries, leading to lots of spurious fusion genes are discovered and defined as fusions. Here, we present a novel fusion gene detection tool, GFusion, a powerful and efficient fusions detection method for both paired-end and single-end RNA-seq data, to predict fusion genes by comprehensive analysis of realigned split reads or spanning fragments. **Methods**: In this work the GFusion pipeline performs multiple alignments and strict filtering algorithm to improve sensitivity and reduce the false positive rate. The main steps of GFusion procedure involves with following seven steps: Initial alignment, Creation of anchors, Aligning and locating anchors, Locating fusion boundaries, Confirming fusion models, Aggregating candidate fusions, Constructing fusion index and realigning. One of powerful features in GFusion is that it focuses more on the mapped mate reads and split reads. GFusion requires the anchor length more than 20bp to minimize impact of short sequence multi-mapping. **Results**: GFusion successfully detected 26 from 30 previously reported fusions in four cancer datasets (three breast cancer cell lines BT474, SKBR3, MCF7 and a normal breast cell line). We also demonstrated the effectiveness of GFusion using 24 million 76bp paired-end reads simulation data which contains 42 artificial fusion genes, among which GFusion successfully discovered 37 fusion genes. Compared with existing methods (FusionSeq, FusionMap, Tophat-fusion), GFusion presented higher sensitivity and lower false positive rate. The GFusion pipeline can be accessed freely for non-commercial purposes at: http://bioinfo.nuaa.edu.cn/GFusion.html.

2832F

Germline mutations in cancer-predisposing genes: Clinically actionable genetic variants in 300 patients. *A.P. Tafti[1], K.Y. He[2], M.M. He[1,3,4].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; 3) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; 4) Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA.

Statement of purpose: Detection of clinically actionable genetic variants across cancer-predisposing genes using whole-genome sequencing (WGS) data and electronic health records (EHRs) may expand our understanding of carcinogenesis, direct personalized patient care, and enable genetic counseling for patients and families. Our objective is to identify clinically actionable germline mutations in cancer-predisposition genes using patients' WGS data and comprehensive EHRs. Methods used: We analyzed genetic mutations in 565 cancer-predisposition genes in 300 deceased patients with complete lifetime medical records. The 565 genes include 60 autosomal dominant cancer predisposition genes, 29 autosomal recessive genes, 58 tumor suppressor genes, 23 kinase genes, and additional 395 cancer-associated genes [1]. To infer biological insights from massive volume of WGS data and comprehensive clinical data in a short amount of time, we developed an in-house analysis pipeline within a software framework called SeqHBase [2] to quickly classify genetic pathogenicity based on variant biological function, frequency, ClinVar, and the latest ACMG guidelines [3]. Clinical data of the carriers of pathogenic and/or likely pathogenic variants were reviewed to assess their clinical conditions using their lifetime EHRs. Summary of results: We identified 33 pathogenic and/or likely pathogenic variants including 12 novel loss-of-function variants. Among the 300 participants, 44 (14.7%) carried one or more presumed pathogenic or likely pathogenic variants in at least one of the 565 cancer-predisposing genes. More importantly, 38 patients (12.7% of the population) had a genetic profile likely to cause a cancer predisposition syndrome. When assessing extensive clinical data, a majority (38/44; over 86%) of the 44 patients had different types of cancers. This work shows an estimate of clinically actionable genetic mutations in cancer predisposing genes, which could be potentially used for clinical diagnosis, prevention, and/or personalized treatments, expected from WGS. Reference: 1. Zhang et al. The New England journal of medicine 2015, 373(24):2336-2346. 2. He et al. Journal of medical genetics 2015, 52(4):282-288. 3. Richards et al: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine 2015, 17(5):405-424.

2833W

Identifying statistically significant SNP combinations from GWAS dataset with confounding variable. *A. Terada[1,2,3], D. du Verle[2,3], K. Tsuda[2,3,4], J. Sese[3,5].* 1) PRESTO, JST, Kawaguchi, Saitama, Japan; 2) Dept of Comp. Biol. and Med. Sci., The Univ. of Tokyo, Kashiwa, Chiba, Japan; 3) BRD, AIST, Koto-ku, Tokyo, Japan; 4) NIMS, Tsukuba, Ibaraki, Japan; 5) AIRC, AIST, Koto-ku, Tokyo, Japan.

Epistatic interactions are associated with complex phenotypes in humans, and detecting the combinatorial effect of SNPs is key to understand causal factors of disease. While various statistical analysis methods have been proposed to detect combination of SNPs significantly associated with a target trait, few methods can simultaneously address the following two issues: (1) difficulty of investigating high-order combination of SNPs (2) risk of spurious results due to confounding variables (e.g. age and gender). Recently, a multiple testing procedure, named Limitless Arity Multiple-testing procedure (LAMP) [Terada et al, 2013], was proposed to address this first issue of combinatorial effect discovery. However, because LAMP cannot consider the effect of confounding variables (the second issue), it may produce spurious results in GWAS. In our research, we propose a novel multiple testing procedure so as to evaluate statistical significance of combinations under confounding variables. Using a new testability bound based on the exact logistic regression model for stratified dataset, our algorithm can efficiently compute the significance with no combination size limit. We applied our method to a postmenopausal breast cancer GWAS dataset [Hunter et al., 2007] containing 2434 individuals. We analyzed 4432 exon SNPs, using age as a confounding variable. Our method confirmed a combination of 6 SNPs statistically significantly associated with breast cancer occurrence, independently of the effect of age.

2834T

Inferring escape and reversion rates of cancer neo-antigens in response to T-cell mediated immune presence. *M. Tutert[1], L. Shlush[5], A. Mitchell[5], J. Dick[1,4,5], P. Awadalla[1,2,3,4].* 1) Informatics and Bio-computing, Ontario Institute for Cancer Research, Toronto, Canada; 2) Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada; 3) St. Justine Research Hospital, Montreal, Quebec, Canada; 4) Department of Molecular Genetics, University of Toronto, Ontario, Canada; 5) Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada.

Neo-antigens are peptides that arise from somatic mutations in the human genome, which induce an anti-tumour immune response and have long been targets of cancer therapies. An immune response is dependent on successful recognition of the presented epitope to T-cells. Hence, mutations in the sequences surrounding epitopes can inhibit immune presentation or recognition and thereby influence the response of the antigen-processing pathway, enabling tumours to escape the immune system response. Inversely, epitopes are also able to revert and mutate into forms that are instead recognized by the immune response. Current methods for identifying immunologically active epitopes are based on DNA sequencing of tumours, but these methods often fail to capture the selective pressure by the immune system on sequences surrounding potential epitopes, as well as account for the variable expression due to the highly polymorphic human leukocyte antigen (HLA) genes. For this study, we use Pancan cancer analysis of whole genome (PCAWG) and exploit Acute Myeloid Leukemia (AML) genomic data including whole genome (WGS), exome and transcriptome sequencing from plasma blood taken at patient diagnosis and relapse, and whole genome sequencing of T-cells. Using existing bioinformatics pipelines, HLA typing based on WGS for these patients was performed and cross-referenced against matched RNA-sequencing data for HLA allele expression. Additionally, a subsequent mutation detection pipeline used these inferred alleles to examine detection of mutations in HLA genes. A hidden Markov model (HMM) approach was then taken to account for immune-mediated evolution of epitopes. Since selective pressure for immune escape causes a reduction in the frequency of a wild-type amino acid, we find that the expression of an HLA allele in the host provides a much better model for epitope prediction than looking at sequence data alone. This can be combined with a phylogenetic HMM which is able to make probabilistic epitope predictions based on sequence evolution. Additionally, ordinary differential equation models (ODE) were used in combination with these phylogenetic models to integrate available sequence data and estimate escape and reversion rates for list of epitopes of each patient. This model applies dynamic programming, phylogenetic, statistical and ODE-based modelling to examine epitope parameters in an effort to understand the dependence on epitopes selective pressure by the immune system.

2835F

Identification of germline copy number variations in hereditary prostate cancer cases using whole-exome sequencing analysis. *K. Wood[1], M. Baddoo[2], D. Mandal[1].* 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, Tulane University, New Orleans, LA.

Family history is one of the most prominent predictors of prostate cancer (PCa) risk, and approximately 10% of PCa cases are attributable to inheritable genetic factors. Next-generation sequencing (NGS), including whole-exome sequencing (WES), provides reliable large scale data that can be used for variant identification in PCa. Genomic copy number variations (CNVs) have been detected in prostate tumors, but changes in copy number have not been studied in germline DNAs from hereditary PCa (HPC) cases. The goal of the present study is to use a CNV-specific algorithm to identify CNVs in WES data on Louisiana HPC families. We hypothesize that CNVs in actionable genes contribute to high-risk PCa cases. We have WES data on 3 affected cases from 3 different high-risk HPC families (≥5 PCa affected/family) from Louisiana. Two of the cases are African Americans and one is Caucasian, and the age at diagnosis ranges from 51 to 60 years of age. A CNV-specific calling algorithm implemented in CANOES was used on WES data in order to successfully call the CNVs. CANOES works by reading the WES data files into R programming language, calling CNVs in the samples by modeling read counts using a precise distribution, and plotting all the CNV calls. Following the detection of CNVs, the variants were visualized using Integrative Genomics Viewer (IGV), which is used to visualize genomic variants, including changes in copy number. CANOES called a total of 446 germline CNVs from the WES data, and 278 of the CNVs were confirmed using IGV. Of the 278 CNVs, 84 were in exonic regions of genes, including the exonic regions of 7 cancer-related genes. Four of these 7 cancer-related genes have previously been associated with prostate tumors, including *SPAG11A*, *DMBT1*, *POTEM* and *CKMT1B*. To our knowledge, this is the first time these 4 genes have been found to be altered in the germlines of PCa-affected individuals. In future studies, we will apply the same bioinformatics pipeline to germline WES data for additional PCa cases and unaffected relatives from the same families and from other HPC families. The identification of germline CNVs within HPC families will contribute to the impending need of cancer biomarker identification in the era of precision medicine, and will allow us to facilitate PCa screening in HPC families.

2836W

TASRcloud: **Targeted** *de novo* **assembly of candidate genes on the cloud.** *Z. Xue, R.L. Warren, E. Gibb, D. MacMillan, R. Chiu, J. Chu, I. Birol.* Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada.

The steep decline in DNA sequencing costs over the last decade has enabled the genome and transcriptome sequencing of large patient cohorts, such as those from The Cancer Genome Atlas (TCGA). Traditionally, laboratories were downloading and using these data on an individual basis, which is often a substantial undertaking and an inefficient practice. We present *TASRcloud*, a novel targeted de novo sequence assembly methodology on the NCI Cancer Genomics Cloud Pilots. Application of targeted *de novo* assembly approaches to large sample cohorts is often overlooked. When a researcher is interested in a specific gene or sequence, a complete assembly becomes unnecessarily complicated and confounding. The retrospective targeted reconstruction of transcripts in large cohorts leverages pre-existing big data as a near-inexhaustible sequence source for focused discovery. The targeted sequence assembly approach uses established bioinformatics software that are already in use at the Institute for Systems Biology (ISB) Cloud Pilots in Seattle, including BioBloom Tools (BBT), and award-winning *de novo* genome assembler ABySS. Over 50% of human genes are estimated to produce two or more transcript isoforms via alternative polyadenylation (APA) [1], indicating that shifts in 3' untranslated region (UTR) lengths may represent a general mechanism for regulating gene expression by post-transcriptional processing. In cancer, a global shortening of 3' UTRs has been observed [2], suggesting a mechanism by which certain genes in cancer pathways may escape the regulatory effects of other molecules. We report on a specific application of *TASRcloud*, by conducting a comprehensive pan-cancer survey of tumor-associated APA events by targeted reconstruction of candidate 3'-UTRs, and demonstrate the utility of cloud computing resources for targeted cancer genomics studies by leveraging TCGA data via the NCI Cancer Genomics Cloud Pilots program. [1] Tian B, et al. 2005. NAR. 33:201-12 [2] Lin Y, et al. 2012. NAR. 40:8460-71.

2837T

Visualization and analysis of cancer genomics data with omics tools. *C. Yan, Y. Hu, Q. Chen, D. Meerzaman.* NCI, Rockville, MD.

Next generation sequencing technologies have generated large amounts of multidimensional genomics data to obtain the complete landscape of tumor alterations. Visualization and network analysis tools are commonly used to uncover the complex interrelationships among mutation, expression, copy number and methylation profiles. We have developed two R packages OmicCircos and OmicPath to generate high-quality circular plots and network graphics in Bioconductor environment. OmicCircos has been downloaded for more than 10,000 times in the past three years. With implementation of users' suggested functions, we updated OmicCircos with the advanced zoom and text manipulation capabilities. OmicPath is a novel R package to perform pathway analysis, construct de novo network and run systematic functional annotation. It performs multiple statistical analyses on several pathway databases such as PID, KEGG, and user-defined networks. OmicPath provides an easy way to examine the network topology characteristics of expression and mutation profiles in cancers. The utilization of these packages will contribute to a better understanding of cancer molecular etiology.

2838F

mTCTdb: An integrative resource for mutation-dependent targeted cancer therapy. *H. Yao[1], M. Li[1], Z. Liu[1], H. Xu[1], P. Sham[1], J. Wang[2].* 1) Center for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong , Hong Kong, China; 2) Department of Health Sciences Research and Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ 85259, USA.

Recent advances on next-generation sequencing technologies have enabled efficient and accurate detection of somatic alterations from tumor tissue or liquid biopsy of a cancer patient. These personalized tumor mutation profiles provide valuable information for targeted cancer therapy, which identifies the optimal drug(s) or treatment(s) to patient. However, mutation-dependent drug response information, or rather mutation therapeutic implications more generally, have not been well collected and structured. Here we developed an integrative database to provide researchers and clinicians with a combination of mutation information, mutation therapeutic implications and corresponding drug information. Mutation information were obtained from various mutation annotations and functional prediction algorithms. Drug information were collected from databases including DrugBank, DGIdb, TTD, ClinicalTrials.gov and then corrected manually through FDA and literature. Mutation therapeutic implications were acquired both from database such as CIViC and our manual curation. Mutation-oriented drug recommendations are offered in two levels, including direct evidence and gene-drug relationship based prioritization. Given a specific mutation, our resource could provide comprehensive mutation annotation and corresponding therapeutic information, which will benefit both clinicians and researchers. The link of this database will be provided by the end of June, 2016.

2839W

Predicting genome-wide DNA methylation of repetitive elements. *Y. Zheng[1,2], L. Liu[1,3], W. Zhang[1,3], B. Joyce[1,4], Z. Zhang[1,5], W. Kibbe[6], L. Hou[1,3].* 1) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Health Sciences Integrated PhD Program, Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Robert H. Lurie Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) 3Robert H. Lurie Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA Division of Epidemiology/Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL; 5) Driskill Graduate Program in Life Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL; 6) Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD.

DNA repetitive elements (REs) account for over 55% of the human genome, half of which can cause genomic instability by proliferating and inserting themselves into new genomic locations. DNA methylation is an epigenetic defense mechanism that suppresses RE mobility and thus maintains genomic stability. Accumulating evidence shows that methylation greatly varies across specific REs, suggesting distinct locus-specific roles. However, whole-genome RE methylation sequencing is challenging and remains prohibitively expensive for large population studies. The widely used surrogate approach of measuring global methylation by simply averaging methylation levels at various RE loci throughout the human genome may lead to massive information loss and substantially inhibit our ability to further elucidate the distinct biological roles of DNA methylation in locus-specific RE. We developed a random forest-based algorithm to computationally predict genome-wide, locus-specific RE methylation levels. With this algorithm we predicted methylation levels of two major REs, Alu or LINE-1, and verified them using cross-platform validation. By applying this algorithm to TCGA data, we identified that both hypo- and hyper-methylated Alu and LINE-1 could occur in tumor tissue, potentially involving different biological pathways. We also observed that using the mean value of Alu or LINE-1 methylation was less effective in discriminating tumor from normal tissue. Our work may lead to an improved understanding of the role played by DNA methylation of REs in human diseases.

2840T

An FFPE-based prognostic signature to predict metastasis in stage I/II microsatellite stable colorectal cancer. *C.H. Blöcker[1], Y.S. Low[2], J.R. McPherson[1], S.A. Tang[2], Y.Y. Cheng[1], J.Y.S. Wong[2], C. Chua[3], T.K.H. Lim[4], C.L. Tang[2], M.H. Chew[2], P. Tan[1,5], I.B. Tan[3], S.G. Rozen[1,5], P.Y. Cheah[1,2,6].* 1) Programme in Cancer & Stem Cell Biology, Duke-NUS Medical School, National University of Singapore, Singapore; 2) Department of Colorectal Surgery, Singapore General Hospital, Singapore; 3) National Cancer Center, Singapore; 4) Department of Pathology, Singapore General Hospital, Singapore; 5) The Precision Medicine Institute of SingHealth and Duke-NUS, Singapore; 6) Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

PurposeApproximately 25% of early stage (I/II) patients with colorectal cancer (CRC) develop metastases despite curative surgery. Here, we aim to develop a formalin-fixed and paraffin embedded (FFPE)-based predictor of metastases in early stage, microsatellite-stable CRC patients.**Patients and Methods**In designing the predictor, we first considered genome-wide mRNA and miRNA expression and the mutation status of 20 genes as assayed in 150 fresh-frozen tumors. Based on microarray analysis of mRNA levels, we selected 193 potentially informative genes for analysis using NanoString nCounter arrays on the FFPE samples corresponding to the frozen tumors. Neither mutation status nor miRNA expression improved estimated prediction compared to prediction based on the mRNA levels of 193 genes alone. Furthermore, addition of genes beyond the top 19 did not improve prediction. The final predictor, ColoMet19, based on these 19 genes' mRNA levels assayed by nCounter in FFPE samples, had an estimated positive predictive value of 0.72 in the tumors on which it was developed.**Results**We tested ColoMet19 on an independent test set of 131 tumors with known metastasis status, and obtained a population adjusted positive predictive value of 0.73 and a negative predictive value of 0.82. Predicted-positive patients had poorer metastasis-free survival (hazard ratios 1.92 and 2.05 in the design and testing data, respectively).**Conclusion**The positive predictive value evaluated in the test set indicates that early stage CRC patients who are positive by ColoMet19 have a 73% risk of developing metastases. This risk is much higher than that of node-positive CRC patients, who are generally treated with adjuvant chemotherapy. Thus, early stage CRC patients who test positive by ColoMet19 may be considered for adjuvant therapy after surgery.

2841F

Integrating genome and transcriptome data to predict functional driver mutations in breast cancer. *K. Chen, Z. Wang, K. Ng, T. Chen, T. Kim, K.R. Shaw, F. Meric-Bernstam, G.B. Mills.* The University of Texas MD Anderson cancer center, Houston, TX.

Accurate prediction of the functional effects of genetic variation in cancer is critical for realizing the promise of precision medicine. Due to genetic heterogeneity across patients, predicting driver mutations using population genomic approaches remains challenging. We developed a novel Bayesian method, termed xDriver that integrates functional properties of mutations (such as evolutionary conservation) with quantitative traits (such as mRNA expressions) to achieve accurate, context-sensitive prediction of driver mutations. We demonstrate using 752 breast cancer samples in The Cancer Genome Atlas that our integrative approach is able to significantly improve the accuracy of driver mutation identification over existing approaches that do not perform such integration. In particular, our approach is able to enhance the functional prioritization of so-called "tail" (rare) mutations and more accurately delineate driver mutations specific to cancer subtypes (such as *PIK3CA* mutants associated with lymph node positive or negative patients). Importantly, scores generated by our model achieve better agreement with *in vitro* functional cell viability data obtained from growth factors-dependent Ba/F3 and MCF10A cell models, compared to predictions from other commonly used algorithms. Our study highlights the importance of integrating quantitative traits in predicting cancer driver mutations and provides a statistically rigorous solution to accelerate cancer target discovery and development.

2842W

Translating cancer genomics research bioinformatics to precision oncology: Accurate identification of somatic mutations in cancer patient specimens by NGS in clinical scenarios. *F.M. De La Vega[1,2], S. Irvine[3], D. Ware[3], K. Gaastra[3], Y. Pouliot[1], Y. Bouhlal[1], A. Vilborg[1], D. Mendoza[1], F. Goodsaid[1], A. So[1], L. Trigg[3].* 1) TOMA Biosciences, Foster City, CA, USA; 2) Stanford University School of Medicine, Stanford, CA, USA; 3) Real Time Genomics, Hamilton, New Zealand.

Tumor molecular profiling is rapidly becoming the standard clinical test for selecting targeted therapies in refractory cancer patients. DNA extracted from patient samples is enriched for cancer genes and sequenced to identify actionable somatic mutations therein. A major challenge arises when tumor-derived data is analyzed in the absence of normal tissue data, as it is common in clinical scenarios. The distinction between somatic and germline variants become difficult, leaving clinicians to resort to crude heuristic filtering. We present here a variant calling software, developed under quality system regulation protocols, capable of accurately identifying somatic mutations from targeted next-generation sequencing data. A novel Bayesian Network approach models the distribution of reads harboring germline and somatic mutations, estimates the contamination from normal tissue in the sample, scores somatic mutations, and imputes germline variants, without matching normal tissue data. This approach also allows joint analysis of multiple specimens from the same patient (e.g. FFPE and ctDNA) when available, improving the limit of detection. To improve specificity, our caller can also utilize prior information from different databases including somatic mutations, germline variation, and healthy controls data. We validated our method by analyzing data from the TOMA OS-Seq 131 cancer gene panel using the Illumina platform. To validate our method, we analyzed somatic mutations simulated on a background of data from reference samples, data on a dilution series from two reference samples, and several commercial control and clinical samples. We show that, compared to other commonly used methods, we can achieve higher true positive rate whilst maintaining a false discovery rate of 1%. In conclusion, we observe that our caller outperforms other software and is particularly well suited for the clinical use cases.

2843T

Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *J. Demeulemeester[1,2], P. Kumar[3,4], E.K. Møller[5,6], S. Nord[5,6], D.C. Wedge[7], A. Peterson[8,9], R.R. Mathiesen[4,10,11], R. Fjelldal[12], M.Z. Esteki[3], A.J. Grundstad[8], E. Borgen[12], L.O. Baumbusch[1,13], A.-L. Børresen-Dale[5,6], K.P. White[8], B. Naume[6,10], V.N. Kristensen[5,6,12,14], P. Van Loo[1,2], T. Voet[3,4].* 1) The Francis Crick Institute, London, UK; 2) Human Genome Laboratory, Department of Human Genetics, University of Leuven, Leuven, Belgium; 3) Laboratory of Reproductive Genomics, Department of Human Genetics, University of Leuven, Leuven, Belgium; 4) Single-cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, UK; 5) Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo, Norway; 6) K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 7) Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK; 8) Institute for Genomics & Systems Biology and Department of Human Genetics, University of Chicago, Chicago, IL, USA; 9) (Present address) Laboratory of Genetics, University of Wisconsin, Madison, WI, USA; 10) Division of Surgery and Cancer Medicine, Department of Oncology, Oslo University Hospital, Radiumhospitalet, Oslo, Norway; 11) (Present address) Department of Oncology, Akershus University Hospital, Lørenskog, Norway; 12) Department of Pathology, Oslo University Hospital, Radiumhospitalet, Oslo, Norway; 13) (Present address) Department of Pediatric Research, Women and Children's Division, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 14) Department of Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital, Lørenskog, Norway.

Background Single-cell micro-metastases of solid tumors often occur in bone marrow. These disseminated tumor cells (DTCs) may resist therapy and lay dormant or progress to cause overt bone and visceral metastases. Unfortunately, the molecular nature of DTCs remains elusive, as well as when and from where in the tumor they originate. Here, we apply single-cell sequencing to identify and trace the origin of DTCs in breast cancer. **Results** We sequenced the genomes of 63 single cells isolated by micromanipulation from the bone marrow of six non-metastatic breast cancer patients using established immunocytochemical markers and morphologic characteristics for epithelial tumor cells. Comparison of the cells' DNA copy number aberration (CNA) landscapes with those of the primary tumors and lymph node metastasis established that almost half of the cells classified as tumor cells are indeed DTCs disseminating from the observed tumor. The remaining cells represented non-aberrant 'normal' cells and 'aberrant cells of unknown origin' that have CNA landscapes discordant from the tumor. Genotyping somatic mutations called on bulk tumor exomes in the single-cell sequences confirmed that these cells did not derive from the same lineages as the observed breast cancers. Intriguingly, their prevalence tends to increase with patient age. Evolutionary reconstruction analysis of bulk tumor and DTC genomes enabled ordering of CNA events in molecular pseudo-time and tracing the origin of the DTCs to either the main tumor clone, primary tumor subclones, or subclones in an axillary lymph node metastasis. **Conclusions** Single-cell sequencing of bone-marrow epithelial-like cells, in parallel with intra-tumor genetic heterogeneity profiling from bulk DNA, is a powerful approach to identify and study DTCs, yielding insight into metastatic processes. A heterogeneous and age-correlated population of CNA-positive cells of unknown origin is present in bone marrow.

2844F

Interactome-based approach to identification of cancer genes with low mutation frequency. *I. Glotova[1], D. Kural[2].* 1) Seven Bridges Genomics UK Ltd., London, United Kingdom; 2) Seven Bridges Genomics Inc., Cambridge, United States.

Large-scale cancer projects are generating constantly growing amounts of genomic data in many cancer types. One key challenge in the detection of functional cancer driver genes is the identification of genes with low mutation frequency. These genes are usually near significance when scored by driver gene callers, and could be identified as new drivers when more individuals are sequenced for the corresponding tumor. We applied a systems biology approach that integrates mutation data from 32 tumors with a human protein-protein and protein-DNA network (interactome). Using mutation data as input, four driver gene callers were employed for obtaining a list of predicted significant and near significance genes (NSGs) for each tumor: MutSigCV, DOTS-Finder, OncodriveFM and OncodriveCLUST. Simple mapping of known genes (Cancer Gene Census), significant genes and NSGs to the interactome shows that some NSGs do interact with at least one known/significant gene in the corresponding tumor network, and/or are involved in the same pathway(s) with certain known/significant genes. To further analyze the network properties of NSGs, we applied network-based localization measures: module size (largest connected component in a tumor network) and shortest mean distance (mean of all the shortest paths for all gene pairs in a tumor network). Although NSGs increase module size in most tumors that have an observable module, module size was found to be a degree-dependent property in 47% of these cancers (becomes non-significant if node degree is considered). However, mean shortest distance is not degree-dependent and is significantly lower (compared to random networks) in 26 tumors (even when NSGs are added to the tumor network). This observation shows that NSGs have a tendency to agglomerate in the same interactome neighborhood as known and predicted driver genes. Moreover, in 61.5% of tumors the addition of NSGs to the tumor network decreases its mean shortest distance. Thus, if combined with additional data sources like sequence, expression and genotypic data sets, mean shortest distance can be used as a predictive measure for low-frequency driver genes in different cancers. We also analyzed network-based separation between two tumor networks (shows if the gene neighborhoods of two tumors significantly overlap or are separated). NSGs were found to increase the overlap between some tumor neighborhoods, as well as decrease the overlap between others and make it non-significant.

2845W

Comprehensive molecular profiling and subtyping of lung adenocarcinoma in Asian population. *S. Hwang[1], J. Byn[2], C. Seo[3], J. Kim[1], Y. Jun[1], Y. Jung[1], C. Park[1], S. Kim[1], P. Kim[1], J. Seo[1], Y. Jang[1], N. Yu[1], Y. Jang[1], M. Lee[1], Y. Jung[1], S. Park[4], J.O. Yang[4], B. Lee[4], J. Lee[3], J. Kim[1], D. Hwang[2], J. Kim[5], S. Lee[1].* 1) Life Science, Ewha Womans University, Seoul, South Korea; 2) New Biology, Daegu-Gyungbuk Institute of Science and Technology, Deagu, South Korea; 3) DNA Link Inc., Business Cluster Center, 150 Bugahyeon-ro, Seodaemun-gu, Seoul, South Korea; 4) Korean Bioinformation Center (KOBIC), Korea Institute of Bioscience and Biotechnology, Deajon, South Korea; 5) Lung Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea.

Lung cancer is the most common cause of global cancer deaths, resulting in over a million deaths each year and adenocarcinoma is the most common subtype of lung cancer. Here we report molecular profiling of 114 lung adenocarcinoma (LUAD) patients in Korea using whole exome sequencing, transcriptome sequencing, and copy number variation analysis. Comparative analysis with ~1,000 cases from previous studies, which are mostly on Caucasian patients, showed a dramatic difference in mutation frequencies of EGFR and KRAS mainly due to the ethnic factor. EGFR mutations were more frequent in Asian patients, whereas mutations in KRAS were more common in Caucasians. In Asian LUAD patients, the major alterations were EGFR mutations or indels such as EGFR L858R and Exon 18 deletion, but many other driver alterations were also identified including fusion events of ALK, ROS, and RET genes and aberrant splicing in MET, TP53 and PTEN genes. Notably, we were able to identify actionable targets and therapies for ~80% of our LUAD patients. We have also developed a new patient stratification strategy combining the mutation and gene expression profiles. Mutation-based patient clustering suggested three subtypes – EGFR-only, EGFR and TP53, and high mutation (including KRAS) groups. The EGFR-only subtype showed significantly better prognosis than two other subtypes (p-value < 0.03). Patient clustering based on mRNA expression, using fold-change value between tumor and matched normal tissues, identified another three subtypes whose signature genes were enriched in the metabolism and PD-1 signaling, oncogenic pathways such as RTK-PI3K-AKT-mTOR, and metastatic processes. These three groups were in partial agreement with the TCGA subtypes of terminal respiratory unit, proximal proliferative, and proximal inflammatory subtypes, respectively. The metastatic subtype showed worse prognosis than two others. The subtypes from mutation-based and expression-based clustering were distinct and complementary. In conclusion, our comprehensive molecular profiling of lung adenocarcinoma among Asian population not only identified the ethnic difference in driver mutations and copy number alterations but also devised a new patient stratification scheme combining mutation and expression profiles, which might be valuable in suggesting personalized precision care for lung adenocarcinoma patients.

2846T

Don't assume infinite sites! Detection of intratumoral divergence of copy number alterations. *Y.A. Jakubek[1], H. Kadara[2,3], P. Scheet[1].* 1) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX; 3) The American University of Beirut, Beirut, Lebanon.

Increased intratumoral heterogeneity is associated with poor prognosis for cancer patients. Genetic heterogeneity in the context of tumor genomes can generally be classified into two categories: single nucleotide variants (SNVs) and copy number alterations (CNAs); in the later category we include copy neutral loss of heterozygosity. Currently most methods for tumor heterogeneity detection focus on SNVs and net copy number differences among samples from the same tissue or tumor. These methods often apply an infinite sites model for both SNVs and CNAs. While this may often be appropriate for SNVs, we demonstrate its violation for CNAs in highly unstable cancer genomes. Here, we propose a novel method to identify divergent CNAs in multiple SNP DNA microarray data sets from the same tumor. In this context divergent is defined as a CNA identified in multiple samples from the same tumor but with different chromosomal changes giving rise to the CNA. Our algorithm uses B allele frequency (BAF) data from Illumina SNP arrays to detect regions where BAFs from germline heterozygous sites shift in opposite directions between pairs of samples (indicative of divergent tumor clones). We applied our method to data from core needle biopsies extracted from the tumors of 30 non-small cell lung cancer (NSCLC) patients. We observe instances of divergent allelic imbalance changes in 16 of the 30 cases comprising 2% of CNAs tested. Strikingly, one tumor had a rate of 16% divergent CNAs. These divergent genomic alterations represent about half of the total number expected since imbalances of the same haplotype will not be observable in such data. Some of the identified divergent events encompassed entire chromosomal arms (5p, 7q, 18p, 18q, 20p, 21q, 22q). Our method allows for the detection of genomic regions that are divergently lost or gained. This information can help identify CNAs under positive or negative selection in the tumor microenvironment. This approach provides an added dimension to tumor heterogeneity analysis for a more comprehensive characterization of cancer genomes.

2847F

GtExCLDB – A database of genotypes and gene expression for cancer cell lines. *I. Jeong, J. Kim, I. Choi, Y. Jun, S. Kim, M. Lee, S. Lee.* Ewha Womans University, Daehyeon-dong, Seodaemun-gu, Seoul, South Korea.

Cancer cell lines are essential resources for biomedical researches such as functional studies and screening experiments for developing therapeutic targets. Proper choice of cell lines that are suitable for experimental purpose is often difficult because the genotype and/or expression data are missing or scattered in diverse resources. Here we report the development of GtExCLDB, an online database resource of human cancer cell lines that provides the genotype and expression information. We have collected the mutation, gene expression, and copy number variation (CNV) data from three representative databases on cell lines – CCLE (Cancer Cell Line Encyclopedia), COSMIC (Catalogue of Somatic Mutations in Cancer), and NCI60. In total, GtExCLDB includes 1,787 cell lines from 170 cancer types and 29 tissues. Mutation, gene expression, and CNV information is available for 1,369, 1,318, and 1,613 cell lines, respectively. Since the microarray platforms are different for each database, we removed the batch effect using Combat software and re-processed the whole data of gene expression and SNP chips. The cell line information was standardized using Cellosaurus from ExPASy. Our user interface supports diverse queries where cancer types, tissues, genotypes, and gene expression conditions can be combined. For example, users can easily identify lung cancer cell lines with EGFR L858R mutation and browse the gene expression or CNVs in specific genomic region for those cell lines. Providing molecular characteristics as well as the clinical information, we believe that GtExCLDB would be a valuable resource for many biomedical researches for functional or screening studies. GeExCLD is available at http://gtexcldb.ewha.ac.kr.

2848W

Impact of genetic and epigenetic alterations in prostate cancer on tissue-specific regulatory network. *E. Khurana, P. Dhingra, A. Martinez Fundichely, A. Neil Forbes, E. Minwei Liu, D. Liu, A. Sboner, D. Rickman, M. Rubin.* Weill Cornell Medicine, New York, NY.

Prostate cancer is characterized by multiple recurrent genomic and epigenomic alterations. We investigated the combined effects of single nucleotide variants (SNVs), large structural variations (SVs) and DNA methylation changes on the prostate transcriptional machinery. We integrated whole-genome sequencing (WGS), RNA-seq and DNA methylation data from 521 primary prostate tumor samples with functional genomics data from the ENCODE and Roadmap Epigenomics projects. We developed a three step computational model: (a) construction of prostate regulatory network using DNase I hypersensitivity data, (b) identification of significantly mutated, rearranged and differentially methylated coding and non-coding regions in prostate cancer and, (c) interpretation of the effects of these alterations on prostate regulatory network. We identified 153 transcription factor (TF) hubs in the network, which include known prostate cancer genes (such as ERG) and novel candidates. Our results show that although more regions show significant differential methylation than SVs and SNVs in the regulatory network, a larger number of TF hubs are affected by SVs. This shows SVs are likely to play a stronger role in prostate cancer development by causing larger perturbations in the network as compared to SNVs and methylation changes. We hypothesized this may occur via crosstalk between TF hub expression (modulated by SVs) and DNA methylation changes. Consistent with the prediction, functional validation using Enhanced Reduced Representation Bisulphite Sequencing (ERRBS) shows global methylation changes caused by ERG overexpression. This work provides a novel framework for using tissue-specific regulatory networks to understand the global impact of genetic and epigenetic alterations in tumors.

2849T

Identifying transcriptomic mechanisms of gemcitabine resistance and counter-resistance pathways for pancreatic cancer using public data. *R. Kusko[1], Y. Cha[1], J. Perez-Rogers[2], N. Caballero[1], G. Koytiger[1], J. Shankar[1], A. Lysaght[1], B. Weiner[1], S. Kolitz[1], F. Towfic[1], K. Fowler[1], B. Vardarajan[1,3], M. Artyomov[1,4], B. Zeskind[1].* 1) Genomics, Immuneering, Cambridge, MA; 2) Computational Biomedicine, Boston University School of Medicine, Boston, MA; 3) Columbia University, New York, New York; 4) Washington University in St. Louis, St Louis, MO.

Although pancreatic cancer is increasing in incidence, therapeutic options remain limited. Treatment with gemcitabine or a gemcitabine combination therapy is most common, although current combination therapies are challenged by toxicity. The urgent unmet clinical need for promising novel combination therapies to enhance the efficacy of gemcitabine is suggested by the alarmingly high 5 year mortality rate. We studied gene expression data from 3 pancreatic cancer cell line datasets (Klijn et al, Collisson et al, and CCLE) to query druggable genes and pathways which mastermind gemcitabine resistance or encourage gemcitabine sensitivity. First we processed data and removed low quality samples, then utilized a t-test to pinpoint genes differentially expressed (DE) between cell lines that are sensitive vs resistant to gemcitabine as dictated by Collisson et al's IC50 values. The overlap of DE genes between all three studies revealed 83 genes upregulated and 48 genes downregulated in gemcitabine resistant cell lines. Specific genes of interest changing include upregulation of apoptosis suppressor TMBIM6 (p<0.003) and pro-survival gene IL20RA (p<0.003) in resistant cell lines. Enrichment of genes upregulated in resistant cell lines includes GEO Kinase perturbation ERBB3 (p.adj=3.28E-08) and GO cyclin-dependent protein serine/threonine kinase regulator activity (p.adj=0.038). GSEAP was conducted in each dataset separately, and top shared significant pathway results include upregulation of KEGG drug metabolism cytochrome P450 (p.adj<0.0003) as well as Retinol metabolism (p.adj<0.0004). Drugs which downregulate these genes and pathways may enhance the effect of gemcitabine. Enrichment of genes upregulated in sensitive cell lines includes GEO Kinase Perturbation ALK (p.adj=1.16E-05) and IRAK4 (4.93E-06), as well as PPI Hub protein PRKACA. Drugs which enhance these kinases and hub proteins may increase effectiveness of gemcitabine. Ongoing work includes analysis to pinpoint drugs which may appropriately enhance or suppress above mentioned master regulators of resistance. Functional genomic studies are underway to validate key targets. By leveraging publicly available data to detailing a consistent gene set which separates gemcitabine sensitive from resistant pancreatic cancer cell lines, we have characterized numerous potential driver candidates and pathways which could be the key to a novel gemcitabine combination.

2850F

An analysis of expression of druggable targets and prostate cancer disease signatures in 2,978 patients with adverse pathology after radical prostatectomy: Initial results from the Decipher GRID. *J. Lehrer, M. Alshalalfa, N. Erho, H. Ashbab, M. Takhar, P. Wood, L. Lam, E. Davicioni.* GenomeDx, Vancouver, BC, Canada.

   INTRODUCTIONProstate cancer patient management has been enhanced with several commercial genomic prognostic tests such as the Decipher® prostate cancer classifier. Unlike other tests, Decipher generates genome-wide expression data for each patient. This data has been anonymized and made available for research in the Decipher GRID™. Here we report an initial analysis further characterizing the genomic landscape of localized prostate cancer patients most at risk for recurrence after radical prostatectomy. **METHODS**We conducted a literature search to identify genes and signatures of potential clinical relevance to prostate cancer. The expression and distribution of 697 genes and 31 prostate cancer disease signatures for metastasis risk, proliferation, luminal/basal, small cell and AR signalling were examined across 2,978 radical prostatectomy tumors with adverse pathology tested with Decipher and available in the GRID. For the 697 genes, expression distribution was characterized and high and low expression were defined using thresholds based on median +/- 1.5*1.48*MAD (median absolute deviation). Genes from 31 published prostate cancer signatures were adapted to the Decipher platform and scores were calculated. **RESULTS**For various druggable targets including immune checkpoint inhibitors (PD1, PDL1) and growth receptors (c-MET, EGFR, HER), between 2-11% of patients had high expression above right threshold. Clustering of scores for 31 signatures revealed several clear groupings of patients. About 20% of patients consistently had high scores for all the metastasis risk signatures and low-average AR signalling scores. About 10% of patients had high scores for the proliferation signatures but low metastasis risk scores. Tumors with low AR signalling scores were enriched with high basal and small cell signature scores whereas most luminal tumors had higher AR signalling. **CONCLUSIONS**Since every patient who has received the Decipher test also has a genome-wide expression profile, the Decipher GRID allows researchers to evaluate on a systematic, population-level the expression of genes and signatures that may guide therapies. Such information may be useful for selection of optimal systemic therapy and inclusion into clinical trials of novel targeted agents. This rich genomic resource is being made available on a research use only basis to prostate cancer researchers and to clinicians seeking to better understand prostate cancer in order to advance precision medicine.

2851W

RNA sequencing-based cell proliferation analysis across 19 cancers identifies a subset of proliferation-informative cancers with a common survival signature. *R. Ramaker[1,2], B. Lasseigene[1], A. Hardigan[1,2], S. Cooper[1], R. Myers[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Alabama at Birmingham, Birmingham, AL.

   Despite advances in cancer diagnosis and treatment strategies, robust prognostic signatures remain elusive in most cancers. Cell proliferation has long been recognized as a prognostic marker in cancer, but it has not been thoroughly investigated across multiple cancers using tissue-based RNA sequencing. Here we explore the role of cell proliferation across 19 cancers (n=6,581 patients) from The Cancer Genome Atlas project by employing a 'proliferative index' derived from gene expression associated with *PCNA* expression. This proliferative index is significantly associated with patient survival (Cox, FDR<0.05) in 7/19 cancers, which we have defined as 'proliferation-informative cancers' (PICs). In PICs the proliferative index is strongly correlated with tumor stage and nodal invasion. PICs paradoxically demonstrate reduced baseline expression of proliferation machinery relative to non-PICs suggesting that non-PICs may saturate their proliferative capacity early in tumor development and allow other factors to dictate prognostic outcomes. We find that tumor proliferative index is significantly associated with gross somatic mutation burden (Spearman, p=1.76×10^{-23}) as well as protein altering mutations in well-characterized driver genes such as *TP53*, *PIK3K*, and *RB1*. In breast cancer, protein altering mutations in *RELN* were associated with increased proliferative index (FDR<0.1) across all PAM50 subtypes and decreased expression of *RELN* was associated with poor prognosis (p=0.08) in patients with basal breast cancer. This analysis provides a comprehensive characterization of tumor proliferation rates and their association with disease progression and prognosis across cancer types and highlights specific cancers and mutation events that may be particularly susceptible to improved targeting of this classic cancer hallmark. .

2852T

Patterns of somatic genomic variation in premalignant lesions inform insights into early pathogenesis of lung adenocarcinomas. *S. Sivakumar[1,2], J. Fowler[1], Y. Jakubek[1], C.L. McDowell[3], W. Lang[3], L. Xu[3], J. Fujimoto[3], E. Ehli[4], Y. Yatabe[5], H. Kadara[3,6], P. Scheet[1,2].* 1) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) The University of Texas Graduate School of Biomedical Sciences, Houston, TX; 3) Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Avera Institute for Human Genetics, Sioux Falls, SD; 5) Department of Pathology and Molecular Diagnostics, Aichi Cancer Center, Nagoya, Japan; 6) The American University of Beirut, Beirut, Lebanon.

Lung cancer is the second most prevalent form of cancer and the leading cause of cancer-related deaths worldwide. Its high mortality is due to late diagnosis and further compounded by a dearth of knowledge about its molecular evolution and pathogenesis, especially for the major histological type, adenocarcinoma (LUAD). Here we profile mutations and gene expression in a cohort of 25 patients with matched tumor, premalignant and normal tissues, to identify signatures indicative of the neoplastic progression of premalignant tissues to invasive LUAD. The premalignant tissues (atypical adenomatous hyperplasia; AAH), are considered precursor lesions to LUAD. We carry out a genome-wide assessment of single nucleotide variants and copy number alteration profiles through a cross-platform analysis comprising Ion Torrent deep targeted sequencing of cancer associated genes and broader-scale whole-genome SNP genotyping arrays in order to identify events in AAHs that overlap or differ from LUADs. We will also assess associations with smoking. As Ion Torrent is challenged to resolve homopolymers, we developed an algorithm that considerably reduced artifacts and is now part of our automated QC variant pipeline for this technology. From our preliminary mutation analysis of 10 patients (6 ever-, 4 never-smokers), we observed 16 genes, including known lung cancer drivers such as EGFR, TRIM24 and ABL2, from a within-patient between-tissue analysis; with up to 9 identical variants shared per patient. Moreover, at the tissue level, we found 22 genes common to AAHs and LUADs, including oncogenes such as EGFR, ABL2, RARA, JAK3 and tumor suppressor genes such as STK11 and SOCS1, offering evidence of AAH progressing to LUAD. Among non-synonymous variants unique to the AAH, we observed mutations in KRAS for two patients, both smokers, which may suggest early stages of a second primary tumor. Other variants unique to AAH include genes such as BRAF, SOX11, CSMD3, CDKN2A, NF1, and MAP2K2, all previously implicated in LUAD. Further, C:G>T:A transitions were the most common among all base substitutions, which is typical of EGFR mutants and non-smoker lesions. Our encouraging results offer a better understanding of the neoplastic progression mechanism in LUAD, thus aiding novel biomarker discovery for their early detection and possible chemoprevention.

2853F

Bioinformatics design and analysis approach for targeted NGS tumor immune response expression profiling. *Y. Sun, A. Atkins, A. Pankov, T. Looney, Y. Ku, W. Tom, F. Hyland, J. Au-Young, A. Mongan.* Thermo Fisher Scientific, South San Francisco, CA.

Targeted panel for transcriptional profiling offers increased sensitivity and higher degrees of sample multiplexing for selected gene targets of high interest. In particular, interleukins and other cytokines regulating the immune response are typically expressed at fairly low levels and would be undetectable without the use of targeted panels. Here, we present a 400-gene panel to evaluate the expression of genes involved in immune checkpoints, interferon signaling, cytokine signaling, T and B cell signaling, tumor antigens, and tumor proliferation markers. This panel is designed to be compatible with FFPE samples with low input requirements of 10 ng total RNA. The 400-gene amplicons are designed using AmpliSeq designer (https://www.ampliseq.com/browse.action). When multiple isoforms exist for a gene, the design targets multiple isoforms and allows for the quantitative measurement of the gene expression profile. To streamline the data processing and analysis, we developed an analysis package that automates QC assessment, read alignment, and read counts. Additionally, we implemented a normalization strategy beyond simple accounting for sequencing depth using a composite of 12 housekeeping genes. This approach leads to more accurate estimates of gene expression and fold changes. The analysis package is integrated into the Torrent Suite™ Software for automated sequencing data analysis and analysis report generation. We performed expression profiling analysis on 16 non-small cell lung cancer (NSCLC) FFPE samples, and demonstrated robust detection for expression of key cytokines including interferon gamma.

2854W

Regulation of the immune response by long non-coding RNAs in healthy subjects and stage IV melanoma patients. *Y. Zhang[1,2], L. Wang[1,2], S. Felts[3], V. Keulen[3], L. Pease[3].* 1) Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD; 2) Division of Biostatistics and Bioinformatics, University of Maryland Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD; 3) Department of Immunology, Mayo Clinic College of Medicine, Rochester, MN.

Tremendous advancements in genome-wide transcriptome annotation over the last decade have revealed that long non-coding RNAs (lncRNAs) are pervasively transcribed in human genome, which are emerging as key regulators of diverse cellular processes. However, functions and expression profiling of lncRNAs still remain elusive in the human immune system. In this study, we investigated the expression profiles of lncRNAs in CD4+, CD8+ and CD14+ peripheral blood cells from both healthy subjects and stage IV melanoma patients. First, we performed RNA-seq on 134 blood samples from 10 healthy subjects and 10 stage IV melanoma patients. In total, we reconstructed 346,746 RNA transcripts, 781 of which were significantly differentially expressed between healthy subjects and melanoma patients, including 512 in CD4+ cells, 304 in CD8+ cells and 79 in CD14+ cells. Among these RNA transcripts, we further identified 12,946 expressed lncRNAs including 9,296 annotated by GENCODE V23, 3,319 overlapping with known lncRNAs in GENCODE V23, and 331 novel long intergenic non-coding RNAs (lincRNAs) using our computational bioinformatics pipeline. 25 lncRNAs showed differential expression patterns between healthy and melanoma patient groups in at least one of CD4+, CD8+, and CD14+ cells. The co-expression analysis indicated that lncRNAs were correlatively expressed with genes involved in immune response and cancer progression. Our studies suggest that lncRNAs play crucial regulatory roles in human immune system and may shed light on the pathogenesis of melanoma.

2855T

A text mining and temporal analysis of cytogenetic data correlated with gene expression profiles reveals a dependence network of chromosome aberrations outlined by early and late cytogenetic events in prognostic subgroups of diffuse large B-cell lymphoma. *R. Garcia, P. Koduru.* UT Southwestern, Dallas, TX. 5323 Harry Hines Blvd, Dallas, TX 75390.

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous group of neoplasms characterized by distinct morphological, immunophenotypic, cytogenetics and molecular features. In the present study, we examined the Mitelman and NCBI databases to construct a dependence network of recurrent cytogenetic aberrations (RCAs) in prognostic subgroups of DLBCL and correlate findings with gene expression. In addition, a temporal analysis was performed to outline early and late cytogenetic events in tumor progression. The R package containing the *tm* library was used to analyze the cytogenetic nomenclature and obtain chromosome aberrations and the *pvclust* library was used to design a hierarchical cluster of RCAs. Clusters with p-value <.05 were considered significant. Clusters were assigned based on established RCAs in DLBCL subtypes or literature review. To characterize RCAs at the molecular level, a gene expression profile was performed. The GSE26812 dataset from NCBI was used for this purpose and the GEO2R was used to obtain differently expressed genes. To outline clonal evolution, a principal component analysis and temporal analysis was performed in R. We identified 1433 cases of DLBCL. There were 235 RCAs that appeared at least 5 times but only 195 were used. Among these included: +18, t(14;18), -15, -17, -13, t(8;14), -10, -16, -20, -11,-18, add(14)(q32), -19, t(3;14). In terms of a dependence network of RCAs, we identified 61 significant clusters. Of the most prominent included: +18, +16; i(18)(q10),i(21)(q10); +19, +11; add(18)(p11), +i(18)(q10), del(9)(p), del(1)(q42) in ABC-DLBCL and t(14;18), dup(1)q; +X, +12; add(10)(q24), add(12)(q24); add(14)(q32), add(10)(q22); +15, del(1)(q32) in GCB-DLBCL. A cluster containing add(16)(p13) was assigned as PMBCL. Furthermore, our analysis showed correlations between clustered RCAs in DLBCL subtypes and gene expression, mainly described by dosage compensation in some instances or overexpression and under expression in others. Early cytogenetic events in tumor progression included largely translocations and gains of chromosomes, while monosomy of chromosomes was characterized by late events. In summary, our results demonstrate clonal evolution of unique groups of RCAs in DLBCL subtypes that may be used to predict outcome and further stratify DLBCL. Key words: DLBCL, RCAs, text mining, clonal evolution, gene expression.

2856F

Machine learning to prioritize cancer variants for screening of cell free DNA (cfDNA). *C. Cario, L. Leong, J. Witte.* Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California.

Cell free DNA (cfDNA), a form of 'liquid biopsy', has recently emerged as a promising technology to screen, diagnose, and monitor many types of disease. In the context of cancer, it provides several benefits over traditional biopsy in examining the genetic structure of a tumor that include a simple and non-invasive sampling nature and the ability to detect subclonal variation. One unfortunate complication of cfDNA as a biomarker technology is a weak signal-to-noise ratio that arises primarily from contamination by DNA from healthy tissue. To circumvent this issue, many researchers have narrowed focus to probe a limited set of hand-picked coding variants within a single gene or a small gene panel, trading breadth for depth to improve signal. This approach, while sensible, can lead to detection issues in highly heterogeneous diseases-- like prostate cancer-- in which variants are not universally present at high proportion across patients. It also presents a scenario where bias can be induced when elucidating the underlying biological mechanisms of disease since hand-picked variants often exclude intergenic regions of the genome which may have functional importance. Our group proposes an unbiased approach to the design of a variant panel to screen and monitor heterogeneous cancers. Using prostate cancer to demonstrate this approach, a support vector machine (SVM) is trained on known genome wide tumor variants from the International Cancer Genome Consortium (ICGC) and an equal number of randomly simulated mutations. Using biologic annotation from a dozen databases, the machine learns the weighted combination of annotations most predictive of prostate cancer, and creates a model that can be used to prioritize variants for screening. Through cross validation, we demonstrate superior performance of this SVM directed panel over two hand-curated panels and a frequency-based approach in detection ability. To further assess the performance of this panel, samples from a cohort of over 100 prostate cancer patients undergoing treatment at the UCSF hospital system are being collected to determine the panel's sensitivity and specificity in detecting tumor variants. These results are being compared to identified variants from sequence data of multiple tumor foci in each patient. Clinical applications of the panel, such as the monitoring of residual disease, are also being evaluated.

2857W

Identify circulating DNA's tissue-of-origin in cancer by whole genome sequencing. *Y. Liu[1,2], S. Reed[2], G. Ha[2], S. Freeman[2], V. Adalsteinsson[2], M. Kellis[1,2].* 1) Massachusetts Institute of Technology; 2) Broad institute of MIT and Harvard, Cambridge, MA.

Circulating DNA (ctDNA) has been shown as an emerging non-invasive bio-marker to monitor the tumor progress in cancer patients. In late stage cancer patients, elevated ctDNA has been found not only from tumors, but also from normal tissues. Thus, the identification of ctDNA's tissue-of-origin is critical to understand the mechanism of tumor progression. Recent efforts to identify ctDNA's tissue-of-origin begin to utilize ctDNA's epigenomic status, such as DNA methylation and nucleosome spacing. However, both of these methods have their own limitations: (1) For nucleosome spacing, lack of reference nucleosome maps in different tumor and normal tissues has limited its application to tissue-of-origin deconvolution; (2) For DNA methylation, large DNA degradation during whole genome bisulfite sequencing (WGBS) library preparation, even with current low-input DNA technology, is still the major hurdle for its clinical application, although extensive DNA methylation studies by WGBS in tumor and normal tissues during the last decade have already provided enough reference maps. Very recently, a pioneer study showed significant differences between DNA fragment lengths of methylated and unmethylated ctDNA. Here, taking the advantage of this experimental observation, we developed a computational approach to identify ctDNA's tissue-of-origin by inferring its DNA methylation pattern from fragment information in whole genome sequencing (WGS). We systematically explored the correlation between DNA methylation measured by WGBS and fragment information measured by WGS from ctDNA of the same breast cancer patients. Furthermore, we deconvoluted ctDNA's tissue-of-origin status by using only fragment information from WGS in ctDNA and DNA methylation level from public available tumor and normal WGBS datasets. Overall, our methods here paved the road for ctDNA's application in clinical diagnosis.

2858T

Mammalian species conservation data and the implication for clinical variant classification. *P. Taylor, A. Alvarez-Buylla, Y. Kobayashi, S. Yang, K. Nykamp, S. Topper.* Invitae, San Francisco, CA, USA.

Functionally critical amino acid residues are subject to evolutionary pressure and are expected to be conserved, whereas non-critical positions may be more tolerant to variation. As such, evolutionary conservation data may be a powerful component of predicting the deleterious effect of missense mutations. In practice, however, the predictive power of existing conservation-based in silico algorithms has been inadequate for broad clinical application; the 2015 ACMG guidelines for the interpretation of sequence variants (ISV) consider consensus among multiple algorithms to be only supporting evidence for variant classification. However, the 2015 ACMG ISV also discusses an alternative application for conservation data: the presence of the variant amino acid change in multiple non-human mammalian species is designated strong evidence for a classification of benign. This distinction reflects the assumption that variation within the mammalian clade speaks more directly to questions related to mammalian physiology and is therefore more relevant to the question of human disease. To assess the predictive power of this type of data, we selected 6572 missense variants from our database and examined conservation across 61 non-human mammalian species. We identified 308 missense variants classified as pathogenic (P) or likely pathogenic (LP), and only two (0.6%) of these were present in multiple species; both were in genes associated with cardiovascular defects (*TTR* p.Val142Ile; *KCNH2* p.Ala913Val). By contrast, in silico algorithms predicted eight (2.6%) of the P/LP missense mutations to be tolerated. In 228 of 1427 (16.0%) missense variants classified as benign, the variant amino acid change was present in multiple species, which indicates a clear depletion of these events for deleterious variants. Our analysis confirms and extends the approach proposed in the 2015 ACMG ISV guidelines and suggests that non-conservation among mammalian species may be more accurate than complex in silico algorithms as an indicator that a variant is benign.

2859F

Improving somatic mutation identification in highly variable genomic regions using graph-based reference genomes. *H. Gibling[1,2], J.T. Simpson[1,3], P. Awadalla[1,2].* 1) Informatics and Bio-computing, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Computer Science, University of Toronto, Toronto, Ontario, Canada.

The traditional method of comparing high-throughput sequencing (HTS) reads to a single haploid reference genome is useful for discovering somatic mutations, but less effective in regions of high variability between individuals in a population. One of the most polymorphic regions in the human genome is the major histocompatibility complex (MHC), containing over 200 human leukocyte antigen (HLA) genes which encode proteins that present foreign antigens to immune cells. There are thousands of known HLA alleles throughout the human population. Similarly, over 30 *PRDM9* alleles have been identified thus far, allowing for variation in the location of double strand breaks during meiotic homologous recombination. Variation in cancer HTS data make identifying somatic mutations difficult, and valuable information is lost in unmapped reads, particularly in regions of high variability. Both HLA genes and some alleles of *PRDM9* play roles in cancer development or progression, emphasizing the importance of accurately mapping sequencing reads to these regions when studying somatic mutations. To tackle difficulties in mapping HTS reads to highly variable regions, we exploit the flexible properties of sequence graphs. These graphs represent a genome by combining the reference sequence with known variation within the same data structure. HTS reads can align to variants including single nucleotide polymorphisms, insertions, and deletions by following a path through the graph. This flexibility resolves the issues that arise from mapping to a single haploid reference sequence, including uneven coverage in polymorphic regions and a bias towards overestimating reference allele frequencies. In addition, novel variants detected from HTS reads can be added to the graph to maintain a dynamic reference graph. Using the vg software package, we built two sequence graphs: one for the MHC locus using the reference and alternate haplotypes and SNPs from the 1000 Genomes Project, and one for *PRDM9* using allelic sequences from the polymorphic zinc finger array. Using whole genome sequencing data from the Pan Cancer Analysis of Whole Genomes (PCAWG) project, we intend to evaluate whether reference graphs improve coverage, variant calling, and allele typing compared to traditional mapping methods for these variable regions.

2860W

Discovery of mosaic point mutations in pediatric cancer using whole-exome sequencing of DNA extracted from blood. *E. Karlins[1,2], L. Mirabello[1], M.J. Machiela[1], R.A. Kleinerman[1], J.N. Sampson[1], B.L. Gallie[3,4,5], D.E. Rushlow[5], B. Hicks[1,2], B. Zhu[1,2], M. Dean[1], M. Yeager[1,2], S.J. Chanock[1], M. Tucker[1], S.A. Savage[1], L. Morton[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD 20877, USA; 3) Department of Ophthalmology and Vision Sciences, The Hospital for Sick Children, 555 University Ave., Toronto, ON M5G 1X8, Canada; 4) Departments of Ophthalmology & Vision Sciences, Molecular Genetics and Medical Biophysics, Faculty of Medicine, University of Toronto, Toronto, Canada; 5) Impact Genetics, Bowmanville, Ontario, Canada.

Germline mosaic mutations are known to contribute to pediatric cancer risk. In retinoblastoma it is estimated that ten percent of germline *RB1* mutations are mosaic. When a cancer risk gene is known, as is the case for retinoblastoma, deep sequencing or other techniques can be used to detect mosaic mutations, however when a specific gene is not known, deep sequencing on an exome level to detect mosaic mutations would be extremely expensive. Mosaic mutations are likely to be missed by traditional whole-exome sequencing (WES) variant callers since the mutant allele will be at too low frequency for variant discovery without also generating an excess of false positive variant calls.   Here we explore computational methods for mosaic mutation discovery from WES data using a set of validated blood DNA *RB1* variants from 89 retinoblastoma cases with Illumina WES. Eight of these 89 variants were classified as mosaic by the premier retinoblastoma genetic testing facility, and three of the mosaic mutations were not discovered by traditional exome variant calling. Relaxing variant caller parameters identify these mosaic variants, along with many false positives. We create a model of true positive mosaic variants using sequencing read level metrics as input to the model. Included in the model are metrics that are associated with a WES variant being a true positive and independent of proportion of reference alleles to alternative alleles: position of alternative alleles on the read, mapping quality of reads with alternative alleles, and base quality of alternative alleles as compared to reference alleles, for example. We then apply this model to a test set of 360 osteosarcoma cases with both Illumina WES data and targeted *TP53* Ion AmpliSeq data. For each input variant, the output of the model is a probability of the variant being a true positive. The user can set a probability threshold as desired based on specificity and sensitivity levels appropriate for his or her study.   Our methods can aid variant discovery after WES and rare variant filtering highlight candidate genes. In some instances, adding variants in one or two additional cases may be all that is needed to drive a burden test over the significance threshold. Using techniques like these to discover new variants from existing data can be a cost-effective step before variant validation and help inform decisions on whether to perform follow-up candidate gene deep sequencing.

2861T

AmpliconArchitect: A computational approach to elucidate complex rearrangements in genomes of virus-driven cancers. *V.B. Deshpande[1], N. Nguyen[1], H. Carter[2], R.B. Schwab[3], V. Bafna[1].* 1) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA; 2) Department of Medicine, Division of Hematology/Oncology and Moores Cancer Center, University of California San Diego, La Jolla, CA; 3) Department of Medicine, Division of Medical Genetics, University of California San Diego, La Jolla, CA.

Viruses have been implicated in 15% of all human cancers (11% of all cancer deaths). The integration of the viral genome into the host genome of a normal cell often precedes its transformation into a tumor cell. Viral integrations are reported to cause copy number amplification and structural variations near the integration site on the human chromosome. Existing tools detect integration sites but are unable to resolve the architecture of the human-viral chimeric amplicon, which consists of complex tandem duplications due to looping/episome formation of the chimeric region. Here we present a novel computational approach, AmpliconArchitect(AA), to automatically identify the amplified genomic regions and reconstruct plausible amplicon architectures. AA takes as input whole genome sequencing (WGS) paired-end reads aligned to a combined human-virus reference, using the following steps: (a) Use discordant read-pair alignments and coverage information to iteratively visit and extend connected genomic regions with high copy numbers. (b) Segment amplified regions using a mean-shift approach to detect copy number changes. (c) Use discordant read-pair clusters to construct a breakpoint graph connecting segments. (d) Compute a maximum likelihood network flow to estimate copy counts of genomic segments. (e) Report paths and cycles in the graph that identify the dominant linear and circular structures involved in the chimeric human-viral amplicon. We applied AA to 21 matched cervical cancer samples (20 TCGA+1 anonymous). All samples contained HPV DNA (HPV 16:11; HPV 18: 7; HPV 33, HPV 45, HPV 70: 1). Host integration was detected in 20/21 tumor samples but in no matched normal sample. All amplicons contained the *E6/E7* oncogenes, and 9/21 samples showed deletion in all or part of the *E2* gene, a key repressor of *E6/E7*. AA reported architecture for 16/21 samples, after discarding samples with low copy counts. Of these 16 samples, 3 displayed a simple viral insertion at the site; 5 displayed a unique simple cycle with one human and one viral segment; and 8 samples displayed complex cycles or heterogeneous mixtures of multiple structures. 13 amplicons contained parts of human genes, with 2 containing truncated forms of the oncogene *ERBB2*. Conclusion: AA is a useful tool to understand amplification, heterogeneity and rearrangements at viral integration sites. This will provide crucial insight into the mechanistic/functional aspects of viral integration in cancers..

2862F

Predicting regulatory variants with composite statistic. *J. Wang[1], M.J. Li[2,3], Z. Pan[3,6], Z. Liu[3,5], J. Wu[2], P. Wang[1,3], Y. Zhu[3,4], F. Xu[3,4], Z. Xia[5], P.C. Sham[3,6], J.P.A. Kocher[1], M. Li[3,6], J. Liu[2].* 1) Health Science Research, Mayo Clinic Arizona, Scottsdale, AZ; 2) Department of Statistics, Harvard University, Cambridge, Boston, MA 02138-2901, USA; 3) Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 4) School of Biomedical Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 5) Department of Anaesthesiology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 6) Department of Psychiatry and Centre for Reproduction, Development and Growth, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

Motivation: Prediction and prioritization of human noncoding regulatory variants is critical for understanding the regulatory mechanisms of disease pathogenesis and promoting personalized medicine. Existing tools utilize functional genomics data and evolutionary information to evaluate the pathogenicity or regulatory functions of noncoding variants. However, different algorithms lead to inconsistent and even conflicting predictions. Combining multiple methods may increase accuracy in regulatory variant prediction. **Results:** Here, we first compiled an integrative resource for predictions from eight different tools on functional annotation of noncoding variants. We further developed a composite strategy to integrate multiple predictions and computed the composite likelihood of a given variant being regulatory variant. Benchmarked by multiple independent causal variants datasets, we demonstrated that our composite model significantly improves the prediction performance. **Availability:** We implemented our model and scoring procedure as a tool, named PRVCS, which is freely available to academic and nonprofit usage at http://jjwanglab. org/PRVCS. **Contact:** junwen@uw.edu, jliu@stat.harvard.edu, limx54@gmail. com.

2863W

Accelerating medical genomics through functional mapping of clinical genomic variant data to underlying disease biology. *L. Brody, A. Ajetunmobi, V. Dillard, N. Humphreys-Kirilov, R. Doyle.* Genomic Services, Desktop Genetics, London, United Kingdom.

The wealth of data output from next-generation sequencing (NGS) technologies have led to tremendous progress in the study of the evolutionary origins of human genetic variation and its contribution to human disease. In particular, several consortia have been developed to map human genetic variation, providing genome-wide association studies and rare variant analyses in the context of common and rare diseases. However, many of these variant databases do not provide information on the function of these variants and their contribution to the pathophysiology of human diseases. Without elucidating the functional contribution of these variants to cellular activity, these studies will yield little benefit to human health. It is clear that high-throughput characterization of the molecular and cellular activity of genetic variation is becoming increasingly important for understanding the mechanistic function of the human genome. At Desktop Genetics, we have developed an integrated approach for end-to-end NGS data analysis; from variant detection to functional profiling. We have developed a custom bioinformatics pipeline which can be implemented for genome alignment including, single nucleotide polymorphism (SNP), small insertion/deletion (InDel), and copy number variation (CNV) detection of whole genome/exome sequencing (WGS/WES) data from the Illumina platform. Using sequencing data, and our optimized gRNA design rules, we have implemented a bespoke CRISPR genome editing methodology for *in vitro* functional assays in cultured human cells of a defined genotype. CRISPR/Cas9 in conjunction with customizable guide RNA is used to precisely target genetic and epigenetic regions of interest to induce loss-of-function alterations and validate predicted variant function. This may provide a unique approach for downstream variant interpretation and interrogation of their role in the pathogenesis of rare and common genetic diseases.

2864T

Consideration for applying machine learning methods to GWAS data to predict disease patients. *P. Ruan, J. Sese.* Artificial Intelligence Research Center, National institute of advanced industrial science and technology, Tokyo, Japan.

The conventional analysis in GWAS computes a P-value on each SNP, and a chromosomal region around a statistically significant SNP is supposed to contain a causal mutation of the target disease. However, GWA studies for common or complex diseases often found many SNPs whose odds ratios are moderate, and summation of them can explain only part of the cause of the disease. To overcome the problem, we here test the use of machine learning methods to the analysis of GWAS. We apply cutting-edge machine learning techniques such as deep learning and gradient tree boosting as well as conventional methods such as Support Vector Machine, random forest and logistic regression to the prediction of disease/non-disease patients. In this study, we use a GWAS dataset consisting of 1, 231 postmenopausal breast cancer patients and 1,203 control subjects. To calculate the accuracies for the method comparison, we perform 100 times of 10-fold cross-validations. In our procedure, SNPs whose P-values are less than 0.05 are selected, and discriminant analysis is performed with the SNPs. Then, accuracies (F-measures) on test data are calculated with the discriminant function. We report the comparison of the prediction accuracies of the results including how the results are different/similar among the machine learning methods and how sensitive to parameters in each technique.

2865F

Identification of target lincRNA associated with prostate cancer risk loci. *S. McDonnell, A. Nair, Z. Fogarty, S. Riska, M. Larson, A.J. French, L. Tillmans, M.S. DeRycke, D.J. Schaid, S.N. Thibodeau.* Dept Health Sci Research, Mayo Clinic, Rochester, MN.

Multiple risk loci for prostate cancer have been identified; however determining which genes the loci influence is difficult. Previously we identified 51 regions with significant eQTL signals associated with 88 genes for prostate cancer risk while creating a normal prostate tissue eQTL database from 471 normal prostate tissue samples. Long intergenic non-coding RNAs (lincRNAs) are an abundant class of non-coding RNAs that serve as key regulators for many cellular processes and thus can be potential diagnostic markers and therapeutic targets for diseases such as cancer. Expanding on our earlier work, we have quantified lincRNA in our mRNAseq dataset and performed eQTL analysis correlating expression of lincRNA with 220 prostate cancer (PCa) risk variants. LincRNA were quantified using an Ultrafast and Comprehensive lincRNA pipeline, called UClincR, to identify known and novel lincRNAs using the transcriptome sequencing data. In total, 54,004 lincRNAs were identified; of which 11,212 have median raw count > 7. All PCa-risk SNPs and those SNPs within +/-100Kb and in LD (r^2>0.5) (n=8,073 SNPs) with the PCa-risk SNP were tested against all lincRNA located within 1.1Mb (n=1,169 genes) adjusting for histologic characteristics and lincRNA expression principal components. 64 lincRNA in 23 risk-regions demonstrated a significant eQTL signal using a Bonferroni significance threshold (P≤ 2.5E-07). Future planned studies include expanded analysis of the 64 lincRNA to identify all potential eQTL variants as well as utilizing public datasets to map the significant SNPs to prostate-specific known and predicted regulatory elements, including promoters, enhancers, transcription factor binding sites, transcription start and stop sites.

2866W

Statistical approaches for network-driven discovery and interpretation of cancer driver mutations. *J. Reimand[1,2].* 1) Department of Informatics and Biocomputing, Ontario Institute for Cancer Research, Toronto, Canada; 2) Department of Medical Biophysics, University of Toronto, Toronto, Canada.

   Identifying driver mutations in whole genome sequencing data of cohorts of cancer samples is essential to deciphering tumour biology and designing precision treatments. Information on pathways and molecular interaction networks can improve interpretation of cancer mutations and associating mechanism and clinical information. We hypothesize that many cancer driver mutations precisely modify network interfaces encoded in small sites in proteins, RNA and DNA, leading to interaction losses and gains in networks. We have developed novel computational strategies to find network-associated cancer driver mutations and infer their impact on network topology that are complementary to state-of-the-art approaches. Our mutation enrichment algorithm ActiveDriver-2 discovers coding and non-coding cancer driver elements from whole-genome sequencing data. It assumes that frequent mutations such as SNVs and indels in network-associated sequence sites indicate interaction hotspots in cancer. This approach is applicable in signalling pathways and gene regulatory networks where site-specific interactions between proteins, RNA and DNA are known or predicted computationally. ActiveDriver-2 determines genome elements and sites with statistically significant mutation recurrence by accounting for trinucleotide mutation rates using a Poisson regression model. Our machine learning method MIMP further complements this strategy by annotating mutations that erase existing binding sites and/or creating new binding sites and thus cause rewiring of interaction networks. After validating the statistical properties of our model with synthetic genome mutation data, we analysed the collection of 2,800 cancer whole genomes in the ICGC Pan-cancer Analysis Working Group (PCAWG) effort. We revealed mutation hotspots coinciding with network-associated sites, including post-translational modification sites of proteins, microRNA target sites in gene UTRs, and transcription factor binding sites in gene promoters and enhancers, suggesting novel mechanisms of action of cancer mutations. Our study emphasizes the importance of network-related variation in cancer by providing novel interpretation to known cancer mutations, helping find new cancer drivers and risk modifier alleles, and characterising their biological mechanisms.

2867T

Inferring evolution structure based on copy number aberration from head and neck cancer via penalized two-way mixture Poisson model. *A. Tai[1], W. Hsieh[1], C. Peng[2], S. Peng[2].* 1) Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan, R.O.C; 2) Departments of Resource Center for Clinical Research, Chang Gung Memorial Hospital, Taoyuan, Taiwan, R.O.C.

   Somatic evolution describes the accumulation of mutations in the cells during an individual's life time. The mechanism is important for the development of cancer. The development of cancer is a multistep process and it results in tumor heterogeneity, meaning that each tumor might consist of genomes from multiple clones. There are many computational approaches developed to infer the clonal evolution from the sequencing data. Most of them are based on single nucleotide variants (SNVs). However, the accuracy of these algorithms is usually affected by the copy number aberration. How to effectively identify the copy number aberration is an issue that must be addressed. Furthermore, copy number aberration itself forms an even more complicated problem of tumor heterogeneity. We developed a penalized two-way mixture Poisson model to study the cancer evolution based on copy number aberration. In our model, the clonal architecture is automatically resolved by the penalized function in the likelihood and it is a clustering method that can determine the corresponding copy number state. Finally, we applied our model to the sequencing data of head and neck cancer from TCGA to understand the distribution of copy number aberration and clonal structures.

2868F

Epigenome-based drug repositioning for acute myeloid leukemia. *A.S. Brown, C.J. Patel.* Biomedical Informatics, Harvard Medical School, Boston, MA.

   Repositioning approved drugs for the treatment of new indications is a promising avenue to reduce the burden of drug development. We recently described a novel repositioning method, ksRepo, that enables investigators to move beyond microarray-based gene expression and utilize a variety of other sources of molecular evidence, such as DNA methylation differences. In this pilot study, we applied ksRepo to DNA methylation information from three independent acute myeloid leukemia (AML) datasets. We ranked genes according to their degree of differential methylation between AML blasts and normal bone marrow aspirates. Next, we used ksRepo to predict drug candidates derived from a list of curated gene-drug interactions from the Comparative Toxicogenomics Database. In addition to recovering the most commonly prescribed frontline therapy, cytarabine, ksRepo also found significance for decitabine, an epigenetic drug commonly used in AML treatment. We further observed enrichment for sex hormone-related drugs among the KS-significant drugs, congruent with recent findings in AML. In conclusion, ksRepo provides a platform for deriving repositioning hypotheses from differential DNA methylation and yields clinically relevant predictions.

2869W

Accurate prediction of breast cancer risk based on flow variant analysis of circulating peripheral blood B cells. *K. Upadhyay, M. Syeda, J. Loke, H. Ostrer.* Albert Einstein College of Medicine, Bronx , NY.

   Introduction. About 10% cases of breast cancer occur to women from high-risk families with multiple affected relatives. Heritable mutations in genes in the double strand break (DSB) repair pathway can be detected in over 25% percent of these cases. In up to 40%, a variant of uncertain significance (VUS) is identified and commonly 2 or more VUS are observed. In addition to the issue of prevalent VUS, most sequencing methods are biased in scope and depth, because certain genomic regions can be difficult to capture, amplify or assemble. Previously, we demonstrated that flow variant assays (FVAs) that quantify native and phosphoproteins, protein nuclear localization and protein-protein interactions in the DSB repair pathway were altered in LCLs from individuals with mutations in *BRCA1* and other genes in this pathway. Development of an accurate risk classifier based on logistic regression analysis of multiple FVAs of circulating B cells could improve risk assessment for these individuals and their family members. **Materials and methods.** Nuclear localization and phosphoprotein FVAs were performed on 37 LCLs (discovery cohort) with mutations and benign variants in *BRCA1, BRCA2* and other genes in this pathway and on circulating B cells from 29 women (replication cohort), 20 of whom had breast cancer or ovarian cancer, no *BRCA1* nor *BRCA2* mutation and 2 or more affected relatives, and 9 healthy women with 1 affected relative. Logistic regression classifiers were calculated for 2 FVA and 3 FVA models. Accuracy was calculated for single FVAs and 2 FVA and 3 FVA classifiers in discovery and replication cohorts. In addition, whole genome sequencing was performed on all of the women with breast cancer in the replication cohort. **Results.** Logistic regression classifiers improved the accuracy for prediction of breast cancer risk. The range of accuracies was 86-89% in the discovery cohort for single FVAs, 92% for the 2 FVA classifier and 95% for the 3 FVA classifier. The range of accuracies for single FVAs was 60-90% for single FVAs in the replication cohort, 93% for the 2 FVA classifier and >99% for the 3 FVA classifier. Notably, all women had only VUS. **Conclusion.** The high degree of accuracy of 2 and 3 FVA classifiers suggests that they could be used as an independent risk assessment tool, if validated in larger, independent cohorts. The ready accessibility of analyzing circulating B cells would facilitate clinical use of FVA classifiers for prediction of breast cancer risk.

2870T

Gene set analysis methods: A systematic comparison. *R. Mathur[1], A. Motsinger-Reif[1], A. Shojaie[2].* 1) Bioinformatics Research Center, North Carolina State University, Raleigh, NC; 2) Department of Biostatistics, University of Washington, Seattle, WA.

Pathway or gene set analysis is a useful method to summarize and interpret high-dimensional genetic data. While numerous methods have been developed to conduct this type of analysis, the results from these methods are often vastly different. Methods comparison studies have typically utilized benchmark datasets only, limiting true evaluation of the statistical performance of the gene set analysis methods. In this study we have devised a systematic simulation study, based on real data, with a range of effect sizes and pathway characteristics, to conduct an extensive comparison of the most commonly used gene set/pathway tools. Our simulated data is based on the true correlation structure of a Prostate Cancer microarray dataset, with signal introduced in the TGF-β signaling and KRAS pathways. Using this data, we present and compare the power and false positive rates of Significance Analysis of Function and Expression (SAFE), sigPathway, Gene Set Enrichment Analysis (GSEA), and Correlation Adjusted Mean RAnk (CAMERA) methods, using the Molecular Signatures Database (MSigDB) gene set database as the knowledgebase for each method. Our estimated power and false positive rates rigorously evaluate the statistical properties of these approaches, and provide guidance for users on choosing and interpreting these methods.

2871F

Extraction of long-range sequence and phasing information and characterization of structural variation from FFPE samples using a novel *in situ* **digestion and ligation protocol.** *N. Putnam[1], S. Balakrishnan[1], R. Calef[1], P. Hartley[1], J. Stites[1], C. Troll[1], B. Rice[1], M.P. Powers[1], R.E. Green[1,2].* 1) Dovetail Genomics, LLC, Santa Cruz, CA; 2) University of California, Santa Cruz, Santa Cruz, CA.

Presented herein is a novel technique for the generation of megabase-scale phased sequence data via pre-extraction *in situ* DNA preparation from FFPE samples. Until now, FFPE samples typically yielded sub-500bp DNA fragments, confounding the acquisition of long-range sequence information. Long-range sequence information can only be obtained from FFPE tissue in non-repetitive loci where reads overlap unique breakpoints or other unique sequence features. Because the human genome is >50% repetitive, human FFPE samples have to date not provided adequate long-range genome information. In this first of its kind proof-of-principle study we: (i) remove paraffin from FFPE treated cell lines (GM24149, GM24385), (ii) treat tissue fragments *in situ* with restriction endonuclease, (iii) fill-in the resulting recessed 3' termini, and (iv) ligate the resulting ends, creating chimeric molecules. These chimeric molecules preserve the long-range information of the starting molecule due to proximity-based ligation. Sequencing provides data similar to those produced by Hi-C or Chicago™. The data produced demonstrate that phasing and long-range information is preserved from these samples with >90% accuracy up to 1.9 MB. This technique provides a powerful new way to extract genome information for archives of tumor samples and other FFPE samples from bio-banks including the discovery of structural variation, copy-neutral structural variants, and other long-range sequence variants within (or mediated by) repetitive elements.

2872W

Identification of cancer antigens in NSCLC using exome sequencing and mutated antigen screening with gBlock Mini-genes. *Y. Bao[1], L. Zhang[2], C. Qian[2], B. Yang[2], B. Yin[2], W. Li[2], B. Ning[2], F. Wang[2], C. Hu[2], H. Loi[1], C. Chen[1], H. Wang[2], R. Wang[2].* 1) Research and Development, Integerated DNA Technologies, Redwood City, CA; 2) Center for Inflammation and Epigenetics, Houston Methodist Research Institute, Houston, TX.

Tumor-specific mutated antigens may serve as ideal targets for immunotherapy and precision medicine. Although several hundreds of cancer antigens have been identified, only a few are suitable for immunotherapy using T cell receptor (TCR) or chimeric antigen receptor (CAR) based therapy. Thus, identification and selection of cancer antigens are vital to develop personalized immunotherapy in cancer. Targeted exome sequencing has made it possible to routinely detect cancer mutations for each patient. Lung cancer is one of the most frequently mutated cancer types. Thus, we hypothesize that patient-specific cancer antigens can be identified and used as immune targets for personalized immunotherapy of advanced lung cancer patients. In this study, exome sequencing of both tumor and isolated T-cell samples in 30 NSCLC patients and MS analysis were used to detect tumor-specific mutations. Following the identification of somatic mutations, 20-30 mutated antigen candidates per patient were screened by transient expression of mutated antigens cloned into an expression vector through gBlock mini-genes to generate fusion proteins. Of 89 CD4+ T cell clones used for the screening, four mutated antigens were identified. Among all the 89 T cell clones, 83 of them have response to 4 neoantigens, and 67 of them recognized one dominated antigen. Our work demonstrates that the method described above can effectively identify potential targets for T cell based cancer immunotherapy.

2873T

Sensitive and reliable variant detection from challenging samples. *N. Fang, K. Heitz, R. Tolun.* QIAGEN GmbH, Hilden, Germany.

The rapidly developing next generation sequencing (NGS) technologies provide highly sensitive methods in detecting and characterizing variants in clinical samples. However, clinical samples often come in limited quantity as well as compromised quality. Such samples are not suitable for standard NGS library construction methods, which commonly require hundreds of nanograms of good quality DNA. Examples of such challenging clinical samples are laser capture microdissection (LCM) samples, FFPE samples, and circulating DNA. Here we describe an optimized workflow to efficiently convert small amount of DNA samples into sequencing libraries. The library construction protocol is based on the QIAseq UltraLowInput Library Kit, which has optimized enzyme and buffer formulations that enable high library conversion rate as well as unbiased library amplification from as low as 10pg input DNA. In combination with an optimized hybrid-capture-based target enrichment procedure, the workflow described in this poster enables reliable mutation detection even with low DNA input amount, making it feasible to gain sequence insights from challenging sample types.

2874F

Low level somatic variant detection by Sanger sequencing of formalin-fixed paraffin-embedded (FFPE) samples for Precision Oncology. *A. Gerstner, E. Schreiber, S. Schneider, S. Jackson, K. Varma.* Thermo Fisher Scientific, South San Francisco, CA.

   Deleterious sequence variants play an important role in the initiation and progression of many different cancer types. The detection of germline variants by the gold standard Sanger sequencing has been well established, however, the detection of somatic mutations, especially in heterogeneous tumor samples where variants may be present at a lower level, has been more challenging.To facilitate analysis of somatic mutations in tumor samples, we have developed Sanger sequencing panels that cover the entire coding regions of specific genes (e.g. TP53) as well as an extended panel encompassing 66 frequently detected oncogenic alleles from 18 genes. We have also developed companion software, Minor Variant Finder (MVF), that facilitates detection of low levels of somatic mutations by Sanger sequencing. To demonstrate the workflow of these panels, we analyzed FFPE DNAs from 12 different cancer tissue types. We initially determined variants in these samples using Ion Torrent™ Personal Genome Machine (PGM™) next generation sequencing. We confirmed the identity and variant allele frequency of these variants by Sanger sequencing using 1 ng, 0.5 ng or 0.1 ng DNA. Finally, we made serial dilutions of one of these samples to establish limit of detection (LOD) at as little as 3% of a minor variant in an FFPE sample.Sanger sequencing is the gold standard for confirmation of minor variants detected by NGS. In this study, we show that Sanger sequencing of limited number of targets, in conjunction with the MVF software, can also be an ideal first line screening choice for tumor FFPE samples where limited amount of DNA is available.

2875W

Single-cell transcriptome analysis reveals heterogeneity in endometrioid adenocarcinoma tissue. *S. Hashimoto[1,2], Y. Tabuchi[3], T. Torigoe[3], K. Matsushima[4], H. Yurino[1].* 1) Graduate Sch Medicine, Kanazawa University, Ishikawa, Japan; 2) Japan Science and Technology Agency, Core Research for Evolutional Science and Technology (CREST), Tokyo, Japan; 3) Department of Pathology, School of Medicine, Sapporo Medical University, Sapporo, Japan; 4) Department of Molecular Preventive Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

   The cell populations, such as differentiated states or tissues and organs, are thought to be heterogeneous. Therefore any expression profile based on a tissue sample will blend the true expression profiles of its constituent cells. To overcome this problem, single-cell methods have been developed for both microarrays and RNA-seq. Recently several studies have been reported single cell gene expression strategies. Single-cell gene-expression profiling allows identification and characterization of different cell types and cell subtypes. Furthermore, phenotypic heterogeneity can be studied within the same cell type. To understand the cell population status, single cell analysis is required. Although several studies for single cell gene expression analysis have been reported, number of the observed cells is very limited. Therefore, we have developed novel strategy (Nx1-seq) of single-cell transcriptome analysis for thousands of single cells. In our approach, single cells are deposited in a high-density microwell plate and lysed in situ. mRNA is then captured on barcoding microbeads and reverse transcribed. The diversity of barcode on the beads was made by emulsion PCR using randomly synthesized barcode oligo DNA. The pooling of Nx1-seq data of single cell in two homogeneous cell populations provides rich and highly reproducible transcriptional profiles. In addition, gene expression patterns among libraries with high sequencing reads were close (p=0.92). In this study, we applied Nx1-seq to characterize complex heterogeneous samples in myometrial infiltration side and endometrial side of human endometrioid adenocarcinoma with squamous differentiation tissue by examining thousands of cells per experiment. Endometrial cancer is the commonest pelvic gynaecological cancer in the westernised parts of the world. These results showed the distinction of the cancer cell state with tumorigenicity, the EMT and infiltrated leukocytes in two sides. Finally, we show that established endometrioid adenocarcinoma subtype classifiers are variably expressed across individual cells within a tumor and demonstrate the potential prognostic implications of such intratumoral heterogeneity.

2876T

Oncology biomarker development: Simultaneous digital counting of nucleic acids and proteins at 800-plex. *E. Manrao, D. Kim, G. Meredith, P.M. Ross, J. McKay-Fleisch, E. Piazza, C. Ngouenet, M. Liu, P. Danaher, G. Demirkan, B. Filanoski, B. Birditt, G. Geiss, J. Beechem.* NanoString Technologies, Seattle, WA.

With the growth of biomarker discovery, identification of novel cancer signatures has revealed the limitations of relying on a single analyte class to characterize tumors or predict their responses to therapies. NanoString® has developed its nCounter Vantage™ line of products powered by novel 3D Biology™ technology to meet the need for simultaneous analysis of DNA, RNA, and proteins. Based on single-molecule optical barcoding technology, NanoString's products are designed to consume as little as 15,000-25,000 cells while utilizing the experimental power of 800-plex, cross-analyte quantification. We present two case studies of 3D Biology in action. First, we show that the 30-plex nCounter Vantage Protein Solid Tumor Signaling Pathways Panel can be used in conjunction with the 770-plex PanCancer Pathways Panel to monitor expression regulation and protein phosphorylation within clinically relevant samples, including FFPE tissues. Second, we demonstrate that our RNA and protein detection chemistries work in concert with our novel SNV-specific probes for in-depth characterization of genotype-specific expression changes in a cell culture model of BRAF inhibition. Together, these examples highlight the rich data sets and robust discovery potential provided by a 3D Biology approach to cancer biomarker research.

2877F

Pharmacogenomic prediction of cisplatin-induced nephrotoxicity in Mexican patients treated for childhood cancer. *M. Medeiros[1,2], F. Aminkeng[2,3,4], C. Jimenez-Triana[1], R. Rivas-Ruiz[5], L. Jaurez[6], M. Palomo[6], P. Clark[7], G. Castañeda-Hernández[8], C.J. Ross[2,3,4], B. Carleton[2,3,4].* 1) Laboratorio de Investigacion en Nefrologia, Hospital Infantil de México Federico Gomez, Mexico, Mexico City, Mexico; 2) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, Canada; 3) Division of Translational Therapeutics, Department of Paediatrics, Faculty of Medicine, University of British Columbia, Vancouver, Canada; 4) Child&Family Research Institute, Vancouver, Canada; 5) Coordinación de Investigación en Salud, Instituto Mexicano del Seguro Social , México D.F; 6) Departamento de Oncología, Hospital Infantil de México Federico Gómez; 7) Unidad de Investigación en Epidemiología Clínica, Hospital Infantil de México Federico Gómez, México D.F; 8) Departamento de Farmacología, CINVESTAV IPN, México D.F.

Objective: To perform pharmacogenetic associations with nephrotoxicity (NTX) in Mexican patients treated with cisplatin for childhood cancer. Methods: Retrospective study of children treated with cisplatin for solid tumors that were part of a cohort studied for adverse reactions to chemotherapy, approved by the Hospital IRB and Ethics Committee. NTX was graded as follows: normal renal function (Grade 0); asymptomatic electrolyte disorders, including an increase in serum creatinine, up to 1.5 times baseline value (Grade 1); need for electrolyte supplementation < 3 months and/or increase in serum creatinine 1.5-1.9 times from baseline (Grade 2) grade 3: increase in serum creatinine 2-2.9 times from baseline or need for electrolyte supplementation for more than 3 months after treatment completion, grade 4: increase in serum creatinine ≥ 3 times from baseline or renal replacement therapy. DNA was obtained from saliva. A 4608 SNPs ADME panel were assessed by Illumina iScan SNP genotyping platform (Illumina Inc., USA). Results: 73 patients were included, with a minimum follow up of one year after treatment completion. 53 developed NTX (72.6%), No nNTX (grade 0) was observed in 20 patients (27.3%). Grade 1 NTX was observed in 20 patients (27.3 %), grade 2 in 8 patients (10.9%) and grade 3 in 25 patients (34.2 %). Patients with NTX were younger than patients with non-nephrotoxicity, median age 5.9 years vs. 13.4 years respectively (p=0.006). The following polymorphisms were associated to NTX: *MHTFR* rs1801133 (p=0.011, OR 3.3, 95%CI 1.22, 8.98*), EPHX1* rs1051740 (p=0.035, OR 0.34, 95%IC 0.12-0.98), *SLC22A2* rs316019 (p=0.037), no association was found with *ERCC2* rs13181, *ERCC1* rs11615 y *CD3EAP* rs3212986. Conclusion: 72.6% of cisplatin treated children presented NTX. This study replicates the importance of *MHTFR* rs1801133, *EPHX1* rs1051740 and *SLC22A2* rs316019 in NTX, previously reported in adult population.

2878W

Expression of *miR-124* **and** *SPHK1* **in human breast cancer.** *M. Minouei[1], M.A. HosseinPour Feizi[2], S.A. Rahmani[1], V. Montazeri[3], R. Ravanbakhsh Gavgani[2], M. Kazem Zadeh[2], Sh. AzarBarzin[2], A. Fateh[2].* 1) Biology Department, Faculty of Basic Sciences, Islamic Azad University, Ahar Branch, East-Azerbaijan, Islamic Republic of Iran; 2) Animal Biology Department, Faculty of natural Sciences, Tabriz University, Tabriz, Islamic Republic of Iran; 3) Department of Thorax Surgery, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Islamic Republic of Iran.

   Background: Breast cancer is still one of the most leading causes of cancer mortality in the women worldwide. MicroRNAs(miRNAs), are a class of small non-coding RNAs, that post transcriptionally regulate gene expression. *MiR-124* is a brain-enriched miRNA, related to the regulation of growth, carcinogenesis and metastasis. *SPHK1*, a highly conserved lipid kinase, which phosphorylates sphingosine to the formation of sphingosine-1-phosphate (S1P). Evidence show that *SPHK1*, via the production of S1P, is an important signal molecules in the regulation of a wide range of cellular functions including cell proliferation, differentiation, as well as prevention of apoptosis. This study aims to determine and comparison of the *SPHK1* and *miR-124* expression level in tumoral and adjacent normal tissues in Breast Cancer tissues. **Methods:** Total RNA was isolated from 40 breast cancer and adjacent normal fresh tissues. Following the confirmation of the quality and quantity of extracted RNA, Complementary DNA (cDNA) was synthesized. Afterward, the expression of *miR-124* and *SPHK1* was evaluated by Real-Time PCR in 40 paired breast cancer and normal tissues. Variability in the initial quantities of mRNA was normalized to the internal control for both *miR-124* and *SPHK1*; the relative expression levels were evaluated using the $2^{-\Delta\Delta Ct}$ method. **Results:** Our study indicates that the expression level of *miR-124* was significantly reduced and the mean expression level of *SPHK1* was increased in breast cancer tissues compared with the matched normal tissues. Furthermore, correlation between *SPHK1* upregulation and reduced expression level of *miR-124* in tumoral tissues compared with those of adjacent healty tissues was evaluated in the present study. **Conclusion:** In current study increase of *SPHK1* expression was seen in the advanced stage of breast cancer, also this study confirm *SPHK1* oncogenic role. We also, showed that the expression level of *miR-124* was significantly lower in human breast cancer tissues than that of adjacent normal tissues. The decreased expression level suggests an inhibitory effect of *miR-124* on metastasis of breast cancer. In conclusion, according to our findings, *miR-124* and *SPHK1* may be probably used as a new potential therapeutic targets for the treatment of breast cancer patients.

2879T

Effects of AlCl₃ in the lipid profile of a breast cancer cell line detected by Desorption Electrospray Ionization Mass Spectrometry (DESI-MS). *R.M. Rodrigues-Peres[1], A.M. Porcari[2], J.K.R. Heinrich-Muçouçah[1], M.N. Eberlin[2], L.O.Z. Sarian[1].* 1) CAISM – Women´s Hospital, State University of Campinas - UNICAMP, Campinas, SP, Brazil; 2) ThoMSon Mass Spectrometry Laboratory, Institute of Chemistry, State University of Campinas - UNICAMP, Campinas, SP, Brazil.

   Introduction: Aluminum is used in many different chemical presentations and in a wide range of products such as food, paper, dyes, pigments, paints, glass, fuels, textiles, cosmetics and pharmaceuticals, as well as in water purification and oil refining processes. Aluminum compounds are known for biological effects on human health and its linkage to breast cancer has been hypothesized. In vitro and in vivo studies show that this metal might bind estrogen receptors, interfering with cell pathways and phenotype characteristics, as in protein expression, exhibiting neuronal effects in humans and also showing influence in hormone responses. We used the Desorption Electrospray Ionization Mass Spectrometry (DESI-MS) technique to verify the differential expression of lipids in a breast cancer cell line cultivated with and without aluminum chloride. Methods: Breast cancer cell line T47D was grown in HAM F-10 medium containing 10% FBS (control), 10, 15 and 25% of AlCl₃. Pellets were extracted (Bligh & Dyer) and analyzed using a DESI source connected to a Q-Exactive (Thermo) at a resolution of 70,000. The sprayed solvent was MeOH (2 uLmin⁻¹) and the negative ion mode was used in 100 – 1500 *m/z* range. Metaboanalyst 3.0 was used for chemometrics. Results: Mass spectra were mainly composed of ions up to *m/z* 700. Fat acids molecules and dimmers, as well as other small lipids formed the predominant ions. When partial least square discriminant analysis (PLS-DA) was applied for the *m/z* 290 - 470 range, full differentiation of the control and treated groups was achieved. Such discrimination was mainly attributed to the ions of *m/z* 350.246, 364.226, 346.215, 367.215. Based on their exact mass values, these ions were assigned as [M + Cl]⁻ adducts of acylcarnitines, sphingosines, prostaglandins and steroidal compounds. Conclusion: We observed a differential profile for ions that are mainly associated with hormones, in the control group in contrast to the treated group, independently on the AlCl₃ concentration used. These findings endorse the assumption that aluminum might be altering some pathways within the cell, in this case, affecting the lipid profile in breast cancer cells. Further studies are needed to elucidate the possible effects of aluminum in human breast cells, whether it is influencing cell pathways or the tissue microenvironment.

2880F

A flexible single-cell RNA-seq system for the accurate processing of 1,000s of single-cells. *M. Srinivasan.* WaferGen BioSystems, Fremont, CA.

   Genomics measurements on genetically identical single cells have significantly advanced our understanding of deterministic variability in gene expression and how those differences impact the formation and maintenance of cellular communities. Understanding these complex events with enough statistical rigor require the rapid sampling and processing of thousands of individual cells in a high throughput manner. To fulfill this growing need, we have engineered ICELL8, a novel single cell analysis system that can process thousands of cells from several samples for RNA-seq applications. The system consists of: ICELL8 chip - a micro-fabricated metal-alloy scaffold with 5,184 nanowells. MultiSample NanoDispenser - an automation system to dispense nanoliters of reagents into the nanowells of the ICELL8 chip Imaging Station - a microscope fitted with a fast stage to quickly image Hoechst 33324 and propidium iodide stained cells in all 5,184 wells and CellSelect™ - software used to identify single cell containing wells (or wells of researcher's choice) and dispense reagents to make cDNA ONLY to wells either automatically called by the software and/or selected by user. To showcase the ability of ICELL8 system in translational research, 1000s of mesothelioma cells were processed. RNA sequencing and analysis of the mesothelioma cells at the single cell level revealed the heterogeneity and also the genes that are expressed in common between the cells. Principal component analysis of the single cell RNA-seq data showed that mesothelioma cells formed a distinct cluster compared to other tumor types and normal cell. Overall this study significantly expands on previous genomic studies on mesothelioma and identifies several previously unknown alterations and biological pathways that are potential targets for drug discovery and treatment. Incorporating genomic analysis for detection of actionable alterations as part of mesothelioma patient care will help in developing rational individualized therapy.

2881W

Analysis of structural integrity of genomic DNA in tissue sample prep using Bulk Lateral Ultrasonic™ Energy. *V. Vivek[1], P. Sansanwal[1], D. Holley[1,2].* 1) Microsonic Systems Inc., San Jose, CA; 2) Department of Biological Sciences, San Jose State University, San Jose, CA.

   As Next-Gen sequencing throughput continues to accelerate, a serious tissue sample preparation bottleneck is emerging, calling for an automated higher throughput tissue processing technique. The solution must be quick, efficient, and should not affect DNA quality. Microsonics has pioneered the use of a novel ultrasonic technology called "Bulk Lateral Ultrasonic™ Energy." This "Microprocessor for Life Sciences" is the core technology that makes possible a multi-channel front-end tissue sample prep device that is scalable to any number of simultaneous channels, and processes (lyses) tissue samples in less than 10 minutes resulting in high DNA yield and quality. This work reveals how this new form of energy is utilized to quickly and efficiently process tissue samples, and analyzes the effect of BLU energy on the structural integrity of DNA, including: 1) the influence of ultrasound sonication on the structural integrity of pre-purified genomic DNA, 2) the results of tissue lysis with ultrasound sonication, 3) the influence of ultrasound sonication on the structural integrity of genomic DNA during lysis of animal tissue, and 4) the efficacy of BLU energy for complete elution of DNA bound to magnetic beads. Careful optimization of these parameters has led to a successful implementation of Bulk Lateral Ultrasonic energy for automated high throughput and efficient front-end tissue sample prep sufficient to meet the needs of current and emerging Next-Gen sequencing throughput demands.

2882T

Mutation signatures and intratumor heterogeneity of esophageal squamous cell carcinoma in a Chinese cohort. *Y. Wang[1], W.Q. Yuan[2], M.F. Liu[2], J. Bai[3], Q.X. Song[1], A. Abliz[2], Z. Liu[2], J.J. Li[2], C.Q. Zeng[3], H. Cai[2], Y. Ke[2], J. Li[1].* 1) Department of Human Genetics, 4909, Buhl, Medical School of Michigan University, Ann Arbor, MI; 2) Laboratory of Genetics, Key laboratory of Carcinogenesis and Translational Research, Peking University, Cancer Hospital & Institute, Beijing, China; 3) Laboratory of Disease Genomics and Individualized Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

Esophageal squamous cell carcinoma (ESCC) is one of the most common and most aggressive cancers. Epidemiological features of ESCC are extremely complex, with strong geographic differentiation among world's populations. Rural Anyang in the Henan Province of China is a well-known high-incidence area, however the causal factors in this population remain elusive. We performed exome sequencing of 81 tumor-normal pairs, identified TP53, ZNF750 and NOTCH1 as significantly mutated genes, and observed highly recurrent aberrations in several other genes previously reported for ESCC (PIK3CA, KMT2D, FAT2 and FAT1). The total set of ~7,000 SNVs revealed two main signatures: C>T transitions at NpCpG due to spontaneous deamination of 5-methyl-cytosine, and C>T and C>G mutations at TpCpW attributed to the APOBEC family of cytidine deaminases. The latter signature points to HPV infection as one of the potential mutagenic sources, even a possible etiological factor. To characterize intratumor heterogeneity we applied our newly developed method, CHAT, to estimate the clonal frequencies of copy number alternations and single nucleotide mutations in each tumor. Many tumors show a wide, sometimes multi-modal distribution of clonal frequencies, suggesting extensive within- tumor diversity due to coexistence of multiple clones. Survival analysis suggests that patients with complex multi-clonal tumor structure tend to have poorer prognosis. To better understand the patterns of growth, migration and metastatic potential among different cells within a tumor we performing exome sequencing to compare multiple regions in 10 ESCC as well as 2 esophageal neuroendocrine carcinoma tumors. In many ESCC's, each local-region sample still contains multiple clones, which are often shared with a distant region, suggesting extensive dispersal of the clonal populations and relatively slow sweep by the most dominant clone. We constructed "clone trees" to depict the most likely lineage relationship of the clones and the likely driver genes or pathways for each branch. Metastatic samples at lymph node often contain multiple clones, including those appearing in the early portions of the evolutionary tree, suggesting polyclonal seeding to the lymph nodes as well as invasion of early-stage tumor cells. Our analysis of spatial heterogeneity of molecular lesions in ESCC revealed likely temporal progression of tumorigenic events that may have driven the initiation, outgrowth, and metastasis.

2883F

Pro-Seq: A rapid and low cost method for obtaining duplex sequencing data for liquid biopsy applications. *J. Pel[1], W. Choi[1], M. Despotovic[1], L. Gelinas[1], A. Leung[1], G. Shibahara[1], A. Marziali[1,2].* 1) Boreal Genomics, Vancouver, BC, Canada; 2) Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada.

A considerable challenge to the widespread clinical adoption of sensitive sequencing-based liquid biopsy for cancer is the high assay cost compared to potential reimbursement. Assay cost is currently dominated by the large amount of DNA sequencing required to achieve coverage of a broad gene panel, and by the high read depth required for high clinical sensitivity. Attaching unique molecular barcodes to each ctDNA fragment for the purpose of error reduction further increases sequencing requirements, making liquid biopsy commercialization in many clinical applications impractical. We present a novel library construction process for NGS that increases the accuracy of the combined library construction and sequencing process by at least an order of magnitude. Named Proximity-Sequencing (Pro-Seq), the method duplicates the sequence information in each original DNA strand prior to the bulk of library construction in such a way as to provide redundant, linked templates to the sequencer that contain copies of both senses of the starting template (duplex sequencing). The redundant templates remain linked through the library construction process, allowing detection of PCR errors as sequence disagreement between multiple strands. The linked templates are amplified in a single sequencing reaction, such that base quality and incorporation information can be used to determine which bases of the sequence were corrupted during PCR amplification steps. Since multiple strands are sequenced as part of the same sequencing read, sequencing accuracy is improved without requiring use of additional redundant reads used in molecular barcoding techniques. The method is expected to have significant utility in any applications that require detection of rare sequence variants, including analysis of cell free DNA for liquid biopsy. Additionally, the workflow is entirely based on novel reagents and can be implemented without additional instrumentation beyond standard NGS equipment. We present the detailed method on an Illumina platform, and data from sequencing of bacterial and human cell free DNA that demonstrate error rates that are improved by an order of magnitude over the current state of the art NGS chemistries. Plasma-derived mutation detection is also demonstrated.

2884W

Deep learning models of cell type-specific chromatin accessibility profiles elucidate genome-wide tissue-of-origin signatures in cell-free DNA.
C. Probert, C.S. Foo, N. Sinnott-Armstrong, J. Israeli, D. Kim, C. Curtis, A. Kundaje. Stanford University, Stanford, CA, CA.

Cell-free DNA (cfDNA) has recently been shown to contain a nucleosome footprint tissue-of-origin signature; however, current methods for deconvolving tissue-of-origin from this signal use only reads mapping to specific promoter regions and measure correlation to gene expression. Here, we improve tissue-of-origin inference from cfDNA by leveraging large, publicly available DNA accessibility datasets to train a deep latent factorization model to deconvolve tissue-of-origin mixture from genome-wide cfDNA read positioning profiles. We have uniformly processed DNase accessibility data for > 300 human tissues and cell types to create a comprehensive tissue-specific atlas of chromatin accessibility. We use this dataset as the basis for a deep latent factor embedding to infer a mixture of cell types and tissues present in a given cfDNA sample. We apply read positioning profile 'V-plots' to cfDNA, and show that they can be used to learn more discriminative representations of cfDNA profiles that improve tissue signature deconvolution. We demonstrate the concordance of this model with both the tissue-of-origin composition of simulated cfDNA mixtures and also with clinical pathology assessments for both healthy and diseased individuals with high-coverage cfDNA sequencing data. We believe these models of tissue and disease signatures in cfDNA datasets are potentially informative for earlier cancer detection and diagnosis..

2885T

NanoChannel based next-generation mapping for the "dark matter" in the genome- comprehensive genome wide large structural variation detection in population. *H. Cao[1], A. Hastie[1], A.W. Pang[1], E.T. Lam[1], W. Andrews[1], T. Anantharaman[1], T. Liang[1], J. Lee[1], K. Pham[1], T. Chan[1], W. Stedman[1], M. Saghbini[1], X. Zhou[1], J. Wang[1], Z.Y. Zhu[1], M. Austin[1], M. Borodkin[1], P. Kwok[2], M. Xiao[3].* 1) BioNano Genomics Inc, San Diego, CA; 2) UCSF, San Francisco CA; 3) Drexel University.

Association of genome structural variations (SV) with complex diseases and traits has been well established; however, large SV (> 1 kb) analysis of human genomes has been limited to date by technical shortcomings of Microarrays (limited to imbalanced copy number variation (CNV) with a narrow dynamic range, low resolution, and relative readouts), next-generation sequencing (NGS) (primarily CNV, some balanced events but too short to span most repeats), karyotyping and fluorescence *in situ* hybridization (FISH) (both low resolution). Using a single-molecule-based genome analysis system, BioNano Genomics next-generation Irys® mapping (NGM) technology, it is now possible to comprehensively analyze whole genomes for SVs > 1 kilobase pairs (kbp), including balanced events such as translocations and inversions in a cost-effective and high-throughput manner. This technology allows for the *de novo* discovery or confirmation of large SVs in patient cohorts and population studies, which is needed to potentially uncover genomic structural causes of Mendelian and complex diseases such as cancer and neurobehavioral disorders. We demonstrate the robustness of NGM for genome-wide discovery of SVs in the CEPH trio from the 1000 Genomes Project with a 96% Mendelian concordance rate. We uncovered 100s of insertions, deletions, and inversions greater than 5 kbp, 7 times more than the large SV events detected by NGS, a large portion of which was novel, and some are located in the regions likely leading to disruption of gene function or regulation. We have also begun analysis of a trio of Ashkenazi Jewish descent from the NIST GIAB project, where we have found hundreds of inversions, insertions, and deletions, including large deletions in a clinically relevant UGT2B17 gene loci in the mother and son. We have also demonstrated the direct detection of translocation events in clinically validated leukemia and multiple myeloma samples as well as large inversion events frequently occurred in DUF1220-domain repeat region of NBPF gene loci related to brain size development and mental disorders. We show that NGM is a robust and effective method for structural variation detection in the human genome. Systematic whole genome structural variation within healthy and disease population cohorts is needed, in addition to the conventional SNP analysis, to study the effects of a full spectrum of genomic variations in human disease and to decipher complex traits in the era of precision medicine.

2886F

Detecting a novel subset of large genomic rearrangements in primary prostate cancer using next generation mapping. *V.M. Hayes[1,2], W. Jaratlerdsiri[1], C. Yang[3], P. Sheth[3], D.C. Petersen[1], E.K.F. Chan[1].* 1) Human Comparative & Prostate Cancer Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia; 2) Central Clinical School, University of Sydney, Camperdown, Sydney, NSW, Australia; 3) Bionano Genomics Inc., San Diego, CA, USA.

   Whole genome sequencing has recently revealed complex genomic rearrangements to be common molecular events driving prostate carcinogenesis, with clinical significance yet to be fully elucidated. Detecting the full range and subtypes of large structural variants (SVs), greater than one kilobase (kb) in length, is however challenging using clinically feasible short-read next generation sequencing (NGS) technologies alone. Unlike NGS, with single base resolution, next generation mapping (NGM) allows for the interrogation of megabase length DNA molecules outside the detection range of short-read NGS. To the best of our knowledge, this is the first study to use a nanochannel NGM technology, namely the Irys system from BioNano Genomics Inc., to generate a complete tumour-normal genome map from a clinically derived primary prostate cancer. Achieving an effective mapped coverage of 35X for a Gleason score 7 (4+3), ETS-negative prostate tumour, at an estimated 43% tumour purity, we identifying 85 large somatic deletion and insertion rearrangements. Only one-tenth of these large SVs were detectable using high-coverage short-read NGS and automated five-tooled bioinformatic analyses. Armed with our NGM-derived target regions, we performed manual inspection of corresponding NGS reads and *de novo* assembled scaffolds, allowing for a dramatic 94% of the total NGM-called SVs to be verified. Furthermore, while less than 0.5% of SNVs show oncogenic potential, over 50% of the SVs directly impact a gene or gene region. From this single-patient study, we hypothesize that a novel spectrum of large genomic rearrangements largely undetectable using standard short-read NGS may be driving primary in prostate cancer, with additional clinically relevant potential to enhance subtype classification.

2887W

Enabling low input samples for low frequency mutation detection. *J. RoseFigura[1], S. Sandhu[1], C. Schumacher[1], L. Kurihara[1], G. Durin[2], T. Harkins[1], V. Marakov[1].* 1) Swift Biosciences, Ann Arbor, MI; 2) Covaris Inc., Woburn, MA.

   A major goal in the field of next generation sequencing is to increase the number of clinical samples that can be analyzed, particularly at low input. Often, detection of low frequency somatic mutations is needed, which additionally requires higher complexity library construction, demanding deep sequencing. To determine the workflow that supports these requirements best, we compared two library preparation methods coupled to either mechanical DNA shearing or enzymatic fragmentation, by comparing the Covaris Focused-ultrasonicator coupled with the Swift Accel-NGS 2S Kit to the KAPA HyperPlus kit that includes enzymatic fragmentation. Inputs of 100, 10, and 1 ng of high quality DNA were tested with both Swift-Covaris and KAPA. There were notable differences between these two methods in both the workflow set-up and performance. For the Swift-Covaris method, no modifications in set-up were required for either DNA shearing or library prep, spanning a six-order range from 100 ng to 1 ng. This method produced consistent insert size distribution and proportional library complexity across the sample input range of 1 ng to 100 ng, without loss of efficiency at low input. The KAPA HyperPlus workflow was sensitive to DNA quality and varying inputs, requiring optimization of fragmentation time and adapter concentration for each sample change. The insert size distribution had a higher standard deviation compared to Swift/Covaris and library complexity was disproportionally reduced at 1 ng inputs. Some libraries were subjected to hybridization capture using the IDT xGEN lockdown Pan Cancer panel. When comparing sequence data using high quality genomic DNA at 1 ng, duplication rates were > 2 fold higher for KAPA compared to Swift/Covaris. At 1 ng input, the mean target coverage was also reduced approximately 2 fold for KAPA. In addition, KAPA libraries consistently displayed a broader insert size distribution than Swift/Covaris libraries. These results demonstrate that, while the Swift Accel-NGS 2S kit using Covaris DNA shearing can maintain coverage metrics at low inputs, KAPA HyperPlus can not. The benefits of both higher quality and the more robust workflow of the Swift/Covaris process enable successful sequencing of low input clinical samples.

2888T

Enrichment based NGS is superior to amplicon based protocols for hereditary cancer gene analyses in somatic tumor samples of patients at risk for tumor predisposition syndromes. *E. Schrock[1,2], F. Zakrzewski[2], K. Hackmann[1], K. Grueztmann[2], L. Gieldon[1], J. Porrmann[1], S. Zeugner[3], A. Kruger[2], K. Kast[4], P. Wimberger[4], D. Aust[2,3], G. Baretton[2,3], A. Rump[1], B. Klink[1,2].* 1) Institut fuer Klinische Genetik, Medizinische Fakultaet Carl Gustav Carus, Technische Universitaet Dresden, Fetscherstraße 74, 01307 Dresden, Germany; 2) Deutsches Konsortium für Translationale Krebsforschung (DKTK), Dresden; Deutsches Krebsforschungszentrum (DKFZ), Heidelberg; Nationales Centrum für Tumorerkrankungen (NCT), Dresden, Germany; 3) Institut für Pathologie, Universitaetsklinikum Carl Gustav Carus, Technische Universitaet Dresden, 01307 Dresden, Fetscherstraße 74, Germany; 4) Klinik und Poliklinik für Gynaekologie und Geburtshilfe, Universitaetsklinikum Carl Gustav Carus, Technische Universitaet Dresden, Fetscherstr. 74, 01307 Dresden, Germany.

Introduction: Pedigree analysis of families at risk for tumor predisposition syndromes often reveals that the index patient is not available for genetic diagnostics. In a substantial number of cases, however, DNA of formalin-fixed paraffin-embedded (FFPE) tumor blocks could be provided from pathology. Since the analysis performed in the index patient is of particular importance, we compared the overall performance between a novel enrichment-based NGS protocol and amplicon-based NGS protocols with regards to the detection of mutations in 94 hereditary cancer genes, including BRCA1, BRCA2, and TP53, in numerous FFPE tumor samples. **Method:** The detection of mutations (SNVs, InDels, and CNVs) was done using fragmented DNA of poor to very poor quality from more than 50 routine FFPE tumor samples. The Illumina TruSeq Nano DNA Library Prep Kit was used in combination with the TruSight Cancer 94 gene panel and compared to the amplicon-based TruSight Tumor 26 and BRCA1/2 gene panels. The overall performance and the detection rate of SNV- and CNV-mutations were compared between enrichment- and amplicon-based NGS-strategies. **Results:** The enrichment-based panel generated results of much higher quality and higher reproducibility. In particular, distribution of sequencing reads along target regions is very homogeneous and comparable within a sample and between different samples. The detection of CNVs, which is mandatory for analyzing hereditary cancer genes with suspected deleterious mutations, could only be performed robustly using the enrichment-based strategy, since amplicon-based protocols show huge coverage changes within and between samples and drop-out of amplicons. Furthermore, amplicon-based methods caused PCR- artifacts. Consequently, the enrichment-based panel was applied to FFPE tumor samples in families with suspicion of hereditary cancer syndromes were no living index patient was available. We were able to detect the underlying pathogenic mutations, for example in the CDH1 gene in a family with hereditary diffuse gastric cancer. We also could confirm germline mosaic mutations in TP53 identified in peripheral blood to be present in the majority of cells in corresponding FFPE tumor tissues. **Conclusions:** Using our enrichment-based strategy on fragmented DNA from FFPE samples leads to an increased NGS performance and to robust mutation detection even in large tumor suppressor genes (e.g. BRCA1/2) and is therefore recommended rather than amplicon-based strategies.

2889F

Resolving the functional significance of *BRCA1* RING missense substitutions. *A.M. Paquette[1], K. Tao[1], A.W. Stark[1], R. Bell[1], A. Thomas[2], W.D. Foulkes[3,4], D.E. Goldgar[5], S.V. Tavtigian[1].* 1) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT; 2) Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT; 3) Program in Cancer Genetics, Departments of Oncology and Human Genetics, Montreal, Quebec, Canada; 4) Department of Medical Genetics, Jewish General Hospital and McGill University Health Centre, McGill University, Montreal, Quebec, Canada; 5) Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT.

Interpreting variants of uncertain significance (VUS) is a central challenge to clinical geneticists. Two well documented functions of the *BRCA1* (MIM 113705) RING domain are its interaction with *BARD1* (MIM 601593) and subsequent E3 ubiquitin ligase activity of the heterodimer. Although several substrates have been identified as *BRCA1* ubiquitination targets, mice homozygous for the *BRCA1* p.I26A RING mutation are no more tumor prone than their wild type littermates. This missense substitution abrogates the heterodimer's E3 ubiquitin ligase activity without affecting heterodimer formation. In contrast, *BRCA1* variants that disrupt heterodimer formation result in DNA repair defects and loss of tumor suppression. Interestingly, a recent report provided evidence that measurement of *BRCA1*: *BARD1* interaction by itself poorly predicts the pathogenicity of RING missense substitutions. Here we combine these recently published high-throughput functional assay data with data from 68,000 full sequence tests of *BRCA1* to estimate genetic risk attributed to these two functional classes of missense substitutions. Our analyses find minimal evidence of pathogenicity for missense substitutions that are damaging to E3 ligase activity (from phage display assay data), but not the *BARD1* interaction (from yeast 2-hybrid assay data). In contrast, RING variants that disrupt heterodimer formation had strong evidence of pathogenicity. While these results are consistent with the functional activity of the only two known neutral missense substitutions in the *BRCA1* RING domain, multiple RING variants that have already been classified as pathogenic appear to have wild type *BARD1*-binding scores in the yeast 2-hybrid assay data. Mindful of the limitations of yeast two-hybrid screens, we moved this assay into mammalian cells, and found that a system that is more sensitive to the abundance and stability of each protein eliminates this discord. To facilitate clinical classification via the Bayesian Integrated Evaluation, we calibrated this assay by regressing % WT activity for all of the known pathogenic and neutral *BRCA1* RING missense substitutions (plus several cross-species likely neutral substitutions) against their observational data odds of pathogenicity. Moving forward, we envision a high-throughput assay to accelerate the clinical classification of all possible *BRCA1* RING missense substitutions that are the result of a single-nucleotide substitution.

2890W

Using Irys next-generation mapping to identify medically-relevant genomic variation. *A.W.C. Pang, J. Lee, S. Chan, W. Andrews, T. Anantharaman, A. Hastie, E. Holmlin, H. Cao.* Research and Development, BioNano Genomics, San Diego, CA.

Structural variation (SV) detection is fundamental to understand the genetic basis of diseases. While karyotyping, cytogenetic, and conventional molecular detection approaches are robust, they can be manually intensive, biased towards targeted loci, and cannot elucidate the copy number of long repeats. The BioNano Genomics Irys® System offers an unbiased and sensitive method to detect large variants. Based on next-generation mapping (NGM), it uses NanoChannel arrays to linearize DNA hundreds of kilobases in size, high resolution imaging, and haplotype-aware *de novo* assembly to construct genome maps megabases in length. To complement the *de novo* assembly method, the long molecules can be aligned to the *in silico* digested human reference to identify aneuploidy and copy number alterations by depth counts. The combined methods can capture over 4,000 insertions, deletions (>one kilobases), translocations and inversions in a human sample, about 2.5 times more than sequencing alone. We present a novel software that examines genome map calls to identify disease-associated SVs. To search for high-confidence calls, it searches for genomic elements such as centromeres and segmental duplications that may confound SV-calling, and evaluates quality scores of the assembly, alignments, and SVs detected. To further ascertain medically-relevant candidates, it filters the remaining calls with over 36,400 SVs collected from 32 mapping experiments from phenotypically "normal" individuals, and compares with calls from disease-specific databases. Using this pipeline, whose total runtime is only a few hours, we can efficiently focus on several dozens of significant candidates for further analysis. We ran the Irys System on 10 samples with hematological malignancy, applied the candidate-finding software, and identified known rearrangements such as the t(9;11)(q21.3;q23.3) in acute myeloid leukemia AML and t(9;22)(q34.12;q11.23) in CML. Moreover, we uncovered novel mutations ranging in size from a small 4.2 kb insertion disrupting an ALL gene *CNOT3* to a large 9.6 Mb deletion at 17q25. In conclusion, Irys technology may replace conventional detection approaches with one standardized and comprehensive platform, aid the discovery of functional SV, and improve our understanding of the mechanisms of diseases.

2891T

From FFPE tissue to targeted multiplex panel sequencing in less than 60 minutes using Hyb & Seq™ chemistry. *D. Kim, M. Walsh, E. Piazza, R. Khafizov, E. Manrao, J. Beechem.* NanoString Technologies, Inc., Seattle, WA.

Hyb & Seq is a single molecule sequencing method that uses cyclic nucleic acid hybridization of fluorescent molecular barcodes to sequence native single molecule targets. The ability to directly sequence native molecules affords the opportunity for a simple, very rapid and automatable sample preparation workflow for targeted sequencing, even with challenging sample types such as formalin-fixed paraffin embedded (FFPE) tissue. Here, we demonstrate proof-of-principle Hyb & Seq sample preparation by targeting 100 different gene targets, and quantifying the efficiency of the sample preparation on various FFPE tissue types (lung, colon, skin and tonsil). In the Hyb & Seq sample preparation, FFPE tissue is first 'liquefied' (one-pot deparaffinization and tissue lysis) using an aqueous reagent mixture containing lysis buffer with Proteinase K and with heat. Following incubation for 30 minutes at 56°C and for 5 minutes at 98°C, the liquefied FFPE tissue slurry is passed through a syringe filter to separate and collect the supernatant, which contains the genomic DNA released from the tissue. Fragmented genomic DNA is captured directly in a 15 minute hybridization reaction using multiplexed target specific capture probe panel to both enrich for gene targets of interest and to create 'gapped' single-stranded regions in the captured native DNA molecules. These 'gapped' single stranded gene targets are purified, captured onto a flow-cell, and then quantified using Hyb & Seq* to identify and digitally count, yielding both the unique sequence and qualitative copy number of each target gene. Experimental results showed that more than 95% of the 100 targets were successfully captured onto the flow-cell using just one to three FFPE slices of 10 microns each, with no PCR amplification (success criteria of ≥1,500 copies captured per target). The total counts from each target were within one log range, demonstrating good capture uniformity. The total processing time from FFPE curls to start of sequencing was under 60 minutes, with the total hands-on time of less than 15 minutes. Based on these results, Hyb & Seq's simple workflow demonstrates the possibility of a rapid sample-to-answer sequencing instrument well suited for clinical applications. *Detailed description of Hyb & Seq chemistry is available on-line: https://www.youtube.com/watch?v=LUc5Yf9Dlgg.

2892F

miRNA as liquid biopsy biomarkers in breast cancer. *B.M. Dugan, J.M. Shaffer, E. Lader.* Qiagen Sciences, Frederick, MD.

miRNA are small, 21 nucleotide RNA that play in important role in normal and abnormal gene regulation. Changes in miRNA expression have been observed in an assortment of cancers including breast cancer. Recently, increased interest in noninvasive surveillance and detection of biomarkers associated with cancer are driving the discovery of miRNA expression profiles that can be used to detect disease without the need for tissue biopsy. To demonstrate the effectiveness of miRNA as a potential early biomarker for breast cancer, small RNA-sequencing was conducted on matched breast cancer samples, tissue biopsy from a breast malignancy as well as serum/plasma. By comparing non-invasive techniques to standard tissue biopsies will determine if specific miRNA profiles are reliable biomarkers for early detection of breast cancer.

2893W

Differential expression of miRNA in 2D and 3D human colon cancer cell cultures after therapeutic compound treatment. *B. Hook, S. Lewis, T. Schagat.* Scientific Applications Support, Promega Corporation, Fitchburg, WI.

MicroRNAs (miRNAs) are small-noncoding RNA molecules that regulate gene expression post-transcriptionally. Amplification and overexpression of individual oncomiRs (miRNAs associated with cancer) or genetic loss of tumor suppressor miRNAs are associated with human cancer and are sufficient to drive tumorigenesis in mouse models. A global decrease in miRNA levels has been observed in human cancers and linked to genetic and epigenetic alterations in components of the miRNA biogenesis machinery. Fluorouracil is an anti-cancer chemotherapy drug that acts as an antimetabolite in cells and has been previously shown to modify the expression profiles of miRNAs in human cancer cells. In this study, we developed a workflow to enable testing of viability, cytotoxicity and apoptosis in the same cells in multiple culture formats, with downstream analysis of miRNA effects. Two colon cancer cell lines (HCT116 and HT-29) were grown as monolayers in standard 2D culture plates, as Matrigel-embedded 3D cultures, and as large 3D spheroids generated using ultra-low attachment plates followed by treatment with 5-Fluorouracil. Viability and cytotoxicity were monitored in real-time for 48 hours followed by measurement of apoptosis and subsequent extraction and quantification of miRNA expression changes.   We determined that 5-Fluoruracil had differential potency between 2D and 3D culture models for both of the two colon cancer cells lines tested. From the same assay wells that were used to determine potency, total RNA including miRNAs was extracted and three miRNAs were analyzed for expression using RT-qPCR. The impact on miRNA expression was dependent on cell culture format and was lower in the 3D culture models. This complete workflow solution enables assessment of cell health and gene expression from the same samples.

2894T

Multiplex detection of driver mutations in lung cancer: Simultaneous assay of single nucleotide variants (SNV) and fusion-transcripts from small amounts of FFPE samples on the nCounter® analysis system. *G. Meredith[1], J. Kargl[2], J. McKay-Fleisch[1], M. Ross[1], J. Garber[1], G. Ong[1], C. Ngouenet[1], J. Phan[1], M. Krouse[1], M. Pansalawatta[1], L. Dennis[1], A. Kharkia[1], E. Piazza[1], R. Boykin[1], A.M. Houghton[2], D. Kim[1], J. Beechem[1].* 1) R&D, NanoString Technologies, Seattle, WA; 2) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

Worldwide, lung cancer is the most diagnosed form of cancer with a survival rate that is among the lowest. Combined, mutations (in the form of SNVs and InDels) and gene fusions account for the majority of clinically interpretable and actionable genomic alterations seen in the disease's most common form, non-small cell lung cancer (NSCLC). Thus, tools and methods for rapid and robust multiplexed detection of these genomic alterations from formalin-fixed, paraffin-embedded (FFPE) tissue samples are critically needed. NanoString's optical barcoding technology now enables detection of small genomic sequence changes such as SNVs and InDels. A test-panel that targets dozens of clinically relevant variants has been developed that permits multiplex detection of SNVs and InDels. Detailed technical performance of this test panel is described in a companion presentation. Importantly, this panel can be used in a 3D Biology™ workflow when combined with the nCounter Vantage™ Lung Gene Fusion panel to permit simultaneous detection of fusion transcripts from RNA. The SNV assay workflow begins with pre-amplification of 5 ng of FFPE-extracted gDNA via multiplex PCR. Resulting amplicons are then hybridized to nCounter probes. The fusion assay uses 150 ng of total FFPE-extracted RNA which is directly probed in a separate hybridization. These low input requirements support minimal use of precious samples that may obviate the need for microdissection. After hybridization, the two reactions are pooled for quantification within a single cartridge lane on an nCounter system. Each panel has specific controls that ensure QC checks and proper statistical data analysis. Here, we first demonstrate simultaneous assay of both classes of driver mutation using cell line and commercial control samples to verify the analytical performance. Next a cohort of 50 patients was assayed for the presence of SNVs and/or gene fusions and the results compared to matched whole-exome sequencing data. SNV detection via the nCounter assay was highly concordant (>95% cross-validation) with sequencing results. Further, SNVs were accurately detected from FFPE samples with as little as 5-10% tumor cellularity. Likewise, positive gene fusion calls were validated by follow-up clinical assays. Significantly, we show that SNVs and fusion transcripts that are clinically relevant in lung cancer can be assayed together using a single 10 micron FFPE slice.

2895F

Multiplex profiling of cancer driver mutations: Detection of single nucleotide variants (SNV) and small InDels from small amounts of FFPE samples on the nCounter® analysis system. *M. Ross[1], J. Lee[2], J. McKay-Fleisch[1], C.P. Vellano[2], J. Garber[1], A.K. Eterovic[2], E. Manrao[1], M. Krouse[1], J. Phan[1], M. Pansalawatta[1], E. Piazza[1], L. Dennis[1], A. Kharkia[1], D. Kim[1], G.B. Mills[2], G. Meredith[1], J. Beechem[1].* 1) R&D, NanoString Technologies, Seattle, WA; 2) Dept. of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Tools and methods to perform rapid and robust multiplexed molecular profiling of formalin-fixed, paraffin-embedded (FFPE) tissue samples are critical for basic and translational cancer research. NanoString's optical barcoding technology has been adapted to enable detection of small genomic sequence changes such as SNVs, multiple base substitutions, and insertions and/or deletions of less than 20 bases through development of a novel class of variant-specific hybridization probes*. A test panel of variant-specific probes that targets 26 clinically relevant variants has been developed that permits multiplex detection of SNVs from gDNA. This assay can be combined with other NanoString Vantage™ panels for 3D Biology™ to enable, for example, simultaneous detection of gene expression from RNA. The workflow begins with a pre-amplification of gDNA (fresh or FFPE-extracted) via multiplex PCR to enrich targets of interest. Resulting amplicons are then hybridized with the SNV panel of variant-specific nCounter probes. After hybridization, pooling with any other 3D Biology panel can be done for simultaneous quantification within a single cartridge lane on an nCounter system. SNV panel specific controls have been implemented for both thorough QC measures and proper statistical analysis. All variant-specific hybridization probes in the panel were verified to have accurate detection down to 5% variant allele fraction, with many able to go below 1%. As little as 5 ng of FFPE-extracted gDNA is sufficient for the assay, supporting minimal use of precious samples, such as patient-derived FFPE tissues and obviating the need for microdissection. Using 5 ng, simultaneous detection of 15 variants, each present at either 2.4% or 4.8% allele fraction, was demonstrated using a blend of genomic DNA from engineered cell-lines. Most importantly, SNV mutations were detectable from a range of patient-derived FFPE samples that were known to harbor mutations (as low as 5% tumor) or were subsequently verified as such by sequencing. Significantly, we show that the SNV detection chemistry can be combined with gene expression analysis by using an nCounter Vantage RNA Panel in a 3D Biology workflow to provide simultaneous assay of these two important classes of analytes, DNA and RNA, from limited amounts of nucleic acid. *Kim, D. et al. 3D Biology™: Simultaneous single-molecule quantification of DNA (SNVs), mRNA, & proteins. Poster presented at AGBT; February 2016; Orlando, FL.

2896W

Comprehensive identification of piwi-interacting RNAs binding sites associated with hypoxia-inducible genes downregulation. *B.C. Minatel, N. Firmino, E.A. Marshall, K.L. Bennewith, W.L. Lam, V.D. Martinez.* Department of Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada.

   Background: The expression of hypoxia-induced genes is tightly regulated and involves a variety of regulatory mechanisms. Piwi-interacting RNAs (piRNAs) are small (24-32 bp) RNA species able to recognize complementary DNA sequences throughout the genome, inducing epigenetic silencing nearby to target regions. Since the expression of specific piRNAs is responsive to changes in oxygen tension, we investigated if piRNAs play a role in the regulation of hypoxia-induced genes, by predicting their binding pattern to regulatory DNA regions (upstream transcript start site, TSS) of genes downregulated under hypoxic conditions. **Methods:** We determine gene expression profiles, as well as piRNA expression profiles from small-RNA sequencing libraries, derived from hypoxic (n=932) and normoxic (n=894) tumors from 7 human organs, and for normoxic/hypoxic (16h at 1% O_2) cell lines from the same sites. For prediction analyses, we considered a 3.5 kb DNA sequence window upstream of the TSS of previously described hypoxia downregulated genes (n=51). piRNA/DNA binding prediction was performed using two different algorithms (miRanda and ThermoBLAST), with specific parameters defined by piRNA features. **Results:** Our analysis identified that a total of 36 piRNAs were consistently upregulated under hypoxia in tumor samples and cell lines from different organs. Twenty seven (75%) of these piRNAs are predicted to bind the upstream region of genes known to be downregulated by hypoxia. We further determined the concordance of piRNA upregulation with corresponding mRNA downregulation. To identify which of these 27 piRNAs are predominantly interacting with hypoxia related genes, we compared their binding preference between the 3.5 kb window upstream TSS of hypoxia downregulated genes and those of all human protein-coding genes. Five piRNAs were found to have binding sites significantly enriched in upstream regions of hypoxia-downregulated genes. These piRNAs had predicted upstream binding sites in 0.9-1.6% of the hypoxia-repressed genes compared to only 0.1-0.3% of all genes in the human transcriptome (p value $X^2 <= 0.005$). **Conclusion:** Our results indicate that a set of piRNAs may have a role in the regulation of hypoxia induced gene expression. Although further mechanistic validation is needed, the enriched presence of piRNA binding sites in regulatory regions of hypoxia-regulated genes suggests the involvement of piRNA mediated epigenetics mechanisms in hypoxia-induced gene silencing.

2897T

Impact of the cell cycle on mRNA-Seq gene expression profiling for single cancer cells using BD™ Precise assays. *C. Lomas[1], K. Dembski[1], J. Goglio[1], K. Yaccato[2], C. Betts[2], S. Weaver[2], X. Wang[1].* 1) BD Life Sciences, San Jose, CA; 2) BD Life Sciences, Menlo Park, CA.

   Single cell mRNA sequencing has been developed as a powerful tool to examine the heterogeneity of a cell population. However, confounding factors, such as the cell cycle state of the cell, might obscure the observation at the single cell level. In a biological sample, each cell is at an unknown phase of the cell cycle. Cell cycle-dependent gene expression can bias assessment of differential gene expression or even drive cluster analysis when trying to cluster cells by gene expression patterns. In this study, we analyzed the influence of the cell cycle states on gene expression profiles by sequencing mRNAs from single cells sorted by BD's fluorescence-activated cell sorting (BD FACS™ technology). Three different methods of cell preparation were used to study the relationship between known cancer-related genes and cell cycle genes using the T47D breast cancer cell line. For the first method, cells were treated with cell cycle inhibitors to arrest cells in G0/G1, S, and G2/M phases of the cell cycle before sorting. For the second method, cells in culture were sorted based on G0/S1, S, G2/M phases, without cell cycle inhibitors. Finally, a third method was used to sort the cells irrespective of the cell cycle phase and track the cell cycle phase information by using a BD FACS index sorting feature. The cell cycle phase was determined by intracellular staining using Hoechst 34580, a DNA dye that is excited by a violet laser, to estimate the DNA content. 96 cells prepared by each method were sorted into individual wells of a BD™ Precise encoding plate using a BD FACSseq™ cell sorter and assayed using a gene panel containing common cell cycle- and cancer-specific markers. Principle component analysis of mRNA expression of the cell cycle genes was used to cluster the cells, and cancer gene expression profiles were determined. The cell cycle phase of each cell deduced from the targeted gene panel gene expression clustering matched the index sorting data generated by the BD FACSseq cell sorter. In this study, we present a BD Precise assay to study BD FACSseq sorted single cells at different phases of the cell cycle, which allowed us a better understanding of the effects of cell cycle phases on gene expression profiles at the single cell level. The panel and algorithm generated could be applied in cancer tissue samples and other biological tissue sample types to separate the potential confounding effect of cell cycle on single cell gene expression studies.

2898F

Massively parallel functional analysis of missense mutations in *BRCA1*.
L. Starita[1], M. Islam[2], S. Fields[1,3], J. Parvin[2], J. Shendure[1,3]. 1) Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics and The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 3) Howard Hughes Medical Institute.

Interpreting the effects of genetic variation is the central challenge for the progress and prospects of precision medicine. Variant-effect prediction algorithms can classify any possible genetic variation but despite substantial efforts, they do not have the accuracy to support their use in a clinical setting. On the other hand, experimentally testing the effect of each individual variant on protein function or gene expression yields high-quality information, but is low-throughput. To resolve this, we are developing massively parallel functional assays to determine the effects of thousands of genetic variants on protein function and/or gene expression in a single experiment. Using such an assays, we comprehensively evaluated the effects of >1,300 amino acid substitutions on the biochemical functions of the RING domain of BRCA1 as well as the effects of all possible nucleotide substitutions in exon 18 of *BRCA1* on mRNA splicing. We have now successfully parallelized a cellular assay to test the effects of large numbers of missense variants on the tumor suppressive functions of the full-length BRCA1 protein, the best proxy for its *in vivo* function in relation to breast and ovarian cancer risk. The sum of our results-to-date show that predictions based on massively parallel experimental analysis markedly outperform commonly used computational tools in predicting the effects of missense variants on BRCA1 function. We envision that for *BRCA1* and other clinically relevant genes, massively parallel functional assays will facilitate the prospective interpretation of variants observed in clinical sequencing.

2899W

Guiding platinum-based chemotherapy with a somatic mutation signature of *BRCA1/2* **impairment.** *E.Y. Zhao[1], Y. Shen[1], E. Pleasance[1], K. Kasaian[1], M. Jones[1], C. Ch'ng[1], C. Reisle[1], P. Eirew[2], K.L. Mungall[1], N. Thiessen[1], Y. Ma[1], A. Fok[1], A.J. Mungall[1], Y. Zhao[1], R.A. Moore[1], D. Villa[4], T. Shenkier[4], C. Lohrisch[4], S. Chia[4], S. Yip[3], K. Gelmon[4], H. Lim[4], S. Sun[4], K.A. Schrader[2,6], S. Young[5], A. Karsan[1,5], R. Roscoe[1], J. Laskin[4], M.A. Marra[1,6], S.J.M. Jones[1,6,7].* 1) Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada; 2) Department of Molecular Oncology, British Columbia Cancer Agency, Vancouver, BC, Canada; 3) Department of Pathology & Laboratory Medicine, The University of British Columbia; 4) Department of Medical Oncology, British Columbia Cancer Agency, Vancouver, BC, Canada; 5) Cancer Genetics Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada; 6) Department of Medical Genetics, University of British Columbia, Vancouver BC, Canada; 7) Department of Molecular Biology and Biochemistry, Simon Fraser University, Vancouver BC, Canada.

Background: *BRCA1* and *BRCA2* are centrally involved in homologous recombination. Mutations in these genes account for 5-10% of breast cancers and have been associated with sensitivity to platinum-based chemotherapies. Whole genome sequencing has revealed somatic mutation signatures which are molecular phenotypes of genomic instability. A signature associated with *BRCA1* or *BRCA2* deficiency (BRCAd) has been found to predict cisplatin response in pancreatic cancer xenografts. Here, we investigate whether this signature can direct platinum-based therapy in breast cancer patients and quantify the associated clinical benefit. **Methods**: 47 advanced breast cancers underwent whole genome and transcriptome sequencing on an Illumina HiSeq sequencer. Comprehensive genomic characterization involved alignment, *de novo* assembly, and comparison to matched blood sequence to characterize somatic mutations and gene expression. Mutation signatures were deciphered using non-negative matrix factorization. A retrospective chart review was performed to obtain the timing and outcomes of cisplatin and carboplatin regimens. Of 29 platinum treated patients, 24 underwent radiological investigation. 11 were treated exclusively before sequencing, 14 after, one both before and after, and 3 during. **Results**: Six mutation signatures were deciphered, of which four have been previously observed in breast cancer. The BRCAd signature accounted for a range of 12 to 10,246 somatic mutations per cancer. Logistic regression demonstrated 59% increased odds (odds ratio 1.16-2.50) of radiological clinical improvement (CI) on platinum-based therapy for every 1000 mutations attributed to the BRCAd signature (p = 0.01). All patients with CI fell within the top BRCAd quartile. One patient with CI and high BRCAd signature possessed no detectable abnormalities in *BRCA1* or *BRCA2*. Another carried a *BRCA1* missense variant of unknown significance (H1390R). Patients with above-median BRCAd signature also experienced a 69 day extension of time to treatment failure on platinum-based therapy compared to patients with below-median BRCAd signature (p = 0.005). Cox analysis demonstrated 18% decreased daily odds of treatment failure per 1000 BRCAd-attributed mutations (hazard ratio 0.71-0.95, p = 0.007). **Conclusion**: A BRCAd-associated mutation signature predicts clinical improvement on platinum-based therapy in breast cancer and extends time to treatment failure, even in the absence of detected *BRCA1*/*BRCA2* mutations.

2900T

Association of germline variants in the *APOBEC3* region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *A.R. Banday[1], C.D. Middlebrooks[1], K. Matsuda[2], K.I. Udquim[1], O.O. Onabajo[1], A. Paquin[1], J.D. Figueroa[3], B. Zhu[4], S. Koutros[4], M. Kubo[5], T. Shuin[6], N.D. Freedman[4], M. Kogevinas[7,8,9,10], N. Malats[11], S.J. Chanock[4], M. Garcia-Closas[4], D.T. Silverman[4], N. Rothman[4], L. Prokunina-Olsson[1].* 1) Laboratory of Translational Genomics - National Ca, NATIONAL INSTITUTES OF HEALTH, BETHESDA, MD; 2) 2Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 3) 3Usher Institute of Population Health Sciences and Informatics, Medical School, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, USA; 5) Center for Integrative Medical Science, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan; 6) Department of Urology, School of Medicine, Kochi University, Koichi, Japan; 7) CIBERESP, CIBER Epidemiologia y Salud Publica, Madrid, Spain; 8) Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; 9) Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain; 10) National School of Public Health, Athens, Greece; 11) Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

Somatic mutations of a specific type (<u>C</u> to T or G substitutions in the T<u>C</u>A or T<u>C</u>T motifs) have been described in many tumors as APOBEC-signature mutations. This mutation pattern has been correlated with expression of two members of the *APOBEC3* subfamily - *APOBEC3B* and *APOBEC3A*. However, factors affecting APOBEC mutagenesis are largely unknown. Here, we explored the contribution to cancer risk and APOBEC mutagenesis of the two common germline variants in the *APOBEC3* region - a single nucleotide polymorphism (SNP), rs1014971, and a 30 Kb deletion that eliminates *APOBEC3B* and creates *APOBEC3AB* chimera. Imputation and association analyses were performed in 3,732 bladder cancer cases and 10,721 controls. Expression analysis of all genes and their isoforms in 1 Mb region centered at rs1014971 was performed in the Cancer Genome Atlas (TCGA) bladder cancer dataset (N=412) and breast cancer dataset (N=520). Expression was analyzed by linear regression in relation to rs1014971 genotypes adjusting for age, sex, race, DNA CpG methylation status and copy number variation (CNV). APOBEC-signature mutation analysis was performed with genotypes of rs1014971. Functional annotation of rs1014971 and two correlated SNPs ($r^2 \geq 0.8$) was performed *in silico* using ENCODE data, and experimentally, using electrophoretic mobility shift assays (EMSA). Three bladder and three breast cancer cell lines were treated with bleomycin, a DNA-damaging drug or Sendai virus, that induces robust antiviral interferon response, and mRNA expression of *APOBEC3s* was evaluated by qRT-PCR. We found that SNP rs1014971 was associated with bladder cancer risk, increased *APOBEC3B* expression, and enrichment with APOBEC-signature mutations in bladder tumors. In contrast, the deletion was not important in bladder cancer, while being associated with breast cancer risk and enrichment with APOBEC-signature mutations in breast tumors. Experimental analysis showed allele-specific DNA-protein interactions for rs1014971, which is located within a putative enhancer region of *APOBEC3B*. *In vitro, APOBEC3B* was predominantly induced by treatment with a DNA-damaging drug in bladder cancer cell lines and *APOBEC3A* was induced as part of antiviral interferon-stimulated response in breast cancer cell lines. These findings demonstrate contribution of germline variants to APOBEC mutagenesis and suggest tissue-specific role of environmental oncogenic triggers, particularly in individuals with germline *APOBEC3* risk variants.

2901F

Next generation sequencing analysis of *CDH1* gene in Mexican patients with diffuse gastric cancer. *A.R. Bustos-Carpinteyro[1,2], M.G Flores-Miramontes[3,4], V. Ramírez-Rodríguez[5], S.A. Romero-Martínez[5], A. Aguilar-Lemarroy[4], L.F. Jave-Suarez[4], J.Y. Sánchez- López[2].* 1) Instituto de Genética Humana, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) División de Genética, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 3) Doctorado en Ciencias Biomédicas. Universidad de Guadalajara, Guadalajara, Jalisco, México; 4) División de Inmunología, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 5) División Diagnóstica, Roche México, México, Distrito Federal.

Introduction. Gastric cancer (GC) constitutes the third cause of cancer death worldwide; histologically, GC can be classified as diffuse and intestinal. Diffuse GC has a reduced expression of the E-Cadherin cell adhesion protein, which is encoded by the *CDH1* gene; mutations in this gene are one of the main causes for gene inactivation and has been considered as an important factor in cancer progression and metastasis. To date, have been reported a total of 155 *CDH1* germline mutations, of which 126 are pathogenic and 29 are unclassified variants. **Materials and Methods.** We studied 22 DNA tumoral tissues of Mexican patients with diffuse GC. Mutation analysis of the promoter region and exome of the *CDH1* gene was performed by NGS throw the GS Junior 454 platform by Roche® and the GS Amplicon Variant Analyzer software (Roche®). **Results.** We found a total of44 mutations of which 18 are point mutations, 18 deletions and 8 insertions; these mutations were located mainly in the promoter region and introns (47.7% and 36.4%, respectively), whereas only 13.6% were in exons and the remaining (2.3%) was in the 3'UTR. 11 of the 44 variants have been previously reported as SNPs or mutations. The most important variants found were: two polymorphisms in the promoter region (rs5030625 and rs16260), which have been previously reported and associated with gastric cancer. Two intronic variants, IVS1 48+3delA and IVS4 531+10 G>C, that could have pathologic potential due to their localization. Three variants in codifying region: c.202delT (rs786202151) reported as pathogenic for be associated with hereditary cancer-predisposing syndrome; c.360delG (COSM1684458); and c.367insC (not previously reported); these variants alter the reading frame leading to a truncated polypeptide (at 82, 142 and 136 residues, respectively), affecting the precursor protein. **Conclusions.** Most of the mutations identified in this study are located in the gene promoter region; these variants could be relevant for the expression regulation of *CDH1* (mainly down-regulated in GC); for this reason, a complementary study of gene expression (mRNA and protein) would be appropriate. Additionally, further bioinformatics analyses are needed in order to understand the biological meaning of the variants found, and to identify those with potential pathogenicity.

2902W

The genomic landscape of never smoker lung adenocarcinomas. *T. de Silva[1], V. Martinez[1], K. Thu[1], S. Lam[2], W. Lam[2], H. Varmus[3], W. Lockwood[1].* 1) Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, Canada; 2) National Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) NISC Sequencing Program, Sandra and Edward Meyer Cancer Center, Weil Cornell Medical College, New York, NY.

   Although the majority of lung cancers can be attributed to tobacco exposure, approximately 15 to 25% arise in individuals who report smoking fewer than 100 cigarettes during their lifetimes ("never smokers"). Lung cancer may develop in never smokers through mechanisms distinct from those that affect smokers and is strongly associated with female gender, East Asian ethnicity, adenocarcinoma histology, and a significantly higher frequency of EGFR mutations and ALK translocations--- findings that are useful biomarkers for administration of targeted drugs. Our studies build on two important observations. First, a significant fraction of lung adenocarcinomas in never smokers have not been found to have mutations in known oncogenic driver genes. Second, mutations in oncogenic driver genes appear to be insufficient for tumorigenesis, suggesting that additional alterations are required. To address these issues, we have comprehensively studied the genomes of 15 lung adenocarcinoma from never smokers - seven "triple negative" tumors (with normal EGFR, ALK, and KRAS genes) and eight EGFR mutant tumors - with the goal of identifying novel mutant genes in these subsets. Whole exome sequencing revealed a median of 46 coding mutations per tumor with a median mutation density of 2.69 mutations/Mb. To identify mutated genes that confer a selective growth advantage we used a multistep approach, filtering variants based on gene expression level, background mutation rate and gene size. 32 unique genes demonstrated significant evidence of conferring a selective advantage. These included known oncogenes (EGFR, ERBB2 and MET) and tumor suppressor genes (p53, RB1 and ATM). In addition, sequencing of tumor RNA revealed fusions involving RET or ROS1, one fusion in each of two tumors. The variations in MET consisted of truncating and splice-site mutations and have not been previously characterized in this context. We also observed mutation, hemizygous and homozygous loss of multiple genes from the chromosome arm 6q. Pathway analysis indicated frequent mutation in genes implicated in PI3 kinase signaling, RNA splicing and histone modification. Expanded genomic analyses of captured exons from over 180 genes from the initial 15 tumors and an extended panel of 85 tumor/normal pairs have revealed a shared signature, which indicates prevalence in such lung adenocarcinomas. Together, this work will expand our understanding of lung adenocarcinomas arising in never smokers. .

2903T

Identification of functional candidate variants at two ovarian cancer outcome loci. *D. Glubb, S.E. Jonatty, M. Quinn, S. MacGregor, G. Chenevix-Trench.* QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

   Recently, we identified associations between five SNPs and either progression free or overall survival ($p < 1 \times 10^{-5}$) in a genome-wide association analysis of over 4,000 European epithelial ovarian cancer patients. All five SNPs and other candidate outcome variants correlated by linkage disequilibrium ($r^2 > 0.4$) are located in non-coding regions. To identify genes targeted by these SNPs, we have analysed candidate outcome variants using bioinformatic, chromatin conformation capture (3C) and reporter gene approaches.   eQTL data from the Genotype-Tissue Expression Project indicate that candidate variants at the 19p12 outcome locus significantly associate ($q < 0.05$) with *ZNF100* expression in normal ovarian tissue. Through 3C analyses of the 19p12 outcome locus in two ovarian cancer cell lines (OVCAR8 and COV362), we identified two putative regulatory elements (PREs) that physically interact with the *ZNF100* promoter. Allele-specific 3C analysis indicated that the major alleles of two candidate variants (rs10421113 and rs10421342) correlated with the interaction of the *ZNF100* promoter and one of the PREs. Since this PRE was found to increase *ZNF100* promoter activity in ovarian cancer cells, these alleles may associate with greater *ZNF100* expression in ovarian tumors. At the 1q22 outcome locus, candidate variants were located in a super-enhancer that was predicted to regulate *MEF2D* in ovarian tissue. The *MEF2D* promoter was too close to candidate variants for 3C analysis but the super-enhancer and eQTL data from normal ovarian tissue suggested that *MEF2D* is a regulatory target. A PRE coincident with the super-enhancer and containing the major alleles of four candidate variants (rs11264488, rs10159180, rs11264489 and rs942964) acted as an enhancer and increased *MEF2D* promoter activity by 117% ($p < 0.0001$) and 61% ($p = 0.03$) in OVCAR8 and COV362 cells, respectively. In OVCAR8 cells, a haplotypic variant of the PRE, carrying minor alleles of the candidate variants, increased promoter activity by a further 21% ($p = 0.03$ after multiple comparison correction).   In summary, we have found functional evidence that candidate variants at two loci associated with ovarian cancer survival may mediate effects on outcome through regulation of *ZNF100* and *MEF2D*.

2904F

Rapid CNV characterization of clinical cancer samples on the Oxford Nanopore MinION. *S. Goodwin[1], F. Sedlazeck[2], T. Baslan[3], R. Levine[3], W.R. McCombie[1], M.C. Schatz[1,2].* 1) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) John Hopkins University, Baltimore, MD; 3) Memorial Sloan Kettering Cancer Center, New York, NY.

Acute myeloid leukemia (AML) is a complex, devastating cancer of the myeloid blood cells where the cells lose the ability to differentiate normally. While this rare disease only accounts for ~1% of cancer deaths annually, the mortality rate exceeds 50% and left untreated can lead to death within just a few weeks. Approximately 30% of patients present with complex karyotypes upon conventional cytogenetic examination, and these patients have a poor prognosis. Copy number variations are widely recognized as contributing to cancer pathogenesis, patient prognosis and therapeutic response. DNA sequencing can provide molecular cytogenetic information, which in turn provides improved patient classification and can direct sub-type specific therapies. Despite recent advancements in Next-Generation Sequencing (NGS) approaches, clinical genetic sequencing often takes weeks before actionable results are generated due to the substantial investment in infrastructure required for most sequencing pipelines. This time to actionable results is time which AML patients may not have. The Oxford Nanopore MinION device could theoretically provide clinicians with a tool to generate rapid low-coverage copy number profiles, equivalent to cytogenetic examination in a point-of-care setting. We have developed a pipeline for generating copy number profiles from clinical cancer samples in three days from sample collection to analysis. Using the widely rearranged SKBR3 cancer cell line as a model we have been able to generate >300k short reads on during a single MinION run followed by CNV analysis by Ginkgo, software capable of profiling CNV in low-coverage samples. These profiles have been able to identify nearly all CNVs called by Ginkgo from MiSeq libraries. This approach can be used by clinicians to accurately and rapidly to characterize the CNV profile from clinical samples and may help in informing patient treatment. With upcoming improvements to the Oxford Nanopore platform and software we are optimistic that the time from sample collection to results can be decreased even further.

2905W

Genetic revolution or evolution: Will complex single gene variance within breast cancer tumors challenge the long standing genetic principles that assume that universal laws govern genotype to phenotype relationships. *B. Gottlieb[1,2,3], F. Babrzadeh[7], K. Klein Oros[1,6], C. Alvarado[1], C. Wang[7], B. Gharizadeh[7], M. Basik[1,2,4,5], C.M.T. Greenwood[1,5,6], L.K. Beitel[1,2,3,4], M. Trifiro[1,2,3,4].* 1) Cell Genetics, Lady Davis Institute for Medical Research, Montreal, Quebec, Canada; 2) Segal Cancer Center, Jewish General Hospital, Montreal, Quebec, Canada; 3) Human Genetics, McGill University, Montreal, Quebec, Canada; 4) Medicine, McGill University, Montreal, Quebec, Canada; 5) Oncology, McGill University, Montreal, Quebec, Canada; 6) Epidemiology, McGill University, Montreal, Quebec, Canada; 7) Stanford Genome Technology Center, Stanford University, Palo Alto, California, USA.

An additional level of diversity in breast tumors has been revealed by sequencing at much greater than usual depths and by using sequencing analysis tools that have identified for the first time a degree of genetic heterogeneity within individual genes, a condition that we have termed, complex single gene variance (CSGV). An analysis of six breast tumors revealed the presence of CSVG within the androgen receptor gene. A total of 53 mutations were identified, although always only in a minority of cells within each tumor. The presence of different single gene mutations in a minority of cells within a tumor has never been considered a possibility and therefore when identified usually ignored by being assumed to be the result of sequencing or sequence analysis errors. This assumption will be discussed in light of new developments in sequence analysis. Hopefully, the ability to identify CSGV in cancer tissues will further contribute to our understanding of both carcinogenesis as well as resistance to chemotherapy by being able to more closely follow putative mutations in cancer-associated genes through cancer ontogeny and treatment. Although the idea that breast tumors even within individuals may have unique and highly complex genetic profiles could be seen as a daunting obstacle in the fight to cure cancer. In the "age of personalized medicine" we believe that identifying CSGV when and where it occurs in breast tumor development is an important step to understanding breast cancer ontogeny. Finally, CSVG challenges the traditional scientific philosophy of "parsimony" that underlies our desire to find the common gene mutations that contribute to breast cancer, in the belief that doing so is essential if we are to make progress in understanding the role of genomics in breast cancer.

2906T

Genomic survey of microsatellite instability reveals cancer-specific signatures and driver mutations. *R.J. Hause[1], C.C. Pritchard[2], J. Shendure[1], S.J. Salipante[2].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Laboratory Medicine, University of Washington, Seattle, WA.

Microsatellites, 1-5 base pair repetitive sequences present throughout the human genome, can abnormally shorten or lengthen because of defects in the DNA mismatch repair (MMR) system, resulting in a "microsatellite instability" (MSI) phenotype. MSI is a key prognostic and diagnostic tumor phenotype that has been well studied by conventional methods. However, both the genomic landscape of MSI events and differences in MSI among cancer types remain poorly illuminated. To address this, we analyzed over 200,000 incidentally sequenced microsatellite loci across 4,224 cancer exomes spanning 18 different cancer types from The Cancer Genome Atlas, to our knowledge the first comprehensive investigation of the landscape of microsatellite instability in cancer exomes. We constructed a highly accurate global classifier for MSI (MOSAIC) that achieved 95.0% sensitivity and 97.8% specificity compared with gold-standard MSI calls based on the revised Bethesda guidelines. We observed that MSI-low (MSI-L) samples did not display significant differences from MS-stable (MSS) samples in the number of MSI events and support discontinuation of the use of MSI-L as a distinct classification. Comparative examination of MSI revealed both cancer-specific and general loci associated with global MSI, such as a frameshift mutation in an 8-bp polyadenine tract in exon 10 of the tumor suppressor *ACVR2A* which was observed in 56.6% of MSI-high (MSI-H) cancers. Lastly, we observed a correlation between survival outcomes and the overall burden of unstable microsatellites, suggesting that MSI may be a continuous, rather than discrete, phenotype that is informative across many cancer types. Our results provide a comprehensive view of MSI in cancer exomes, highlighting both conserved and cancer-specific MSI properties reflecting selective pressures and identifying candidate cancer drivers. Future work will extend these analyses to whole genomes and attempt to functionally validate unstable microsatellite candidates as causally influencing cancer phenotypes.

2907F

Effects of G6PD activity inhibition on the mechanical properties and cytoskeletons of cervical cancer cells. *W. Jiang, Z. Fang, C. Jiang, Y. Feng, R. Chen, X. Lin, Z. Zhang, L. Han, X. Chen, H. Li, Y. Guo.* Medical genetics, Sun Yat-sen Univeristy, Guangzhou, China.

Glucose-6-phosphate dehydrogenase (G6PD) deficiency has been revealed to be involved in the efficacy to anti-cancer therapy but the mechanism remains unclear. In our previous study, we have observed increased ROS generation caused by the inhibition of G6PD activity using Dehydroepiandrosterone (DHEA) and shRNA technology. We aimed to investigate the further anti-cancer mechanism of G6PD deficiency. In the present study, Peak Force QNM Atomic Force Microscopy was used to assess the changes of topography and biomechanical properties of cells and detect the effects on living cells in a natural aqueous environment. Moreover, Laser Scanning Confocal Microscope was used to observe the alterations in cytoskeleton to explore the involved mechanism. When G6PD was inhibited by DHEA or shRNA interference, the abnormal Young's modulus and increased roughness of cell membrane were observed in HeLa cells. Simultaneously, the inhibition of G6PD activity caused the disorganization of microfilaments and microtubules of cytoskeletons. Our results indicated the anti-cervix cancer mechanism of G6PD deficiency may be involved with the abnormal reorganization of cell cytoskeleton and abnormal biomechanical properties caused by the increased ROS due to the inhibition of G6PD activity. The results suggested suppression of G6PD may be a promising strategy in developing novel therapeutic methods for cervical cancer.

2908W

A recurrent *ERCC3* truncating mutation confers moderate risk for breast cancer.

V. Joseph[1,2], S. Topka[1,2], D. Villano[1,2], V. Ravichandran[1,2], K.N. Maxwell[3], A. Maria[1,2], T. Thomas[1,2], P. Gaddam[1,4], A. Lincoln[1,4], B. Wenz[3], S. Carmi[5], K.A. Schrader[6], S.N. Hart[7], S.M. Lipkin[8], S.L. Neuhausen[9], M.F. Walsh[1,4,10], L. Zhang[1,11], Z.K. Stadler[1,4,8], M. Robson[1,4,8], J.N. Weitzel[12], M.J. Daly[13,14], F.J. Couch[7,15], K.L. Nathanson[3,16,17], L. Norton[1], G. Rennert[18], K. Offit[1,2,4,8]. 1) Dept. Medicine, Memorial Sloan Kettering Cancer Center, NEW YORK, NY; 2) Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center (MSKCC) New York, N.Y., 10065; 3) Division of Hematology/Oncology, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia; 4) Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering, New York, N.Y., 10065; 5) Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, 9112102, Israel; 6) British Columbia Cancer Agency, Canada's Michael Smith Genome Sciences Centre, Vancouver; 7) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905; 8) Department of Medicine, Weill Cornell Medical College, New York, N.Y., 10065; 9) Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, 91010; 10) Department of Pediatrics, Memorial Sloan Kettering Cancer Center (MSKCC), New York, N.Y., 10065; 11) Department of Pathology, Memorial Sloan Kettering Cancer Center (MSKCC), New York, N.Y., 10065; 12) Clinical Cancer Genetics (for the City of Hope Clinical Cancer Genetics Community Research Network), City of Hope, Duarte, CA, 91010; 13) Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, 02142; 14) Center for Human Genetic Research and Department of Medicine, Massachusetts General Hospital, Boston, MA, 02114; 15) Department of Laboratory Medicine and Pathology, and Health Sciences Research, Mayo Clinic, Rochester, MN, 55905; 16) Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, 351 BRB 2/3, 421 Curie Blvd, Philadelphia, PA 19104; 17) Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA; 18) Cialit National Israeli Cancer Control Center and Department of Community Medicine and Epidemiology, Carmel Medical Center and B Rappaport Faculty of Medicine, Haifa, Israel.

   Known gene mutations account for approximately 50% of the risk for breast cancer. However, over half of the heritable risk remains unexplained. Fifteen percent of the heritable risk is accounted for by rare variants, while SNPs explain another 14% of the heritable risk. Many moderate risk variants, being discovered by next generation sequencing, account for an as yet unknown proportion of the remaining heritability. A truncating mutation c.325C>T:p.Arg109* in the ATP dependent helicase ERCC3 was observed recurrently among exomes sequenced in BRCA negative, breast cancer-affected individuals of Ashkenazi Jewish ancestry. Modeling of the mutation in ERCC3 deficient or CRISPR/Cas9 edited cell lines showed a consistent pattern of reduced expression of the protein and concomitant hypomorphic functionality when challenged with UVC exposure or treatment with the DNA alkylating agent IlludinS. DNA damage repair capability and overall survival of cells was compared for the mutant and wild type, demonstrating a significantly reduced survival following DNA damage. Overexpressing the mutant protein in ERCC3 deficient cells only partially rescued their DNA repair-deficient phenotype. Comparison of frequency of this recurrent mutation in over 6500 chromosomes of breast cancer cases and 6800 Ashkenazi controls showed significant association with breast cancer risk (OR_{BC}=1.53, OR_{ER+}=1.73) particularly for the estrogen receptor positive (ER+) subset (P<0.007). Significance: A functionally significant recurrent ERCC3 mutation increased the risk for breast cancer in a genetic isolate. Mutated cell lines showed lower survival after in vitro exposure to DNA damaging agents. Thus, similar to tumors arising in the background of homologous repair defects, mutations in nucleotide excision repair genes such as ERCC3 could constitute potential therapeutic targets in a subset of hereditary breast cancers.

2909T

Evolution of multiple cell clones over a 29 year period of a chronic lymphocytic leukemia patient.

S. Liu[1,2], Z. Zhao[1,3], L. Wu[1], Y. Hou[1,4], M. Dean[5,1]. 1) BGI, Shenzhen, Guangdong, China; 2) School of Life Sciences, Sun Yat-sen University, Guangzhou, China; 3) School of Biological Science and Medical Engineering, Southeast University, Nanjing, China; 4) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, Maryland, USA.

   Chronic lymphocytic leukemia (CLL) a frequent B-cell malignancy characterized by recurrent somatic chromosome alterations and a low level of point mutations. We performed SNP microarray analyses of a single chronic lymphocytic leukemia (CLL) patient over 29 years of observation and treatment, and transcriptome and whole genome sequencing at selected timepoints. We identified the chromosome alterations 13q14-, and 6q- and 12q+ in early cell clones, elimination of clonal populations following therapy, followed by the appearance of a clone containing trisomy 12 and a chromosome 10 copy neutral LOH that marks a major population dominant at death. Serial single cell RNA sequencing revealed an expression pattern with high *FOS*, *JUN* and *KLF4* at presentation, which resolved following therapy, but reoccurred following relapse and death. The pseudo-temporal analysis further characterized the cancer initiation, proliferation and extinction, and relapse, indicating that complex changes in expression occur over time. Several signalings, including B cell receptor signaling, EGFR signaling, ERKs signaling and NF-kappa B signaling, showed expression changes along evolution. In conclusion, we performed a very detailed and extensive analysis of a single CLL patient across the evolutionary lifespan. And we provided the evidence that CLL can evolve gradually during indolent phases, and undergo rapid changes following therapy.

2910F

Splice-site mutations in MET as a potential driver of lung adenocarcinoma. *D.C.H. Lu[1], V. Martinez[1], K. Thu[1], S. Lam[2], W. Lam[2], H. Varmus[3], W. Lockwood[1], NISC Sequencing Program.* 1) Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada; 2) NIH Intramural Sequencing Center, NHGRI, NIH, Bethesda MD; 3) Sandra and Edward Meyer Cancer Center, Weil Cornell Medical College, New York, NY.

Lung cancer is a complex genetic disease, suffering from a lack of curative therapies available for treating advanced tumours that often present upon diagnosis. Despite their high mutational load, many malignancies are driven and sustained by a single oncogenic mutation affecting key cellular pathways. This concept is known as "oncogene addiction", and forms the basis behind the use of targeted therapies to combat the molecular mechanisms driving cancer development. However, many lung cancers do not contain mutations in currently known oncogenic drivers. To uncover new targetable mutations, we performed whole exome sequencing of lung adenocarcinoma (LAC) samples from 15 patients. By stratifying patient samples based on known oncogenic drivers, we discovered that mutations in the hepatocyte growth factor receptor (MET [OMIM 164860]) were exclusive of other known oncogenic drivers. These findings were subsequently confirmed in an expanded panel of 83 LACs and by cross-referencing with recently published LAC sequencing studies including The Cancer Genome Atlas. Our expanded analysis and assessment of paired RNA-seq data revealed that MET is most commonly mutated through a splice site deletion that leads to skipping of exon 14. This exon contains the CBL binding site that is necessary for receptor degradation following activation by HGF, suggesting that splice site mutations affecting MET could act as the driver of tumourigenesis in LACs. To explore this possibility, we performed siRNA knockdowns of MET in LAC cell lines containing either wild-type MET, amplified MET, or the MET splice mutant. We observed cell death only in MET amplified lines, suggesting that cells are not dependent on splice mutant MET for survival. To further explore whether MET splice site mutations can act as an oncogenic driver, we have generated mice which conditionally overexpress MET and its splice mutant counterpart in lung epithelial cells. These mouse models are being used to understand whether the MET splice mutant is hyperactive and capable of driving as well as sustaining lung tumourigenesis, comparing its effects with the overexpression of wild type MET. To supplement our in vivo work, we have also generated human cell lines overexpressing the MET splice mutant. Our in vitro work will serve to help uncover genetic interactions with the MET receptor, and facilitate the testing of drugs to determine whether therapeutic inhibition of MET is capable of inducing tumour regression. .

2911W

Identification of copy number variations in high-risk lung cancer families from Louisiana. *D. Mandal[1], K. Wood[1], K. Broussard[1], A. Bencaz[1], M. Baddoo[2], J. Chambliss[1], R. Govindan[3], S.M. Pinney[4], M. deAndrade[5], C. Gaba[6], P. Yang[5], M. You[7], A.G. Schwartz[8], C.I. Amos[9], M.W. Anderson[4], J.E. Bailey-Wilson[10].* 1) Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, New Orleans, LA; 3) Washington University School of Medicine, St. Louis, MO; 4) University of Cincinnati College of Medicine, Cincinnati, OH; 5) Mayo Clinic, Rochester, MN; 6) University of Toledo Dana Cancer Center, Toledo, OH; 7) Medical College of Wisconsin, Milwaukee, WI; 8) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 9) Geisel School of Medicine, Dartmouth College, Lebanon, NH; 10) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD.

Lung cancer is the leading cause of cancer death among both men and women, and according to the estimates of the American Cancer Society, about 14% of all new cancers are lung cancers. About 10% of lung cancer cases (22,000 cases per year in the U.S.) have at least one first-degree relative affected with lung cancer, and 25% of cases have at least one first- or second-degree affected relative, indicating that family history is a significant risk factor. To identify germline genetic variants we have performed whole exome sequencing (WES) on high-risk lung cancer families (≥7 LC cases/family) from Louisiana. Identification of copy number variations (CNVs) has been reported on lung cancer cell lines and somatic samples; however, no germline specific CNVs have been reported in high-risk lung cancer cases. Studies on CNVs are becoming essential in understanding the genomic variation that might enable prevention and early detection of lung cancer in family members. The objective of the present study is to detect germline CNVs that might be responsible for increased lung cancer susceptibility in high-risk families. Whole exome sequencing (WES) of three selected high-risk families (≥7 LC cases/family) from Louisiana, including affected and unaffected family members, was performed. Phenotype, genotype, and smoking data are available on all family members. CNV specific algorithm incorporated in CANOES is used in order to successfully call and identify CNVs. Identified variants are visualized using Integrative Genomics Viewer (IGV). In addition to identifying CNVs, we will also be able to detect the transmission of any CNVs in multiple generations in these families. Once confirmed by IGV, these CNVs will be further validated in other families with hereditary lung cancer. In the long run, identification and hereditary transmission of CNVs might be useful in conjunction with the identification of other genomic variants in the detection of high-risk individuals with multiple cases of lung cancer in the families.

2912T

Molecular-genetic evaluation of squamous cell carcinoma of the larynx. *C.O.A. Melo[1,4], T.C. Vieira[1,2,3], A.D. Da Cruz[1,2,3], C.C Da Silva[1,2,3].* 1) Departament of Biology, Pontifical Catholic University of Goias, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 3) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 4) Genetics and Molecular Biology Graduate Program, Federal University of Goias, Campus Samambaia, Goiânia, GO, Brazil.

The larynx is a tubular structure of the upper aero digestive tract. Any pathology that affects the larynx can cause several problems in their normal physiological function, with full and direct influence on the decrease in quality of life of the individual. Among different diseases that affect the larynx, the cancer is one of the most serious. The squamous cell carcinoma of the larynx is a multifactorial disease, influenced by environmental factors, behavioral and inherent to the individual. The aim of this study was to describe the polymorphism of *TP53* gene in 20 individuals with the squamous cell carcinoma of the larynx, and investigate the microsatellite instability (MIS) and loss of heterozygosity (LOH). Samples from peripheral blood and tumor tissue biopsies from 20 patients diagnosed with laryngeal squamous cell carcinoma (SCC) were used. The investigation of the chromosome imbalances was performed by comparative genomic hybridization (a-CGH) using the commercial kit CGH/Nick Translation® (Vysis, USA), following the manufacturer's recommendations. The DNA samples were submitted to LOH and MIS analysis using a bank with eight polymorphic microsatellite markers: D17S1678; STS-M955a85; D8S135; RH68036; WI-20580; RH17792; HR 92600 and 027856. The STS-AA microsatellite DNA fragments were amplified using GeneAmp® Thermal Cycler 9700 (Perkin Elmer, USA). Regarding polymorphism *TP53* gene, the results were 65% of homozygous (p53AA) and 35% of heterozygous (p53PA). The genomic instability in the regions 8p21 and 8p22 were found 20% of MIS in both, and 5% of LOH for 8p21. For regions 17q22 and 11p15 has been identified the same frequency of MIS 30%. Found also 20% of LOH and 5% of MIS at 3p13 and 30% of LOH corresponding region 9p21. For the region 17p13.1 was observed that 20% of the samples had LOH and 35% MIS. The high frequency of MIS found in the study suggests that the process of tumorigenesis of the larynx following the mechanism of mutant phenotype. The imbalances in the regions of chromosomal gains were observed in 1q21→qter. The losses occurred in chromosome 3p21→p28, 11q23, 16p16, 16q12, 17p13→pter and 22q13. The techniques of comparative genomic hybridization have improved the understanding of genetic alterations in squamous cell carcinoma of the larynx, especially to characterize the relationship among the early precursors of laryngeal cancer, as well as to identify genomic regions that may contain oncogenes and tumor suppressor genes.

2913F

Analysis of genetic variability and mutational spectrum of ABCG2 transporter in Mexican patients with prostate cancer. *G.M. Mimendi Aguilar[1], E. Sierra[2], M. Villanueva[3], M. Dean[4], S. Sanchez[5], J. Sanchez[6], S. Flores[6], M. Rosales[6], M. Romero[7], M.C. Moran Moguel[1].* 1) Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Servicio de Urología, UMAE Hospital de Especialidades, IMSS, Guadalajara, Jalisco, Mexico; 3) Unidad de Anatomía Patológica, UMAE, Hospital de Especialidades, IMSS, Guadalajara, Jalisco, Mexico; 4) Laboratory of Translational Genomics, Division of cancer Epidemiology and Genetics, NCI, NIH Advanced Technology Center, Gaithersburg, MD; 5) Servicio de Urología UMF No 45, IMSS, Guadalajara, Jalisco, Mexico; 6) División de Medicina Molecular, CIBO, IMSS, Guadalajara, Jalisco, Mexico; 7) Facultad de Medicina, UAG, Guadalajara, Jalisco, Mexico.

Background. Prostate cancer is the most common cancer diagnosed in men and the sixth leading cause of cancer-related death in men worldwide. Most cancers arise from tissue specific stem cells. Several characteristic markers identify cancerous prostate stem cells like CD44, CD133, integrin α and β and ABCG2 transporter protein involved in efflux dihydroxy-androgens such as testosterone, which has been proposed as a mechanism for maintaining phenotype of prostate stem cells (cancer). The Objective of this study was to identify gene variability in *ABCG2* in genomic DNA (gDNA) tumor tissue from Mexican patients with prostate cancer. **Methods.** gDNA was obtained from 32 patients attending at the Urology Service at Unidad Médica de Alta Especialidad, Hospital de Especialidades CMNO with confirmed diagnosis of prostate cancer. 16 exons of *ABCG2* gene were amplified by polymerase chain reaction (PCR). PCR products were analyzed on agarose gels stained with ethidium bromide and purified before sequencing using equipment Life Technlogy®3730 DNA analyzer. The electropherograms were analyzed using the "Surveyor DNA mutation analysis software variant" (Softgenetics®). **Results.** The sequence analysis of 1,024 *ABCG2* gene fragments of 32 mexican patients with prostate cancer (64 chromosomes) revealed the presence of 15 variants located in coding region (exons) and 8 in intronic regions. In exon 9 six different variants were identified. The frequency of the different alleles varies between 3.12% and 68.75%. **Conclusion.** This study is the first in which are identified and described variants of *ABCG2* gene in patients with prostate cancer in Mexican population, most of them have been previously described and variability is similar to that described in other populations, except Asian population. The transition (G/A) is the most frequent substitution found between variants of *ABCG2* gene and exon 9 is likely a hot spot mutation in patients with prostate cancer of the Mexican population.

2914W

Second-hit" genes in lung adenocarcinoma. *M. Oh, W. Lockwood, B. Harbourne.* University of British Columbia, Vancouver, Canada.

   Lung cancer is the leading cause of cancer death worldwide. Despite the continuing efforts to better understand its molecular origins and to develop an effective therapeutic agent, the 5-year survival rate remains poor at 18% for lung cancer patients. Recent studies have uncovered several key genes, like the Epidermal Growth Factor Receptor (EGFR) gene, that are frequently activated in lung adenocarcinoma (LAC) and drive lung tumorigenesis. the advent of small molecule tyrosine-kinase inhibitors (TKI), such as Erlotinib, that target EGFR and interfere with its activation. Despite new targeted therapies being introduced, the long-term prognosis still remains low largely due to the acquired resistance to TKIs in patients.   Our project aims to find secondary mutations in LAC that cooperate with EGFR to help accelerate lung tumorigenesis. To find candidate genes, we have analyzed tissue specimens from LAC patients through integrative genomic analysis. Our studies revealed a copy number gain in chromosome 16p in mutant-EGFR tumors, which corresponded with increased expression in Golgi-localized, gamma-ear-containing, ARF-binding proteins 2 (GGA2). Functionally, GGA2 has been implicated in protein trafficking and sorting. We hypothesize that GGA2 interacts with internalized EGFR and traffics the receptors to the cell surface. Consequently, overexpression of GGA2 in LAC tumors will result in the accumulation of mutant EGFR in cells and increased EGFR signalling, which will help drive tumor progression To determine whether GGA2 interacts with EGFR, we performed co-immunoprecitation assays on cells that overexpress EGFR and GGA2. We also generated GGA2 knockdown and overexpression lines to observe whether GGA2 induces a phenotypic change in EGFR-mutant lung cancer cells. Different techniques like Western blot and flow cytometry were performed to elucidate detailed interaction mechanisms.   Our project aims to demonstrate whether GGA2, a candidate "second-hit gene", helps increase EGFR signalling by trafficking the receptors to the cell surface. GGA2 can further be implicated in acquired resistance in EGFR LAC as GGA2 acts to help EGFR maintain its oncogenic activities. Ultimately, GGA2 can become a potential therapeutic target in LAC and lead to advancements in chemotherapeutic strategies for lung cancer.

2915T

Microsatellite instability in chronic myeloid leukemia using D17S261 and D3S643 markers - A pilot study in Western India. *T.N. Patel[1], M. Chakraborty[2], P. Bhattacharya[2].* 1) Department of Medical Biotechnology, VIT University, Vellore, India; 2) Department of Biomedical Sciences, VIT University, vellore, India.

   Tumor progress through a series of genetic alterations that involve proto-oncogenes and tumor suppressor genes. Microsatellite, the short tandem repeat sequences, present over the span of human genome, can show instability at multiple loci due to error in DNA mismatch repair machinery. The study was aimed to evaluate the association between microsatellite instability (MSI) and evolution of CML.The markers chosen for this study were D17S261 (Mfd41) and D3S643 (CI3-9). Five patient and three control DNA samples were provided by Genexplore Diagnostics and Research Centre Pvt. Ltd. in MoU with VIT University as per ethical guidelines. The DNA was used to polymerase chain reaction (PCR) with normal and fluorescent primers and the PCR products were subjected to bioanalyzer and fragment analysis. Bioanalysis and fragment analysis results on comparison showed deviated results. Bio-analysis did not show any significant variation in (AC) repeats but fragment analysis indicated presence of MSI in most of the patients as compared to normal individuals for both the markers. Although bioanalysis was performed to check the presence of altered peak for the marker chosen, fragment analysis helped in confirmation of MSI. These findings suggest that MSI is an event tat may also contribute towards CML progression and may induce resistance to various drugs.

2916F

Comparative genomic analysis of NF1-associated atypical neurofibromas (ANF) and malignant peripheral nerve sheath tumors (MPNST). *A. Pemov[1], N.F. Hansen[2], R. Patidar[3], J.F. Boland[4], S.C. Chandrasekharappa[2], J.C. Mullikin[2,5], NISC corporate author[5], CGRL corporate author[6], M. Wallace[7], J. Khan[3], E. Legius[8], B. Widemann[9], D.R. Stewart[1].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD; 2) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, Rockville, MD; 3) Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD; 4) CGRL, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 5) NIH Intramural Sequencing Center, National Human Genome Research Institute, Rockville, MD; 6) Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 7) Department of Molecular Genetics and Microbiology, UF Genetics Institute, UF Health Cancer Center, University of Florida, Gainesville, FL; 8) Department of Human Genetics, Catholic University Leuven, Leuven, Belgium; 9) Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD.

   BACKGROUND. NF1-associated ANF are tumors characterized by atypical pathological features that often arise within plexiform neurofibromas (PNF) in NF1 and can transform into MPNST. Currently, it is not clear what causes PNF transformation into ANF and further into MPNST, however several studies identified deletion of the *CDKN2A/2B* locus as the most frequent genetic event distinctive from PNF in ANF. In this study, we performed genomic analysis of 16 ANF and 4 MPNST matched with germline DNA obtained from 14 and 4 unrelated patients, respectively. All tumors but 2 MPNST (cell lines) were primary ones. **METHODS.** We performed Whole Exome Sequencing (WES) on Illumina Hi-Seq 2500 platform (96 Mb SeqCap EZ Exome + UTR Library), copy-number analyses by Illumina HumanOmniExpressExome SNP-arrays, and deep sequencing of *NF1* and validation of select mutations on ThermoFisher IonTorrent PGM platform. **RESULTS. ANF.** We identified *NF1* germline mutations in 94% and somatic mutations in 81% of samples, and heterozygous *CDKN2A/2B* deletion in 75% of samples. In addition, there were 13 other somatic mutations in 9 tumors (median 1, range 0-4; 7 samples could not be validated due to lack of DNA). *NF1* and *CDKN2A/2B* were the only genes having inactivating mutations in multiple samples. We observed an elevated level of genomic instability in the tumors. **MPNST.** We identified *NF1* germline and somatic mutations in 100% of samples, somatic *CDKN2A/2B* locus deletion in 100% of samples (in 3/4 samples the deletion was homozygous) and inactivating mutations in PRC2 genes *EED* and *SUZ12* (both were accompanied by loss of heterozygosity (LOH) thus becoming homozygous). There were no PRC2 mutations in the remaining MPNST, however we observed two deleterious mutations in a histone H4-specific acetyltransferase gene. In addition, there were 69 somatic mutations in 3 tumors (23 mean, 18-31 range). *NF1*, *CDKN2A/2B* and PRC2 complex genes were the only ones inactivated in multiple samples. We observed highly re-arranged chromosomal architecture in all tumors. **CONCLUSIONS.** It appears that PNF-ANF transition is predominantly driven by heterozygous deletion of the *CDKN2A/2B* locus. Further progression to MPNST likely involves an additional inactivation of one of the components of the PRC2 complex or possibly other chromatin modifying genes. In our data, large deletions result in LOH of PRC2 mutations and *CDKN2A/2B* locus deletions and are necessary for malignant transformation of these tumors.

2917W

Insecticide exposure induces gene rearrangements associated with hematological malignancies. *P. Perez-Vera, M.P. Navarrete-Meneses, C. Salas-Labadia, M.A. Sanabrais-Jimenez, J. Santana-Hernandez, A. Serrano-Cuevas, A. Olaya-Vargas.* Genética Humana, Instituto Nacional de Pediatría, MEXICO, Mexico.

   Insecticides are used to control insect pests worldwide, and general population is commonly exposed to these agents. Epidemiological studies have shown that exposure to insecticides is associated with an increased risk of hematological cancer in adults and children. In addition, it was demonstrated that these agents damage DNA. However, there is a lack of biological evidence showing if insecticides can induce genetic aberrations associated with hematological cancer. The aim of this study was to assess if *in vitro* exposure to insecticides induces gene rearrangements associated with hematological malignancies. Mononuclear cells from two healthy volunteers were cultured and exposed to 200μM of permethrin or malathion for 24h. *MLL*, *ETV6-RUNX1* and *IGH*, leukemia/lymphoma associated genes, were analyzed independently with FISH. Aneuploidy was assessed with centromeric probes for chromosomes 12, 18 and 14/22. *MLL-AF4*, *ETV6-RUNX1* and *IGH-BCL2* oncogenic fusions were detected by RT-PCR in mononuclear cells from GMCSF-stimulated peripheral blood (5μM, 96h), bone marrow (0.1μM 72h), peripheral blood (200μM 24h), as well as in K562 cell line (0.1 μM, 24h). Insecticide exposure increased the frequency of cells with *MLL* structural aberrations (p<0.05). A high diversity of abnormal patterns were observed, the most frequent aberration was the rupture at the specific leukemia breakpoint. Permethrin also induced *IGH* breakage at J segment and deviations in the number of gene signals as a consequence of aneuploidy (p<0.05). Malathion induced *MLL-AF4* and *ETV6-RUNX1* gene fusions, and permethrin induced *IGH-BCL2* gene fusion in K562 cells. Moreover, permethrin exposure induced *ETV6-RUNX1* in peripheral blood, bone marrow and GMCSF-stimulated peripheral blood. Permethrin and malathion exposure can affect leukemia/lymphoma-related genes in structure and number. This study provides biological evidence showing that insecticide exposure is implicated in the generation of leukemia/lymphoma associated gene aberrations, which could drive the development of these malignancies, as epidemiological studies have suggested. Acknowledgements: To CONACYT MPNM-385279, CONACYT CB 2012-183467-M, Recurso Federal 2016   .

2918T

Public resources from the NIH's Genotype-Tissue Expression (GTEx) program: Biospecimens, data, SOPs and more. *A.K. Rao[1], P. Guan[1], S. Koester[2], S. Volpi[3], H.M. Moore[1], The GTEx Consortium.* 1) National Cancer Institute, Bethesda, MD; 2) National Institute of Mental Health, Bethesda, MD; 3) National Human Genome Research Institute, Bethesda, MD.

The Genotype-Tissue Expression (GTEx) project, a National Institutes of Health (NIH) Common Fund program, aims to better understand how genetic variation influences gene expression in normal human tissues. Non-diseased tissues were collected from nearly 1000 postmortem donors for genomic analyses with the resulting data deposited into the database of Genotypes and Phenotypes (dbGaP) at NIH. Additional information imported into dbGaP for each donor includes clinical, genotype, and gene expression data, pathology review, and data from expression quantitative trait loci (eQTLs). Biospecimens were obtained with the consent of donor families through partnerships with Organ Procurement Organizations. Researchers are encouraged to request access to utilize residual GTEx biospecimens. A new, open access GTEx histological image library where images can be viewed with zooming capability is available. The GTEx program developed and has made publicly available the Standard Operating Procedures (SOPs) detailing collection procedures. Public resources from GTEx include: 1. GTEx Portal: an open access database of GTEx expression data and analysis results: http://www.gtexportal.org 2. Access to residual GTEx biospecimens: http://www.gtexportal.org/home/samplesPage and https://specimens.cancer.gov/ 3. dbGaP: controlled access of comprehensive GTEx clinical and raw sequencing data: http://www.ncbi.nlm.nih.gov/gap 4. SOPs from GTEx biospecimen collections: http://biospecimens.cancer.gov/resources/sops/library.asp 5. GTEx histological image viewer: http://biospecimens.cancer.gov/resources/tissue_image_library.asp 6. Training videos for consenting personnel: http://gtextraining.org/ 7. Various manuscripts detailing the GTEx program available through PubMed   Data from GTEx is already being used in numerous projects to better understand predisposition to disease, with several high profile studies already published. The breadth of the program has afforded multiple, varied public resources from biospecimens to molecular analyses datasets. Furthermore, an embedded ELSI study continues to study the factors affecting consent for research tissue donation. The full potential of GTEx resources has yet to be fulfilled; however, it is likely that these resources will be useful tools for medical research immediately and in future years.

2919F

Genetic interaction networks in *S. cerevisiae* identify cancer relevant Synthetic Cytotoxicity (SC) interactions with cohesin mutations. *S. Reytan[1,2], H. Li[2], M. Kofoed[2], N. O'Neil[2], P. Hieter[2].* 1) Medical Genetics, The University of British Columbia, Vancouver, BC, Canada; 2) Michael Smith laboratories, The University of British Columbia, Vancouver, BC, Canada.

Whole genome sequencing technologies are advancing at an unprecedented pace, opening new opportunities for the genotype-driven personalized treatment of cancer. Synthetic Lethality (SL) based therapeutics have emerged as promising approaches to target cancer specific somatic mutations. SL exploits the genetic distinctness of the malignant cell by targeting a second gene that is required for viability in the presence of a tumour specific mutation, resulting in its specific killing while the normal cells remain viable. We are expanding the list of SL partner genes by screening for conditional SL interactions, in which loss of function of 2 genes results in sensitivity to low doses of a DNA damaging agent, a concept we have called Synthetic Cytotoxicity (SC). SC has the potential to extend the number of genotypes that can be treated with existing chemotherapeutics and to improve the efficacy of these therapeutics.   We have focused our efforts by screening for SC interactions with cohesin-mutated genes. The cohesin complex is highly mutated across a wide range of tumours, including acute myeloid leukemia (12%), bladder cancer (20%), colorectal cancer (~13%), glioblastoma (~6%), Ewing's sarcoma and melanoma (~17%). We used Synthetic Genetic Array (SGA) technology, a high-throughput genetic method in yeast, to screen cohesin-mutated strains against an array of 310 deletions affecting DNA damage response genes. The screens were done in the presence and absence of four distinct genotoxic agents: benomyl (microtubule inhibitor), bleomycin (radiomimetic), camptothecin (TOP1 inhibitor), and methyl methanesulfonate (alkylating agent). We screened and analyzed ~12,400 potential genetic interactions, identifying hundreds of SL and SC interactions. The corresponding genes, when knocked-out, lead to specific killing of cohesin-mutated yeast cells in the presence of sublethal concentrations of a genotoxic agent. Many of these genes belong to conserved biological pathways such as DNA damage-response and chromatin modification, which are known to have effects on genome instability. We created a map of these interactions that can further expand our knowledge about the biological importance of these genes and their potential role in cancer.   These SC genetic interactions can inform and improve individual cancer therapies as personalized medicine treatments, and lead to discovery of new pathways or candidates for anti-cancer treatments.

2920W

Evaluation of mtDNA alterations and *BCL-2* **prompter polymorphism in involvement of metastasis on Iranian breast cancer patients.** *E. San-jari[1], M. Ghaffarpour[1,3], S. Esalamizadeh[2,4], M. Arabzadeh[2,4], B. Motahari[2,5], M. Houshmand[1,2].* 1) National Institute of Genetic Engineering and Biotechnology, Tehran, Iran; 2) Special Medical Center, Tehran, Iran; 3) Iranian Research Organization for Science and Technology; 4) Islamic Azad University,Tabriz,Iran; 5) Islamic Azad University and Research Complex, Tehran, Iran.

Breast cancer is the first most common cancer among the females in Iranian cancer patients. The role of mtDNA alterations is recognized as being in carcinogenesis, tumor progression and prognosis in breast cancer. MtDNA alterations G13397A(ND6),G8993T(ATP6) and alteration of 12308 in tRNA Leu gene have been reported in involvement in metastasis in vivo but to date there was not shown on clinical breast cancer samples as a prognosis factors. It has been also shown that BCL-2 prompter (-938C>A) polymorphism, mtDNA alterationA15326G and G10398A (ND3) correlated with increased breast cancer risks and metastasis respectively. The aim of this study was to address the question if mtDNA alterations of G13397A, G8993T, 12308G, G10398Aand BCL-2 prompter (-938C>A) polymorphism can play a role in promoting of tumor and leading to metastasis. 69 paired fresh tumor and adjacent normal samples were obtained from patients with Breast Cancer (31 of 69 who are experienced metastasis and 38 of 69 non metastasis) .Fresh tumors and their adjacent were extracted for genomic DNA. Whole mtDNA sequencing was used for investigation of mtDNA alterations. The G13397A and G8993T mutations were not seen for all patient including metastasis and non-metastasis. There was an association between 12308 G alteration and metastasis patients. It was also a role for mtDNA alterations G10398A and A15326G and an increase Breast Cancer risk. There was no association between BCL-2 prompter (-938C>A) polymorphism and Breast Cancer risk and metastasis.

2921T

Genome wide copy number profiling detects significant deletion (4q35.1) and amplifications (20p12.1, 12q14.2 and 12q14.3) in gallbladder carcinoma. *A. Sharma, A. Kumar, N. Kumari, N. Krishnani, B. Mittal.* SGPGIMS, Lucknow, Uttar Pradesh, India.

Gall bladder cancer (GBC) is a malignancy of biliary tract having an unusual geographic distribution. Late diagnosis and poor prognosis are typical features of this disease. Recent studies have shown that molecularly targeted therapies are critically dependent on genetic profile of tumour which can lead to refinement of prognostic information. Available molecular profiling data in GBC is limited due to lack of enough high-throughput screening regarding its genetic information. So understanding molecular features of GBC is critical towards improving treatment paradigm for this disease. Present study was aimed to determine mutations, copy number variations (CNVs) in 890 major cancer related genes in GBC patients. 13 histopathologically confirmed GBC cases were included in this study. GBC cells were micro-dissected from formalin fixed paraffin embedded tissue blocks having >80% of tumour cells. DNA extraction was done by standard protocols. Copy number/mutation profiling was done by "OncoScan™ FFPE Assay Kit" (Affymetrix, CA). Data analysis was done by SNPFASST2 algorithm using the Nexus Express for Oncoscan software version 3.0 and 7.5 (Biodiscovery, Inc., CA USA). A substantial number of CNVs detected were recurrent. Recurrent gains were at chromosome 12q (53%), 20p (69%) and recurrent losses were at 4q (46%). The common genes affected in CN loss process were *NPP6, IRF2, CASP3, PRIMPOL, MLF1IP, CENPU, ACSL1, SLED1* in chromosome 4, whereas genes involved in copy number gain process are *HMGA2*, *RPSAP52* on 12q14.3, *RASSF3* gene on 12q14.2 and *MACROD2* on 20p12.1. Present study provides data of copy number profiling in GBC. Significant copy number abrasions were detected in genes involved in tumor suppressing and apoptosis functions. Presence of loss of hetrozygosity was also found to be a significant character in Gallbladder cancer. More studies with greater sample size are required for validation of findings.

2922F

High-throughput functional classification of *BRCA2* **variants of uncertain significance (VUS).** *H. Shimelis, S. Andreas, H. Chunling, F.J. Couch.* Mayo Clinc, Rochester, Rochester, MN.

  Genetic testing of individuals with a family history of breast and/or ovarian cancer has led to the identification of many unique BRCA2 and BRCA1 missense variants of uncertain significance (VUS). Considering that the majority of missense changes in cancer predisposition genes lack sufficient genetic information, cancer risks conferred by many germline missense variants have not been established. Classification of these variants using quantitative models such as the established BRCA1 and BRCA2 multifactorial Likelihood Model has proved challenging due to limited availability of family data for each VUS. In contrast functional analysis of VUS using validated Functional assay may be useful for classifying VUS in the absence of case-control or family data. We have used a homology directed repair (HDR) cell-based assay to characterize missense variants in the DNA binding domain (DBD) of BRCA2. The method has been validated using known pathogenic and known non-pathogenic BRCA2 missense variants and has been associated with 100% sensitivity (95% confidence interval (CI): 75.3%–100%) and 100% specificity (95% CI: 81.5%–100%) for pathogenic BRCA2 variants. We have recently established high-throughput methods for characterization of BRCA2 variants using this HDR assay. Over 300 BRCA2 missense variants within the DBD have been evaluated, of which 80 variants are deleterious and 19 have intermediate activity. When combined with sequence-based in silico prediction models in a probability model, these variants can be classified as pathogenic and moderate risk, respectively. The HDR assay is being applied to all known missense variants in the DNA binding domain in order to provide information on the pathogenicity of all variants identified in this region of the BRCA2 tumor suppressor. In addition, the HDR assay has been used to characterize the influence of missense variants on the activity of proteins encoded by other cancer predisposition genes.

2923W

Integrated genomic characterization of tumors with somatic *NF1* **driver mutation.** *T. Zhang, M. Xu, K.M. Brown.* National Cancer Institute, Bethesda, MD.

  The *NF1* gene encodes a RAS GTPase-activating protein called neurofibromin that influences RAS-MAPK signaling. Somatic loss-of-function driver mutations in *NF1* have bene found in numerous tumor sequencing studies; however, genetic and genomic commonalities between tumor types for which *NF1* is a driver, which may provide mechanistic clues about the role of *NF1* mutations in the development of these tumors, remain unclear. We systematically analyzed and compared all available expression, mutation, CNV, and methylation data for 15 tumor types from TCGA (3,131 sequenced samples). We first used MutSig2CV to identify tumor types for which *NF1* was among the most significantly mutated (driver) genes: cutaneous melanoma (SKCM), lung adenocarcinoma (LUAD), glioblastoma multiforme (GBM), pheochromocytoma and paraganglioma (PCPG), brain lower grade glioma (LGG), and ovarian serous cystadenocarcinoma (OV). Driver mutations in several genes were found only in *NF1*-mutant tumor types (*BRAF*, *EGFR*, *IDH1*, *ARID2*, *EMG1* and *STK19*), while mutations in others (*FBXW7* and *MLL2*) were only found in non-*NF1* tumor types. mRNA expression of *NF1* was significantly lower in *NF1*-mutated samples compared to wild-type samples across all six *NF1*-mutant tumor types. Expression enrichment analysis identified multiple genes whose correlation with *NF1* differed between *NF1* and non-*NF1* tumors, including multiple Ras family genes. Pathway enrichment analysis for these genes further suggested activation of the MAPK-ERK-Rho signaling pathway in *NF1* driver tumors, consistent with the known role of *NF1* as a negative regulator of Ras. While mutation enrichment analysis identified multiple genes with mutations that co-occurred with *NF1* mutations, this co-occurrence was observed in only tumor types with a high mutation burden (SKCM, LUAD and UCEC). Hotspot mutations in four genes were found to be mutually exclusive with the *NF1* mutation in *NF1* tumors (*IDH1* in LGG, *KRAS* in LUAD, *BRAF* in SKCM, and *EGFR* in GBM). However, co-occurrence between less-frequent (non-hospot) mutations of these genes and NF1 was observed, suggesting that *NF1* mutations can amplify the effects of other driver gene mutations. Finally, methylation-based enrichment analysis determined that hypermethylation contributed to *NF1* inactivation in SKCM and LGG. In conclusion, these data suggest that *NF1* inactivation is a common genomic alteration in six different human cancer types, cooperating with other non-hotspot driver gene mutations.

2924T

Detecting cancer-associated structural variants by using megabase-scaled DNA molecules and whole-genome sequencing. *J. Xu[1], D. Bann[2], R. Clark[3], Y. Wang[1], J.R. Broach[1,2], F. Yue[1,2].* 1) Biochemistry Dept, Penn State Univeristy; 2) Institute of Personalized Medicine, Penn State Univeristy; 3) Computer and Info System Dept, Penn State University.

A typical cancer cell manifests abundant genomic structure variants (SV), including copy number alteration (CNA) such as insertions and deletions, and positional rearrangement such as inversions and translocation. It has been shown that a single SV can drive oncogenic transformation, through CNA, gene fusion, or disrupting the linkage between enhancers and their target genes. To assess the global effect of the overall SVs from a cell on its cancerous phenotype, we profiled and characterized all types of SVs in three cancer cell lines Caki2, T47D and K562, by incorporating whole-genome sequencing (WGS) and genome maps generated by Bionano Irys platform, a new technology that uses megabase-scaled DNA molecules and can capture large-scale SVs. Our approach identified genome aneuploidy, hundreds of insertions and deletions, and tens of inversions and translocations from each cell line. We also identified several hundred genes deleted in each cell line, and stratified them by copy numbers retained. Some deleted genes are known to canonically associate with cancer, such as ATM from breast cancer cell T47D and CDH1 from kidney carcinoma cell Caki2. Moreover, the T47D deleted regions are enriched of genes involved in the regulation of mammary gland epithelium proliferation. We further examined SV's impact on gene expression by comparing transcriptomes of T47D and normal breast cell HMEC, and found that expression of many partially deleted genes are gene-dose dependent. Finally, we interrogated its clinical significance by examining the TCGA breast tumor samples and reported luminal and basal-specific copy number loss of genes deleted in T47D cells, coupled with reduced gene expression.   Overall, we have shown that incorporating WGS and Bionano Irys can accurately capture SVs in cancer cells. Our results identified a spectrum of deleted essential genes in cancer cells that are representative for patient cohorts, and we believe that efficient detection of SVs encompassing those marker genes might be critical to cancer diagnosis and subtyping.

2925F

Reclassifying duplication variants in high risk cancer genes by identifying tandem duplication breakpoints. *H. Chong[1], W. Mu[2], S. Willett[1], S. Lam[1], R. Huether[2], A. Elliott[1].* 1) R&D, Ambry Genetics, Aliso Viejo, CA; 2) Bioinformatics, Ambry Genetics, Aliso Viejo, CA.

Alterations in genomic copy numbers may lead to changes in gene expression and function and are known to cause various disease states including congenital anomalies, developmental disorders and numerous cancer syndromes. Most gross deletions in high risk cancer genes, usually larger than 3~5 megabases, fall within microarray reporting guidelines and are reported as pathogenic. However, gross duplications are generally reported as variants of unknown significance (VUSs), which can then subsequently be reclassified as likely deleterious or likely benign as more information becomes available. To determine whether a copy number duplication variant can be reclassified as likely pathogenic, we utilized paired-end Next Generation sequencing (NGS) method that can detect the breakpoints of tandem duplications in high risk cancer genes, including *BRCA1, BRCA2, ATM, CHEK2, CDH1*, and *TP53*. DNA was available for NGS on 103 consenting cancer patients with a variety of gross duplications who had a personal or family history of breast and/or ovarian cancers. We designed probe sets to capture the target regions with the suggested breakpoints identified by array comparative genomic hybridization (aCGH) for 32 different gross duplication variants. The captured DNA was then sequenced by paired-end NGS and mapped to the genome. Our custom pipeline identified a tandem duplication in 79% of samples we tested (81/102). Most tandem duplications in these high risk cancer genes caused a frameshift or structural change, which may result in novel function of the genes (loss-of-function or gain-of-function) and is predicted to be damaging by most *in silico* algorithms examined. Therefore, this study suggests that 32 previously classified VUSs detected as tandem duplications can be re-classified as likely pathogenic. Proper classification of variants detected for these high risk cancer genes will help guide physicians in selecting the best course of treatments for their patients and improve the overall recovery and health of the patient.

2926W

Single-cell whole genome sequencing reveals clonal diversity and evolution in small-cell lung cancer. *D.C.J. Spierings[1], H. van den Bos[1], P. Ferronika[2], A. Saber[3], T.J.N. Hiltermann[4], K. Kok[2], A. Taudt[1,5], D. Porubsky[1], F. Foijer[1], M. Colomé-Tatché[1,5], H.J.M. Groen[4], A. van den Berg[3], P.M. Lansdorp[1,6].* 1) ERIBA, UMCG, Groningen, the Netherlands; 2) Genetics Dept, UMCG, Groningen, the Netherlands; 3) Pathology and Medical Biology Dept, UMCG, Groningen, the Netherlands; 4) Pulmonary Diseases Dept, UMCG, Groningen, the Netherlands; 5) Institute for Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; 6) Terry Fox Laboratory, BC Cancer Research Centre and UBC, Vancouver, Canada.

Small-cell lung cancer (SCLC) is characterized by aggressive growth and high metastatic potential. Intra-tumor genomic heterogeneity can facilitate tumor evolution and has been thought to be relevant to cancer resistance towards chemotherapy. Unfortunately, classic genomic profiling by comparative genomic hybridization (CGH) or whole-exome sequencing (WES) of bulk material can only identify global genomic alterations of a cell population. In the present study, we assessed intra-tumor heterogeneity in genomic copy number variations (CNVs) in a primary SCLC tumor and its metastases at the single cell level using low coverage single-cell whole genome sequencing (scWGS). Single nuclei isolated from frozen tissue sections were flow sorted and scWGS libraries were prepared without upfront whole genome amplification to reduce PCR amplification biases and maintain a direct correlation between sequence reads and genome content. The distribution of reads across the chromosomes was found to faithfully represent the chromosome copy number. An in-house developed algorithm, called AneuFinder, was used to determine the copy number state for each chromosome. Merged scWGS data of all individual cells within a sample resulted in CNV patterns similar to CGH-based CNV patterns obtained from genomic DNA of the same samples. However, at single cell level, a high degree of intra-tumor CNV heterogeneity was observed. The highest level of heterogeneity was observed in the primary tumor and in the lymph node metastasis and the lowest heterogeneity in the liver metastasis. Interestingly, within the heterogeneous population of the primary tumor, one cell displayed a CNV pattern identical to that of the merged lymph node metastasis, whereas two cells resembled the liver metastasis. These rare cells in the primary tumor would not be detected when assessing CNVs using CGH or WES on bulk material. In conclusion, scWGS is a powerful technique to assess intra-tumor CNV heterogeneity providing insight into tumor cell evolution and metastasis.

2927T

Robust whole genome sequencing technology enables structural variant detection from 100ng to 1ng. *S.K. Sandhu, J. Laliberte, C. Schumacher, S. Chupreta, V. Kelchner, L. Kurihara, T. Harkins, V. Makarov.* Swift Biosciences, Ann Arbor, MI.

Detection of structural variants (SV) from whole genome sequencing (WGS), whole exome sequencing (WES), targeted capture and amplicon sequencing data enables access to the 1.2% of the human genome, variations in which, forms the basis of human evolution, genetic diversity among individuals, and genomic disorders/syndromes. Some technical challenges exist owing to the nature of these genomic regions which are often characterized by low complexity and lack sufficient coverage, especially in cases with limiting input material. These problems are further accentuated by PCR bias and a higher frequency of misalignments in these regions. Unbiased, comprehensive, uniform coverage of whole genome ensures even representation of these low complexity regions. We describe a robust, PCR-free method which enables high quality sequencing from 100ng DNA input using Swift Accel NGS 2S technology compared to ten times more input DNA (1000ng) required for an Illumina kit. 100ng PCR-free WGS libraries were compared to 10ng and 1ng plus-PCR generated from the Coriell NA12878 DNA using Swift Accel NGS 2S kit, sequenced on HiSeq 2500 and HiSeq X to 35x coverage for all inputs and 1000ng Truseq PCR-free. Preliminary data from PCR-free sequencing of patient samples from these low inputs was also investigated. A set of curated deletions, insertions, inversions, copy number variants, and translocations reported by the 1000 Genomes Project and confirmed with PacBio and/or Illumina sequencing was evaluated from the data obtained from above libraries using multiple tools including Lumpy, Pindel, Delly, CNVnator and Breakdancer. We found a similar number of total SVs detected across all inputs even from the low input samples with PCR. However, a higher number of insertions and inversions, were detected in the 100ng PCR-free input with high level concordance to the truth list. Further analysis of these SVs is underway.

2928F

Identification of a novel alternative splicing of *PCDH17* **gene: Whole exon-2 skipping.** *T.N.U. Le[1], K. Sakashita[2,3], T.M.T. Ha[1], V.N. Nguyen[1], K. Koike[2].* 1) Medical Genetics, Hue University of Pharmacy and Medicine, Hue, Thua Thien Hue, Viet Nam; 2) Pediatrics, School of Medicine, Shinshu University, Nagano, Japan; 3) Pediatric Hematology and Oncology, Nagano Children's Hospital, Azumino, Japan.

    Introduction *Protocadherin* (*PCDH*) 17 gene is a member of cadherin superfamily, a group of cell membrane proteins that mediates Ca^{2+}-dependent cell-cell adhesion. Recently, *PCDH17* has been reported as a tumor suppressor gene in a number of solid cancers. However, little is known about this gene. Here we report a novel alternative splicing of *PCDH17* gene in peripheral blood cells. **Methods:** DNA and total RNA were extracted from peripheral blood cells. Conventional PCR was used to amplify a *PCDH17* fragment from downstream region of intron 1 to downstream region of exon 3. RT-PCR was used to examine *PCDH17* gene expression. PCR products were extracted and subjected to subcloning for sequencing. **Summary results:** We identified a present of an alternative mRNA isoform of *PCDH17* gene that skipping whole 58-bp exon 2. We also detected an existence of alternative splicing variants mediated by the tandem splice acceptor motif NAGNAG at intron 1-exon 2 boundary. There was no mutation in *PCDH17* examined sequence at DNA level.

2929W

Chromosomal instability and XRCC1 Arg399Gln and XRCC2 Arg188His polymorphism in esophageal cancer patients. *J. Kaur[1], V. Sambyal[1], K. Guleria[1], N. Singh[2], M. Uppal[2], M. Manjari[3], M. Sudan[4].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Department of Surgery; 3) Department of Pathology; 4) Department of Radiotherapy, Sri Guru Ram Das Institute of Medical Sciences and Research Vallah, Amritsar, Punjab, India.

    Background: DNA repair genes continuously monitor chromosomes to correct injuries caused by exogenous agents such as cigarette smoking, a potent risk factor for esophageal cancer. Among various DNA repair genes, XRCC1, located on chromosome 19q13.2, plays a predominant role in both single strand break repair and base excision repair. A genetic polymorphism (rs25487) in codon 399 in XRCC1 gene results into a non-conserved amino acid substitution. Another DNA repair gene, XRCC2, located at 7q36.1, is a member of a family of genes related to RAD51 and is necessary homologous recombination (HR) repair. A relatively rare G to A polymorphism (rs3218536), located in exon 3 of this gene results in substitution of Arginine by Histidine (Arg188His). METHODS: In the present study, approved by institutional ethical committee, comparison of cytogenetic profile and genotype frequencies of the Arg399Gln polymorphism of XRCC1 and Arg188His polymorphism of XRCC2 gene was done in esophageal cancer patients. Genomic DNA was extracted from blood samples of 100 esophageal cancer patients and 100 controls by standard phenol chloroform method. Standard in-vitro cell culturing technique of 72 hours in RPMI-1640 medium was used for cytogenetic analysis. DNA samples were screened for above polymorphisms by PCR-RFLP method. RESULTS: A large number of cytogenetic abnormalities including loss of chromosomes [2,3,4,5,7,9,11,16,17,19,21,22], gain of chromosomes (4,5,20), polyploidy, translocations (rob(13;21), rob(14;22), rob(13;15)), ring chromosomes [46,XY,r?], dicentric chromosomes, breaks [46,XY,chtb(2q),tas(14:21)], acrocentric associations were found in esophageal cancer patients and these aberrations were found to be increased in patients with advanced stages. In addition, the number of terminal associations and triradials were much more in number in patients as compared to controls.However, no significant association was found between Arg399Gln polymorphism of XRCC1 and Arg188His polymorphism of XRCC2 with the advancement of the esophageal cancer. CONCLUSION: The findings demonstrated the potential utility of combining chromosomal instability and gene expression data to determine if these changes are associated with genetic susceptibility to esophageal cancer or might serve as early detection of cancers. Moreover, the study can provide a platform to evaluate the radiotherapy response and survival rate of the esophageal cancer patients in context of the studied polymorphisms.

2930T

Elucidation of a conserved role for *DIS3* in genome stability maintenance. *K. Milbury[1,2], P. Stirling[1,2,3].* 1) Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, British Columbia, Canada; 2) Genome Science & Technology, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

Chromosome instability (CIN) is characterized by an increased rate of the unequal distribution of DNA between daughter cells. These large changes in chromosome structure or number can occur due to both mitotic defects leading to aneuploidy and DNA damage-induced chromosome rearrangements. Previous large-scale screens for CIN genes in the model organism *Saccharomyces cerevisiae* identified *DIS3*, which codes for a catalytic component of the core RNA exosome complex, as a novel CIN gene. Presumed reduction-of-function mutations in human *DIS3* have been identified in roughly 11% of multiple myeloma (MM) cases. We sought to determine the mechanism of CIN in *DIS3* mutants, and to recapitulate MM-associated point mutations at conserved sites in yeast cells, in order to understand potential connections of emergent CIN to MM. We have found that MM-associated *DIS3* mutations induce DNA:RNA hybrid accumulation and increased rate of CIN, although analysis of DNA damage foci by microscopy revealed no increase in double-strand breaks. Yeast *DIS3* mutants experience growth retardation, and cell cycle analyses indicate that these mutants suffer from delayed progress through mitosis. Microarray analysis of one MM mutant has additionally demonstrated downregulation of cell cycle components, consistent with the potential for mitotic defects, in addition of upregulation of a host of metabolic pathways. Further, genetic interaction profiling by synthetic genetic array indicates MM-associated *DIS3* mutations synthetically interact with rRNA processing proteins, as well as a host of mitotic regulators, particularly those involved in the spindle assembly checkpoint. Together, these results demonstrate extensive phenotypic consequences of MM-associated point mutations in *DIS3*, and support a model for CIN in *DIS3* mutants involving defects in cell cycle processes.

2931F

Loss of KMT2D results in PRC2-mediated transcriptional deregulation. *J.T. Topham, A. Gagliardi, D. Trinh, R.D. Huff, A.J. Mungall, J. Schein, M.A. Marra.* Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver BC.

Lysine (K)-Specific Methyltransferase 2D (*KMT2D*) encodes a methyltransferase responsible for mono-methylation of the fourth lysine of histone 3 (H3K4me1), marking active enhancers. Although *KMT2D* is one of the most frequent targets of all types of somatic mutation in human cancers, the cellular consequences of *KMT2D* mutations and their role in cancers are not fully understood. To understand the effects of *KMT2D* mutation on the transcriptional landscapes of cells, we performed differential expression analysis (DEA) using RNA-sequencing data from isogenic *KMT2D*-knockout and wildtype HEK293A cell lines. To investigate whether these effects recapitulate what occurs in *KMT2D*-mutant tumors, we performed subsequent DEA using RNA-sequencing data from 11 different cancer types in which *KMT2D* loss-of-function mutations were frequent (data from The Cancer Genome Atlas (TCGA)). Through integrative analysis, we identified a distinct subset of genes that are consistently altered upon KMT2D loss across both isogenic cell lines and 11 primary tumor contexts. Moreover, we found that these genes are enriched for targets of polycomb repressive complex 2 (PRC2) and are often involved in regulation of extracellular matrix components and cell adhesion/migration. Thus, our results indicate PRC2-mediated deregulation of genes as a consequence of *KMT2D* mutation relevant to oncogenesis.

2932W

Epigenome-wide reprogramming of enhancer elements during *in vivo* **melanoma progression to brain metastasis.** *D.M. Marzese[1], M.P. Salomon[1], J.I.J. Orozco[1], M.A. Bustos[1], S.K. Huang[1], L. Takeshima[1], N. Nellie[2], S.C. Hsu[2], S. Izraely[3], O. Sagi-Assif[3], D.F. Kelly[4], I.P. Witz[3], D.S.B. Hoon[1,2].* 1) Department of Molecular Oncology, John Wayne Cancer Institute (JWCI), Santa Monica, CA, USA; 2) Sequencing Core Center, John Wayne Cancer Institute, Santa Monica, CA, USA; 3) Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel; 4) Brain Tumor Center, Providence Saint John's Health Center, Santa Monica, CA, USA.

The tendency of melanoma cells to metastasize to the brain represents an important unsolved clinical complication. This high affinity has been postulated to be a consequence of a homing effect since both melanocytes and neural cells derive from the neuroectoderm and thrive in similar tissue microenvironments. To better understand the molecular events driving melanoma progression to brain metastasis (MBM), we performed genomic (Exome-Seq), epigenomic (MethEPIC BeadChip and ChIP-Seq for H3K4me1 and H3K27ac) and transcriptomic (RNA-Seq) analyses on three independent human-to-mouse MBM xenograft models with paired cutaneous and MBM variants. Exome-Seq data revealed a negligible number of mutations (mean SNV=15±4) in MBM compared to the paired cutaneous variants. However, integration of epigenomics data revealed a significant redistribution of both, chromatin modifications and DNA methylation patterns. Subsequently, this data was used to train a Hidden Markov Model to systematically annotate regulatory elements and genomic segments enriched for H3K4me1, H3K27ac, and low regional DNA methylation level were annotated as enhancer elements (EEs). When comparing cutaneous and MBM variants, we observed that 427 EEs were lost and 2,938 EEs were gain during melanoma progression to MBM. To identify active EEs, we evaluated changes on the expression of genes encoded ±1Mb of the annotated EE. Active EEs in MBM variants but absent in cutaneous variants were named 'MBM-specific EEs' (n=604; mean size=1,340±53bp). The regional DNA methylation level was significantly lower at MBM-specific EEs compared with non-EEs. To identify putative target gene promoters for MBM-specific EEs, we used long-range chromatin interaction data (ChIA-PET) of human cell lines generated by the ENCODE project. Gene ontology analyses of coding and non-coding genes potentially associated with MBM-specific EEs revealed enrichment in neurological disorders and brain developmental processes. Regional DNA methylation patterns of MBM-specific EEs were further evaluated by generating DNA methylomes of 56 MBM and 90 non-MBM clinical FFPE specimens. Interestingly, we identified a set of EEs significantly hypomethylated in a large proportion of MBM specimens. Together, our data suggest that melanoma progression to MBM is largely influenced by epigenome-wide remodeling specifically affecting enhancer regions and offer a new alternative to understand the complexity of this clinical complication.

2933T

The impact of *PRDM9* **expression among different cancers and cancer patients.** *A. Ang Houle[1,2], H. Gibling[1,2], M. Agbessi[3], V. Bruat[3], L. Stein[1,2], P. Awadalla[1,2,3,4], Pan-Cancer Whole Genome Consortium.* 1) Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Genome Informatics, Ontario Institute of Cancer Research, Ontario, Canada; 3) Research Center of the Sainte-Justine Mother and University Hospital Center, Montreal, Quebec, Canada; 4) Department of Pediatrics, Universite de Montreal, Montreal, Quebec, Canada.

Homologous recombination is a process allowing for exchange of genetic information between homologous chromosomes, and, when impaired, has been linked to multiple genomic failings, ranging from indels to larger scale aneuploidies. In meiosis, PRDM9 binding sites determine positions of double strand breaks by recruiting SPO11, a topoisomerase-like protein that catalyzes double strand breaks. This initiates meiotic recombination, a process essential to proper chromatid segregation. Allelic variation at PRDM9 has been associated with pediatric acute lymphoblastic leukaemia, suggesting the role of PRDM9 in some cancers. In mitotic cells, homologous recombination is a conservative repair mechanism following genetic damage from double strand breaks introduced by either endogenous or exogenous factors. Cancer genomes often show deficiencies in homologous recombination processes. For this study, we exploit more than 2000 samples available in the Pan-Cancer Whole Genome (PCAWG) program of the International Cancer Genome Consortium (ICGC), which generates and hosts whole genome and transcriptome sequencing of tumoral and matching healthy tissue. PRDM9 is a meiosis-specific gene, yet we observed a significant proportion of samples showing PRDM9 expression. Remarkably, PRDM9 expression is found at significant levels in over 260 tumor samples across multiple cancer subtypes, even after applying stringent filtering corrections to ensure its bona fide expression by minimizing signatures from other gene family members. Expression of PRDM9 appears to have an effect on genomic architecture, specifically regarding somatic structural variations and single nucleotide variants. We observed differences in the distributions of somatic genomic alterations, suggesting an association between PRDM9 expression, genomic breakpoints and somatic deregulation. We also identified variations in transcriptome-wide patterns of expression in tumors that are dependent on PRDM9 expression. This study highlights the role of meiosis-related genes, in particular PRDM9, on the transcriptomic and genomic landscapes of tumors, and provides alternative mechanisms explaining homologous recombination deficiencies in cancers.

2934F

Down-regulation of metastasis-associated genes *MTA1* and *MTA2* by resveratrol leads to alteration of epigenetic repression on tumor suppressor genes in chronic myeloid leukemia cells. *C. Biray Avci, Z. Mutlu, C. Caliskan Kurt, B. Göker Bagca, O. Ozalp, Z. Abbaszade, C. Gündüz.* Medical School Department of Medical Biology, Ege University, Izmir, Turkey.

   Chronic myeloid leukemia (CML) is a hematological disease resulting from generation of BCR/ABL oncogene by philadelphia chromosome. Resveratrol is an important phytoalexin in many plants, has cytotoxic and antileukemic effects. Metastasis-associated gene (MTA1) is a member of cancer progression-related gene family. Also, MTA2 gene belongs to metastasis associated family, MTA1 and MTA2 are highly expressed in tumors. In this study we aimed to evaluate the effects of resveratrol on epigenetic alterations in chronic myeloid leukemia cell line and explore a novel therapeutic target of CML. K562 cells were treated with resveratrol time and dose dependent manner and cytotoxicity was evaluated by using WST-1 assay. The RT-qPCR is used for gene expression analysis. Gene expression levels were evaluated by using RT2 Profiler PCR Array. Significant decreases in MTA1 and MTA2 gene expressions were observed in K562 treated with resveratrol. Resveratrol downregulated MTA1 and MTA2 gene expression levels 2.62 and 3.54 folds according to the control cells, respectively. Our findings showed that Resveratrol down-regulated MTA1 and MTA2 gene expression levels. MTA1 and MTA2 modulate the response of several tumor suppressor genes as p53 and oncogenes like c-myc. Therefore, down-regulation of MTA1 and MTA2 gene expression levels by resveratrol could lead to alteration of epigenetic repression on tumor suppressor genes. Thus they could activate themselves. MTA genes could use as a marker for cancer progression and potential target and Resveratrol might be a new target for CML therapy.

2935W

Epigenome dynamics of cultured breast cancer cells after transient treatment with magnetic nanoparticles. *R.S. Bonadio[1], E.C.O. Lima[2], R.C. Silva[1], F.H. Horst[1], M.J.P. Fonseca[1], J.P.F. Longo[1], R.B. Azevedo[1].* 1) Universidade de Brasilia, Brasilia, Brazil; 2) Universidade Federal de Goiás, Goiânia, Brazil.

   Introduction: Several studies have reported changes in gene expression of cells exposed to nanomaterials, but to date there is no toxicity study at epigenetic level that describes the effects of the 'epigenotoxicity' through successive cell generations induced by nanoparticles. Therefore, it becomes necessary to study these phenomena in order to contribute to the development of more appropriate nanoparticles for biological applications. **Objective:** To evaluate the global DNA methylation profile in MCF-7 cells in culture during and after the treatment with maghemite nanoparticles (MNP). **Methods:** The MNP were synthesized by Fe (II) and Fe (III) coprecipitation method and direct addition of citric acid to stabilize the particles. To identify the IC-20 sub-lethal concentration, the MCF-7 cells were treated with MNP during 24h and it was performed the Trypan Blue and the lactate dehydrogenase (Cytotox 96® Non-Radioactive Cytotoxicity Assay, Promega) assays. In order to investigate the real time cell proliferation through 96h after treatment it was performed an assay in the xCELLigence™ system (Roche / ACEA). The global DNA methylation was performed by colorimetric assay with the MethylFlash™ kit (Epigentek) and Whole Genome Bisulfite Sequencing in the HiSeq 2500 platform (Illumina). **Results:** MNP caused a cytostatic effect on MCF-7 cells treated with 60µgFe/ml during 24h. However, after the 24h treatment it was found that the proliferation rate of treated cells was higher than untreated cells. This phenotype could be associated with the DNA methylation pattern that we observed. During the treatment, cells undergo hipomethylation and after the treatment cells restore their DNA methylation status. The whole genome bisulfite sequencing is still being analysed, but the preliminary data indicates that the differentially methylated regions (DMR) among treated and control samples occur mostly on the intergenic regions, introns and CpG islands. **Discussion:** The influence of nanomaterials in the epigenetic status of a biological system is still poorly understood. Thus, nanomaterials have a risk in biological applications, even when administered in non-toxic concentrations considered by conventional techniques. This is because their effects can be extended to multiple generations, even during transient exposure. **Conclusion:** The MNP promote significant changes in global DNA methylation in MCF-7 cells and this event can be exploited for future biomedical applications.

2936T

Resistance to BET inhibitors in lung adenocarcinoma is mediated through a MYC independent mechanism. *J. Calder, W. Lockwood.* Department of Pathology, BC Cancer Research Center, Vancouver, BC, Canada.

JQ1 is an inhibitor of the bromodomain and extraterminal (BET) family proteins, which function as important reader molecules of acetylated histones and recruit transcriptional activators to specific promoter sites. The down-regulation of c-MYC has been linked to JQ1 inhibition in many sensitive cancer cell lines, while reactivation of c-MYC expression has been shown to induce JQ1 resistance. Our lab has shown that lung adenocarcima (LAC) cells are inhibited by JQ1 through a mechanism independent of c-MYC down-regulation, identifying FOSL1 as a possible target in LAC cells. This suggests that the epigenetic landscape of cells from different origins and differentiation states influences response to JQ1. This study aims to identify potential mechanisms regulating resistance to JQ1 in LAC in order to determine if the epigenome affects different processes in different cancer types. LAC cell lines sensitive to JQ1 treatment, H23 and H1975, were passaged with increasing concentrations of JQ1 until the cells were resistant to high doses of the drug. Expression profiles were generated for parental and resistant cells and genes differentially expressed between the states for each cell line were identified and compared across both H23 and H1975 to identify candidate genes. Gene Set Enrichment Analysis (GSEA) was used to identify potential cell pathways and programs driving JQ1 resistance in LAC. Protein expression was evaluated through Western Blot analysis to confirm gene changes associated with resistance. Initial morphological and western blot analysis showed resistant H1975 cells underwent EMT transition with significant decrease in E-cadherin and increase in Vimentin. Initial analysis of differentially expressed genes between the parental and resistant pairs identified 101 significantly differentially expressed genes (corrected p-value <0.005) common between the H1975 and H23 lines; however, MYC was not significantly altered. Preliminary GSEA results identified a few enriched pathways including Wnt signaling pathway and ERK/MAPK targets. The discovery and optimization of small-molecule inhibitors of epigenetic targets is a major focus of current biomedical research. We determined that LAC cells, unlike those from other cancers, develop JQ1 resistance through mechanisms independent of c-MYC, suggesting the epigenomic landscape of a cell can influence both sensitivity and resistance to BET inhibitors.

2937F

The role of the immune system in field cancerization: Progressive inactivation of immune related pathways in esophageal cancer. *L. Cordeiro, L. Pinto.* INCA, Rio de Janeiro, RJ, Brazil.

The concept of field cancerization was introduced over 60 years ago and its presence has been described in many cancer types. It is believed that progressive molecular changes affect cellular functioning spatially, giving rise to an expanding field of heterogeneously altered cells. Unhindered, such progression may lead to cancer. We present *in silico* evidence that field cancerization occurs in the most prevalent subtype of esophageal cancer. Esophageal cancer is the 8th most incident cancer globally and the 6th leading cancer in mortality (5% of all cancer deaths). Esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC) are the main histological types of esophageal cancer, the latter responsible for over 85% of cases. Prognosis is grim: the 5-year survival rate ranges from 5% to 15%. Despite being one of the most lethal cancers, little is known about its carcinogenic mechanisms. We systematically mined public gene expression data and built a unique dataset of over 400 samples of ESCC, tumor-adjacent tissue and esophageal cells from healthy individuals. We analyzed the progression of changes in global gene expression from healthy to tumor-adjacent to ESCC samples and found 17 consistently differentially expressed genes from "healthy" to "tumor" state. Five genes were consistently overexpressed and four of them – *RIPK4*, *FGFR3*, *IRF6*, *EPCAM* – are directly associated with deregulation of epithelial proliferation pathways. Twelve genes were consistently underexpressed and nine of them – *HLA-DPA1*, *IGHM*, *ATP6V1A*, *FAS*, *HLA-G*, *HLA-DRA*, *IGLC1*, *HLA-B*, *POU2AF1* – code for crucial elements in both humoral and cell-mediated immune responses. Network and pathway enrichment analysis on differentially expressed genes in each state transition show a distinctive pattern of changes in global expression. In the first transition, from "healthy" to "tumor-adjacent", several immune-related pathways are inhibited causing grave deregulation in antigen processing and presentation processes as well as decreased immune-related signaling. In the second transition, from "tumor-adjacent" to "tumor", the inhibition of these immune-related pathways is sustained, with respect to the "healthy" state, and epithelial–mesenchymal transition pathways became deregulated. Taken together, these findings suggest the existence of field cancerization in esophageal squamous cell carcinoma and highlight a possible role of global immune system inhibition on its carcinogenic mechanism.

2938W

Integrated gene-level analysis of molecular features in high-grade serous ovarian cancer. *J.M. Cunningham[1], M.A. Earp[2], S.M. Armasu[2], B.L. Fridley[3], K. Kalli[2], M.C. Larson[2], R.A. Vierkant[2], C. Wang[2], S.J. Winham[2], E.L. Goode[2].* 1) Dept Lab Med & Pathology, Mayo Clinic & Fndn, Rochester, MN; 2) Dept. Health Sciences Res., Mayo Clinic & Fndn, Rochester, MN; 3) Biostatistics and Informatics Shared Resource, University of Kansas Medical Center, Kansas City, KS.

High grade serous ovarian cancer (HGSOC) is a complex disease in which initiation and progression have been associated with gene mutation, DNA methylation changes, genetic variation, and environmental processes. To reveal novel genomic mechanisms underlying HGSOC development, we conducted an integrated analysis of 54 HGSOC tumors to determine how tumor gene expression (RNA-seq) may be influenced by germline genetic variation (Illumina platforms imputed to 15.5M variants using 1000 Genomes) and tumor DNA methylation (Illumina Meth 450K) while adjusting for somatic copy number (Illumina OncoArray). Elastic net (ENET) penalized regression methods were used for model fit, and a permutation-based step-down multiple testing procedure (MaxT) was used to derive adjusted *p*-values. Each of the ~20,000 Ensembl IDs (hg19:GTF75) corresponding to protein-coding genes was analyzed using a window 500kb up- and downstream of the gene start and stop. Three nested models were compared: a) genotype and DNA methylation (full model), b) genotype only, and c) DNA methylation only. No model for gene expression was significantly different from the null after adjustment for multiple testing (MT). Based on a p<0.01 (unadjusted for MT) for the full, genotype only, and methylation only models: 161, 135, and 537 genes, respectively, were found to have a model for expression significantly different from the null. No genes were overlapping in all three lists, 13 overlapped in the full vs genotype-only comparison, 22 overlapped in the full vs methylation only comparison, and 6 overlapped in the methylation only vs genotype only comparison. Gene set enrichment analysis (GSEA) of these gene-lists reveals expected and novel signatures. For example, as expected, the gene list derived from the tumor DNA methylation only model analyses is most enriched for genes known to be altered in cancer (false discovery rate [FDR] q-value=1.4×10^{-28}), and involved immune and inflammatory responses (FDR q-value=1.4×10^{-43}), suggesting this method is capable of detecting true biological signals. Genes identified by the full and genotype only model were most enriched for genes involved in signal transduction (FDR q-value= 6.7×10^{-7}) and with promoter regions containing the binding motif (~2kb from transcription start site) for SP1 (FDR q-value=9.2×10^{-7}), respectively. All genes with an ENET model for expression significant at p<0.05 (unadjusted for MT) will be tested for replication using TCGA HGSOC tumor data.

2939T

The effects of malign breast tissue stromal cells in proliferating of MDA-MB-231 cancer cells and miRNA relation. *I. Değirmenci[1], O. Saglam[1,2], Z.S. Unal[2], G. Posteki[3], E. Karaoz[2], H.V. Gunes[1], Z. Utkan[3].* 1) Eskisehir Osmangazi University, Medical Faculty, Department of Medical Biology, Eskisehir/Turkey; 2) Kocaeli University, Center for Stem Cell and Gene Therapies Research and Practice, Department of Stem Cell, Kocaeli/Turkey; 3) Kocaeli University, Medical Faculty, Department of General Surgery, Kocaeli/Turkey.

Purpose: MicroRNAs (miRNAs) are endogenous non-coding RNAs which regulate gene and protein expression through the RNA interference (RNAi) mechanism. Oncogene or tumor suppressor miRNAs play an important role in tumorigenesis. It has been reported that stromal cells in tumor microenvironment gained similar features of tumor-associated cells. Detailed analysis of miRNA functions should be investigated to discover novel markers and to determine the prognosis through better understanding of molecular basis of cancer concept, and to develop new treatment approach consequently. Thus in the current study, we aimed to investigate the effects of co-culture of cancer cells and stromal cells, isolated from normal and malign breast tissues, on each others, and the possible effects of miRNAs in these cross-effects. **Methods:** Subsequently after isolation of stromal cells from healty and malign breast tissue, appropriate culture condition will be maintained. Characterized stromal cells that are derived from normal and malign breast tissue will co-cultured with MDA-MB-231 cancer cell line in the presence of semi-permeable membranes. After the co-culture, proliferation capacity of the experimental groups were evaluated with WST-1 assay. Expression levels of breast cancer specific miRNAs and related genes will be evaluated by real-time PCR. After the co-culture, ELISA test will be performed to determine some of cytokine and chemokine levels. **Results:** As a result of experiments we found that increased proliferation capacity of the cancer cells after co-culture with malignant stromal cells. Also we found that microenvironment is important in the formation of cancer in support of the changes in the expression levels of oncogenic and tumor suppressor miRNA and their targeted genes after the co-culture with malignant stromal cells. As a conclusion, the possible roles of miRNAs on the alterations in cancer cells by paracrine mechanisms of both normal and malign breast tissues stromal cells are attempt to reveal. In the result of the performed studies with this aim, specific gene expressions of related pathways are going to be detected correlated with miRNA changes, and the effects of the tumor microenvironment in tumorogenesis will be revealed in detail.

2940F

Identification of 102 new candidate target genes for prostate cancer risk-SNPs utilizing a normal prostate tissue eQTL dataset. *M.S. DeRycke[1], M.C. Larson[1], Z.C. Fogarty[1], S.M. Riska[1], S. Baheti[1], L.S. Tillmans[2], S.K. McDonnell[1], D.J. Schaid[1], S.N. Thibodeau[2].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Dept Lab Med & Pathology, Mayo Clinic, Rochester, MN.

Multiple risk SNPs for prostate cancer have been identified, however determining which genes the SNPs influence is difficult. Previously we analyzed 146 risk SNPs in 100 regions and identified 51 regions with significant eQTL signals associated with 88 genes for prostate cancer risk while creating a normal prostate tissue eQTL database. Expanding on our earlier work, we have reanalyzed the data increasing the number of risk SNPs analyzed to 202 in 107 regions and using ENSEMBL gene annotation. All risk SNPs, and those SNPs within +/-100Kb and in LD (r^2>0.5) (n=8,073 SNPs) with the PC-risk SNP, were tested against all genes located within +/-1.1Mb from the SNPs (n=3,225 genes). Of the genes tested, 187 demonstrated a gene-level false discovery rate (FDR) <=0.01 (p<1.69e-4). 102 genes not previously reported were identified. Additionally, 85 of the original 88 genes were also reconfirmed; the three genes not reconfirmed included TUBA1B; (FDR=0.0026) and two uncharacterized genes not found in the ENSEMBL gene annotation (LOC284578 and LOC284581). Further studies, including utilizing public datasets to map the significant SNPs to prostate-specific known and predicted regulatory elements, including promoters, enhancers, transcription factor binding sites, transcription start and stop sites, are currently in progress.

2941W

Secreted miRNA: A dual role in promoting oral carcinogenesis. *C.T.D. Dickman[1,2], J. Lawson[1], K. Bennewith[1], C. Garnis[1,3].* 1) BC Cancer Research Center, Vancouver, British Columbia, Canada; 2) Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Surgery, University of British Columbia, Vancouver, British Columbia, Canada.

There has become a growing interest in the role of exosomes in cancer biology, particularly due to their involvement in communication between cancer and stromal cells. Exosomes are small membrane bound vesicles that are known to contain a wide variety of proteins and nucleotides such as miRNAs. Exosomal contents can be a source of biomarkers and their transfer system could be exploited for drug delivery. As the miRNA content of exosomes is non-randomly packaged we hypothesize that miRNAs are selectively retained or excreted into exosomes in oral cancer cells to promote growth within the cell and to create a favorable microenvironment for tumor growth. By examining a panel of oral cancer and dysplasia cell lines we identified several miRNAs under selection for exosomal packaging and release. When exosome release is inhibited by the knockdown of the protein Rab27A (a protein involved in exosomal trafficking) we were able to increase the intracellular concentration of excreted miRNAs and we were able to determine that increased intracellular miR-142-3p targeting TGFBR1 and thereby decreasing cancer cell growth. Over-expression of miR-142-3p in exosomes was shown to increase the tumor supporting effects of endothelial cells with a subsequent decrease in TGFBR1. Mouse tumor injections showed that cells over-expressing miR-142 grew at a slower rate but were more vascular and as a result less hypoxic. Together these data suggest that miR-142-3p excretion plays a dual role in tumorigenesis: the removal of a tumor suppressive effect within the cell and the creation of a pro-carcinogenic niche in the stromal cells.

2942T

Alterations to the *ELF3* oncogene circumvent upstream regulation in lung cancer. *K.S.S. Enfield[1], D.A. Rowbotham[1], S. Rahmati[2], C. Anderson[1], J.Y. Kennett[1], E.A. Marshall[1], R. Chari[3], D.D. Becker-Santos[1], K. Ng[1], C.E. MacAulay[1], S. Lam[1], W.W. Lockwood[1], I. Jurisica[2], W.L. Lam[1].* 1) Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada; 2) Princess Margaret Hospital, Toronto, Ontario, Canada; 3) Harvard Medical School, Boston, MA, USA.

   Introduction: The transcription factor ELF3 has been implicated as both an oncogene and a tumour suppressor. In lung cancer its overexpression is in part due to SMAD4 inactivation and upregulation of ERBB2 and PI3K signaling. Here we perform a large scale multi-omics investigation to determine the genetic mechanisms underlying ELF3 overexpression in lung adenocarcinoma (LUAD) as well as the tissue-specific interactions that may explain its divergent functions. **Methods:** Multi-omic data was analyzed from the BCCRC (n=83), TGCA (n=420), and four transgeneic murine models of LUAD tumourigenesis. Associations of *ELF3* with patient survival were assessed by log-rank method. Lentiviral delivery of shRNA knock-down and overexpression vectors was performed in cell lines; cell proliferation, cell viability and anchorage independent growth was assessed. Physical protein-protein interaction (PPI) networks of 65 cancer and normal expression datasets across 16 tissues were interrogated to investigate tissue-specificity of ELF3 interactions. Pathway analysis was performed using pathDIP. **Results:** *ELF3* overexpression was most frequently associated with DNA copy number gain, focal amplification, and hypomethylation at the *ELF3* locus observed in 82% of LUAD, rather than mutation of upstream regulators. LUAD with *ELF3* DNA-level alteration also harbored known driver mutations, with a statistical enrichment of *KRAS* and exclusion of *MET*. High *ELF3* expression was observed in all transgenic models and associated with poor survival. ELF3 regulated cell proliferation, viability, and the ability to form colonies in soft agar. PPI networks indicated highly tissue and disease-specific interactions. Lung tumour-specific ELF3 networks included gain of NOTCH and TGFβ, and loss of TNFα and MAPK signaling. **Conclusions:** While ELF3 is downstream of several key lung cancer signalling pathways, upstream regulation is often bypassed due to frequent alteration to the *ELF3* locus. Our cell models show ELF3 regulates cancer phenotypes and provide biological evidence for its associations with survival. The biological significance of *ELF3* alteration in LUAD driven by different mutations remains to be deciphered. The tissue-specific PPI networks observed suggest distinct roles of ELF3 across cancers from diverse tissue origins. Lung cancer specific PPI networks reveal ELF3 function in this deadly disease. Due to the nearly ubiquitous overexpression, ELF3 represents a promising therapeutic target in LUAD.

2943F

A combined miRNA-piRNA signature derived from the DLK1-DIO3 imprinted locus predicts lung adenocarcinoma patient outcome. *J.R. Enterina, K.S.S. Enfield, V.D. Martinez, W.L. Lam.* Integrative Oncology, BC Cancer Agency, Vancouver, BC, Canada.

   Introduction: Deregulation of small RNAs at the imprinted DLK1-DIO3 locus has been linked to lung adenocarcinoma (LUAD) patient outcome. While the contribution of microRNAs (miRNA) is established, the role of Piwi-interacting RNAs (piRNA), small RNAs involved in epigenetic regulation, is unknown. In this study, we quantified expression of piRNAs and miRNAs mapping to this locus in two independent cohorts of LUAD and assessed the ability of a combined miRNA/piRNA signature to improve patient outcome stratification. **Methods:** Expression levels (RPKM) for miRNA and piRNA were determined from small RNA sequencing experiments from two cohorts (TCGA n=156; BCCA n=75). Associations with patient overall survival were calculated by inputting miRNA and piRNA expression combinations into a Cox proportional hazard model. Risk scores were calculated by multiplying the expression value for a given gene by its hazard coefficient, and then summing the transformed gene expression values per model. Risk scores were ranked and divided into tertiles of high, medium, and low risk. Risk group survival curves were compared by the log-rank (LR) method. **Results:** A total of 138 piRNAs mapped to the imprinted locus; only 7 were expressed. The survival model was improved in TCGA cohort following the addition of four of the seven piRNAs (p=1.725e-04) to the original miRNA-based model (p=6.053e-04). By LR analysis of risk groups, the miRNA/piRNA model better predicted outcome (p=0.0038) than the miRNA model (p=0.0506). In the long-term follow up (>8 years) BCCA cohort, the miRNA/piRNA model (p=0.04915) out-performed the miRNA model (p=0.4493). Strikingly, LR survival analysis showed only the combined miRNA/piRNA signature effectively separated high and low risk groups (p=0.0019). The median survival for the low risk group was undefined up to 100 months, vs. 45.3 months in the high risk group. **Conclusions:** We find the combination of miRNA and piRNA expression derived from the DLK1-DIO3 locus produces a superior stratification of patient outcome than either metric alone. While the contribution of miRNAs to patient risk stratification is established, the improved model performance derived from the addition of piRNAs indicates another level of gene regulation (piRNAs mainly target DNA sequences). Model performance is optimal when these two small RNA species are considered simultaneously; suggesting coordinated biological effects as a result of locus deregulation. JRE & KSSE contributed equally.

2944W

Hypoxia-regulates piwi-interacting RNAs in human tumor. *N. Firmino, E.A. Marshall, B.C. Minatel, W.L. Lam, K.L. Bennewith, V.D. Martinez.* Department of Integrative Oncology, BC Cancer Research Center, Vancouver, Canada.

   Background: Tumor hypoxia promotes cancer aggression by inducing the expression of genes related to angiogenesis, metastasis, and cell survival. MicroRNA expression patterns have been shown to change under hypoxic conditions suggesting potential involvement in post-transcriptional regulation of hypoxia-inducible genes. PIWI-interacting RNAs (piRNAs) is another class of small non-coding RNAs. piRNAs facilitate gene silencing by recruitment of protein complex components to sequence specific gene targets. piRNAs are primarily expressed in germ cells. However, we recently showed that a subset of piRNAs are expressed in somatic tissues, and might be linked with epigenetic mechanisms in cancer development. In this study, we determine if hypoxia-regulates piRNA expression in human tumors. **Methods:** We classified human tumors from The Cancer Genome Atlas (TCGA) Research Network according to their oxygenation state using expression signatures derived from established hypoxia-associated gene expression changes. We analyzed 3,020 human tumors derived from cervix (N=306), esophagus (N=196), head and neck (N=522), kidney (N=534), lung (N=1,018), ovaries (N=265) and pancreas (N=179). We extracted piRNA sequences from miRNA sequencing libraries using a custom pipeline and quantified high-quality reads that mapped to the piRNA annotation database derived from piRNABank (http://pirnabank.ibab.ac.in/). **Results:** Our analysis revealed that piRNA expression is selectively deregulated by hypoxia in human tumors. We deduced expression levels from 32,237 human piRNAs on a per tumor basis, and identified the most robust changes in piRNA expression between hypoxic and non-hypoxic groups. We included piRNAs in our analysis that: 1) have a median expression ≥ 10 RPKM in at least one of the groups (hypoxic and/or normoxic), and 2) a minimum of 2-fold variation in median RPKM expression values. Considering piRNAs meeting these requirements, we performed the comparison using the non-parametric Mann-Whitney U test. We found that 40 piRNA were consistently deregulated across human tissues. Of those, 36 were upregulated, while 4 were downregulated. In-vitro tumor cell models recapitulated hypoxia-regulated piRNA expression patterns. **Conclusion:** Our data provide the first indication that piRNAs can be up-regulated by hypoxia in patient samples and in cell cultures, and suggest that hypoxia-induced piRNAs may represent an undiscovered mechanism of gene repression in hypoxic tumor cells.

2945T

Methylome-wide profiling of DNA from paired diagnostic prostate tumour and adjacent benign tissue. *L.M. FitzGerald[1,2], E.M. Wong[3], C. Jung[4], J. Joo[3], J. Pedersen[5], G. Severi[6], G.G. Giles[1], M.C. Southey[3].* 1) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia; 2) Cancer, Genetics and Immunology, Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia; 3) Genetic Epidemiology Laboratory, University of Melbourne, Melbourne, VIC, Australia; 4) Victorian Life Sciences Computation Initiative, University of Melbourne, Melbourne, VIC, Australia; 5) TissuPath, Mount Waverley, VIC, Australia; 6) Human Genetics Foundation Torino, Turin, Italy.

   A number of studies have compared genome-wide DNA methylation profiles in prostate tumor and adjacent benign tissue with the aim of identifying diagnostic biomarkers or therapeutic targets. Several differentially methylated CpG (dmCpG) sites have been shown to differentiate tumor from benign tissue but these studies have been conducted using radical prostatectomy (RP) samples, thus their relevance at the time of diagnosis is not known. Here, we aimed to investigate differential methylation in diagnostic tumor and benign prostate tissue samples to identify biomarkers relevant at the time of diagnosis. Our study included 138 DNA samples from FFPE diagnostic trans-rectal ultra sound or trans-urethral resections of the prostate, including 33 paired prostate tumor and adjacent histologically benign tissue samples. DNA methylation was measured using the Illumina Infinium HM450K BeadChip. Linear models to account for the paired study design and an adjusted Bonferroni p-value of 0.01 were used to evaluate differential methylation. After quality control measures and ComBat normalisation, analysis of 31 tumor-benign pairs identified 32,104 significantly dmCpG sites, 2,146 of which had a mean methylation difference (delta-beta) of >0.4. The majority (74.7%) of dmCpG sites were hypermethylated in tumor samples, including 98.1% of dmCpG sites with strong methylation changes (delta-beta >0.4). Using methylation patterns at the dmCpG sites, tumor samples were distinguished from benign samples, including 67 unpaired tumor and benign samples. In addition to confirming several previously identified dmCpG sites from RP studies, our study of diagnostic tissue identified dmCpG sites that had not been previously described in this context. Preliminary gene expression data from the same diagnostic tissues demonstrates that 337 of 550 differentially expressed genes (FDR<0.01) are associated with dmCpGs, the majority of which (78%) are down-regulated in tumor. Our study used diagnostic FFPE samples and applied a genome-wide strategy to identify dmCpG sites that distinguish prostate tumor from adjacent benign tissue. In addition to confirming previously identified dmCpGs from RP studies, we report additional dmCpG sites that appear to be specific to diagnostic tumor tissue. This study provides insight into some of the DNA methylation changes that occur in the early stages of this disease and could potentially lead to improved diagnostic biomarkers and novel therapeutic targets.

2946F

Repair DNA pattern expression and ribosomal genes are associated with adult medulloblastoma subtype. *A.M. Fontes[1], F.B. Barbosa[1], A.L. Simoes[1], C.E.V. Wiezel[1], F.S. Ramalho[2], H.R. Machado[3], S.L. Gerson[4], K.J. Abraham[5].* 1) Genetics, University of Sao Paulo, Ribeirao Preto, SP, Ribeirao Preto, Brazil; 2) Department of Pathology and Legal Medicine, FMRP-USP, Brazil; 3) .Department of Surgery and Anatomy, FMRP-USP, Brazil; 4) Case Comprehensive Cancer Center, School of Medicine, CWRU, USA; 5) Estácio- Uniseb, Ribeirao Preto, Brazil.

   Introduction: Medulloblastoma (MB) represents a genetic and epigenetic heterogeneous group of neuroepithelial primary tumors arising from the cerebellum. MB can be classified in four major subtypes: WNT, SHH, Group C and Group D. With current treatment, some MB patients still do not respond to treatment suggesting heterogeneity whithin these subtypes. To better understand these tumors we investigated five groups of genes associated with cancer and other phenotypes. **Methodology:** In this study we analyzed 4 frozen tumor samples from 4 adult MB patients and 3 control cerebellum tissues. Two color microarray-based gene expression profiling and human comparative genomic hybridization (CGH) (2x 400K) were performed. Bioconductor and Agilent Cytogenomics were used for the analysis. **Results:** We performed hierarchical clustering with 22 transcripts and observed the existence of three subtypes among the specimens analyzed (WNT, SHH and Group D). Next, we analyzed 294 genes, which were categorized in five groups (1. Classic oncogenes, classic tumor suppressor and apoptosis related genes; 2. Repair DNA genes (including POLB and XRCC5); 3. Epigenetic regulator genes; 4. Hox genes including long noncoding associated-Hox and 5. Lysosomal and Ribosomal-genes). We found these tumors have altered expression in 46-62% of these genes with upregulation much more frequent than downregulation. Compared with control samples, *POLB* mRNA is up to 164 times more expressed in MB-SHH, and 44 times less expressed in Group D. *XRCC5* mRNA is 143 times lower regulated in MB-SHH and 126 times up-regulated in MB-WNT. *RPL15* mRNA is up-regulated 140-390 times in all MB-tissues compared with control tissues and *RPS20* mRNA is downregulated 61-156 times. We also observed the DNA CNVs in these tumors and found 3-26 CNVs among these 294 genes. The MB-SHH sample has 22 CNVs, mostly copy number loss, in DNA repair or epigenetic regulator genes. The MB-Group D sample has 26 CNVs mostly copy number gain, in HOX genes or epigenetic regulator genes. The epigenetic genes and long non-coding RNAs show both copy number gain and loss. In some repair genes downregulation and copy number loss were jointly observed while in oncogenes upregulation was observed along with copy number gain. **Conclusion:** This study analyzed the correlation between copy number variation and up or down expression in several cancer-associated genes which can provide useful insights for biomarkers and therapeutic intervention. .

2947W

ETV6/TEL plays a key role in KAP1/TRIM28-dependent DNA damage response in acute lymphoblastic leukemia. *R. Gioia[1], C. Drullion[1], S. Drouin[1], S. Langlois[1], J. Larose[1], C. Jimenez-Cortez[1], D. Sinnett[1,2].* 1) Division of Hematology-Oncology, CHU Sainte-Justine Research Center, 3125 chemin de la Côte-Sainte-Catherine, Montréal, Québec, H3T 1C4, Canada; 2) Department of Pediatry, University of Montreal, 2900 Édouard-Montpetit, Montreal, QC, H3T 1J4, Canada.

   Pre-B acute lymphoblastic leukemia (ALL) represents ~25% of all pediatric cancer cases, which makes it the most common hematological malignancy and the leading cause of disease-related death in children. The most frequent genetic alteration associated with pre-B ALL is the t(12;21) translocation, which results in the formation of the ETV6-AML1 (TEL/RUNX1) chimera and the frequent deletion of the residual ETV6 allele, which leads to complete ETV6 inactivation. ETV6 had been previously described only as a ubiquitously expressed transcriptional repressor. We showed here that ETV6 plays an important role in genotoxic stress cytoprotection by repressing DNA damage response (DDR). Moreover, mass spectrometry and co-immunoprecipitation data indicated that KAP1/TRIM28 is an ETV6 physical interaction partner implicated in this novel function. Indeed, we have characterized through gain- and loss-of function assays ETV6 as a negative regulator of apoptosis and DDR and KAP1 as a critical regulator of ETV6 activity. Finally, we have shown that ATM-dependant phosphorylation of KAP1-S824 modulates ETV6's DDR activity. This study describes a novel role for the ETV6 transcriptional repressor in DNA damage response and apoptosis regulation and suggests new leads as to its role in childhood leukemogenesis..

2948T

Genome-wide differential methylation profiles in circulating cell free DNA for pancreatic cancer early detection and prognosis. *R.J. Hung[1], G. Fehringer[1], R. Singhania[2], S.Y. Shen[2], P. Zuzarte[3], D. Chadwick[2], A. Borgida[1], S. Gallinger[1], D.D. De Carvalho[2].* 1) Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada; 2) Princess Margaret Cancer Centre, University Health Network, Toronto, Canada; 3) Genome Technologies, Ontario Institute for Cancer Research, Toronto, Canada.

   Background: Around 90% of pancreatic cancer patients die within the first year, underlining the importance of early detection and the need for prognostic marker. Methylation aberrations have substantial impact on tumorigenesis and disease progression, but biomarkers that require tumor tissues have limited clinical value for early detection and unresectable cases. On the other hand, circulating cell free DNA (cfDNA) is ideal for biomarker discovery as it was shown to reflect genetic alterations in tumor tissue and can be obtained minimally invasively. Thus we investigate whether aberrant methylation related to pancreatic tumors can be detected in cfDNA, and if so, the clinical utility for early detection and prognosis. **Methods**: To overcome the low input and fragmented nature of cfDNA, we developed a new method specifically for cfDNA genome-wide methylation sequencing with ultra-low input . Based on tumor-normal pairs and cfDNA from a total of 157 individuals (71 cases and 86 controls), we identified differential methylated regions (DMRs) and linked back to epigenetic changes in the tumors. We conducted motif analysis to identify tissues of origin. We assessed the predictive performance of the DMRs based on cross-validation and penalized regression with Lasso penalty. We applied multivariate Cox regression to identify prognostic markers, and network analysis to identify core survival gene networks. **Results:** We identified 12085 DMRs that distinguish pancreatic cancer patients from healthy individuals. The cfDNA methylation profile in patients significantly resembled that of matched resected tumors from the same patients (p= 3.12^{-15} for hypermethylation; p=2.67^{-109} for hypomethylation). Based on 2485pancreatic cancer-specific DMRs (qFDR<0.05) detectable in cfDNA, we observed an increase of area under receiver operating characteristic curve (AUC) of additional 0.39 when comparing to a model with only personal health history and clinical variables, from AUC of 0.59 to 0.98 in the *validation set*. In addition, 59 DMRs were found to be associated with pancreatic cancer survival (qFDR<0.05), and they were located mainly in genes related to apoptosis and tumorigenesis network. **Conclusions**: We demonstrated that differential methylation patterns specific to pancreatic tumors can indeed be detected in the cfDNA, and further identified potential DMRs for pancreatic cancer early detection and prognosis. These results will be further validated in a larger cohort of patients.

2949F

Molecular characterization of the role of RUNX1 in Notch signaling in T-cell Acute Lymphoblastic Leukemia (T-ALL). *R. Islam[1,2], C. Jenkins[3], M. Bilenky[5], A. Carles[2,4], L. Li[2], A. Lorzadeh[2], V. Giambra[3], S. Lam[3], C. Hoofd[3], M. Belmonte[3], X. Wang[3], A. Weng[3,6], M. Hirst[2,4,5].* 1) Bioinformatics, University of British Columbia, Vancouver, BC, Canada; 2) Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada; 3) Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada; 4) Department of Microbiology & Immunology, University of British Columbia, Vancouver, BC, Canada; 5) Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada; 6) Department of Pathology & Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

   T-cell acute lymphoblastic leukemia (T-ALL) is a hematopoietic malignancy driven by oncogenic activation of Notch signaling. Runt-related transcription factor 1 (RUNX1) is hypothesized to participate in supporting Notch signaling, which is well characterized as an oncogenic pathway in T-cell leukemia. In support of this hypothesis we find that human T-ALL patient-derived samples and cell lines (e.g., KOPT-K1) are sensitive to RUNX1 depletion mediated by lentiviral shRNAs and pharmacologic Notch pathway inhibition. In order to dissect the molecular mechanism(s) underlying these phenotypes, we have performed mRNA-seq and ChIP-seq against a panel of histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3 and H3K9me3) for samples that have either been depleted of RUNX1 (shRNA) or NOTCH1 (pharmacologic inhibition). RNA-seq analysis revealed that 14 genes (e.g., CR2, DTX1, EGR1, FGR, HES4, MYC, NOTCH3 etc.) with NOTCH1 and RUNX1 binding site at proximal region (+-5Kb of TSS) are down regulated in absence of either NOTCH1 or, RUNX1. We found NOTCH1 and RUNX1 sites localized to H3K27ac and H3K4me3 enriched promoters of those 14 genes, whereas these regions lose enrichment when NOTCH1 or, RUNX1 is turned off. These NOTCH1 and RUNX1 co-regulated sites also intersect with other co-activators e.g., P300, ETS1, BRD4, MED1, GABPA, RBPJ (source: Hongfang W. et al. PNAS,2014;111(2):705-10). Kitabayashi I. et al. EMBO Journal, 1998,17,2994-3004, showed that RUNX1 interacts with p300 and overexpression of p300 stimulates RUNX☐dependent transcription to induce myeloid cell differentiation. We observed a global loss of H3K27ac upon RUNX1 silencing which is possibly the absence of RUNX1 halts the recruitment acetylation through P300. This result suggests the potential role of RUNX1 in recruitment of histone acetyl transferases in T-ALL. Simultaneously, there is gain of H3K27me3 in proximal regions and ~30% active enhancers become poised in absence of RUNX1. The loss of H3K27ac and gain of H3K27me3 indicates a broad chromatin re-structuring in absence of RUNX1. KEGG pathway enrichment analysis shows RUNX1 target genes are involved in cell cycle, receptor sinaling and metabolic pathways. Consequently, using BrdU incorporation assay we observed a proportion of cells stopped replication and blocked cell cycle. Together these results we suggest an essential role for RUNX1 in regulating the expression of genes involved in maintaining T-ALL through recruiting H3K27ac.

2950W

Epigenetic regulation of *SMPD3* **alters cell stress response in human oral cancer cells.** *J. Jabalee, R. Towle, C. Garnis.* Dept of Integrative Oncology, University of British Columbia, Vancouver, Canada.

Oral squamous cell carcinoma (OSCC) displays a dismal five-year survival rate of ~50%, one of the worst for all cancer types. To significantly improve survival rates we must better understand the molecular events driving the disease. To identify genes involved in OSCC progression, we performed genome-wide DNA methylation profiling of patient-derived normal, pre-malignant, and tumor tissue and identified the *SMPD3* promoter as a site of frequent hypermethylation in pre-malignant (6/10 patients) and tumor (10/10 patients) tissues. Hypermethylation correlates with a decrease in *SMPD3* expression in 4/6 pre-malignant samples and 6/10 tumor samples. In cancer, gene silencing via promoter hypermethylation is a common feature of tumor suppressors. The product of the *SMPD3* gene, neutral sphingomyelinase 2 (NSMASE2), degrades sphingomyelin to ceramide, a key lipid second messenger involved in diverse processes such as apoptosis, growth arrest, and cell signaling via exosomes. Tumor-derived exosomes contain various signaling molecules, including microRNAs, which can influence recipient cells and promote tumorigenesis. We find that *SMPD3* is methylated and not expressed in 5/6 oral cancer cell lines examined; however, the role of *SMPD3* in cancer is not well understood. In order to determine its functional role, we overexpressed *SMPD3* using lentiviral vectors in the oral carcinoma cell lines SCC-25 and Cal27. Our preliminary results suggest that overexpression decreases growth rate compared to control cells. This effect has been observed in other cancer types, including breast, bone, and liver. *SMPD3* is thought to mediate the response to various stress conditions, including serum starvation and radiation, via the generation of proapoptotic ceramide. Interestingly, *SMPD3* overexpression protected Cal27's against serum starvation, in contrast to recent results in mouse fibroblasts. Furthermore, *SMPD3* appears to sensitize Cal27 cells to radiation treatment, in agreement with previous studies. We are currently repeating these experiments in other cell lines. In addition, we are investigating the role of *SMPD3* in regulating microRNA packaging into exosomes under stress conditions Our results suggest that *SMPD3* acts as a tumor suppressor and alters cancer cells' response to stress. Further research is required to determine the effectiveness of *SMPD3* expression level in guiding clinical decision making, particularly with respect to radiation therapy.

2951T

Analysis of long-range gene regulation by the lncRNA MALAT1 genomic locus using enhanced chromosome conformation capture & sequencing (e4C) in lung cancer cells. *J. Kim[1], R. Myers[2], D. Dorset[2], E. Ahn[1].* 1) University of South Alabama, Mitchell Cancer Institute, Mobile, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL.

Diverse oncogenic promoters create aberrant DNA loop structures and chromatin interaction with other genomic loci which play a key role in long-range regulation of histone modification and co-expression of gene pairs. Despite technical advances in chromatin structure analysis, only a limited number of long-range chromatin interactions have been identified related to oncogenic potential in cancers. During our analysis of ChIP-seq data of genome-wide binding sites of SON, a nuclear protein that interacts with both DNA and RNA, we identified a strong peak in a putative enhancer region downstream of the *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1) gene. MALAT1 is a long non-coding RNA (lncRNA) originally described as abundantly expressed in lung cancer. While the function of MALAT1 has been extensively studied in metastasis of diverse cancers, any role of the *MALAT1* genomic region in long-range chromatin regulation has not been explored. To examine detailed features of the sequence downstream of the *MALAT1* gene, we analyzed several ChIP-seq and ChIA PET-seq results from the ENCODE (Encyclopedia of DNA Elements) database. Interestingly, the data revealed that the genomic region downstream of the *MALAT1* gene represents high levels of H3K4me1 and H3K27Ac, numerous transcription factor binding sites, and potential interactions with other genomic loci. Since these features strongly suggest a potential enhancer function of this DNA region, we identified genome-wide chromatin interaction partners of the *MALAT1* downstream sequence using enhanced chromosome conformation capture and sequencing (e4C) method in A549 lung cancer cell line. Interestingly, our e4C revealed specific interactions between the *MALAT1* locus and many target genes which are closely associated with ion transport and GTPase regulation, such as *CACNA2D1* and *HVCN1*. Furthermore, among the interacting genes, we also found *VAV2* and *SMAD6*, well known metastatic genes, as potential co-expression gene pairs with *MALAT1* in lung cancer cells. Taken together, our study demonstrated that the DNA sequence located downstream of the *MALAT1* gene forms long-range chromatin interactions with several genetic loci which are critical for tumorigenesis and co-regulated by transcription factors and SON. Furthermore, our findings strongly suggest a potential role of the *MALAT1* genomic locus as a long-range regulatory *cis*-element that facilitates aberrant transcriptional regulation in lung cancer.

2952F

Frequent silencing of the candidate tumor suppressor *TSLAC1* **by promoter methylation in early stage lung adenocarcinoma.** *T. Kohmoto, K. Kajiura, T. Narut, K. Masuda, K. Kondo, A. Tangoku, I. Imoto.* Tokushima University, Tokushima, Tokushima, Japan.

Transcriptional repression through aberrant DNA methylation of promoter of specific tumor suppressor genes (TSGs) play an important role in carcinogenesis. In order to identify novel TSG candidates implicated in the early development of lung adenocarcinoma (LAC) regardless of smoking status, we performed a genome-wide screening of aberrantly methylated CpG-islands in 12 stage I LAC (6 smokers and 6 non-smokers) with paired non-cancerous tissues using the Illumina Human Methylation 450K BeadChips. A series of criteria targeting novel genes frequently silenced in LADC primary tumors and cell lines possibly through hypermethylation were applied to prioritize the selection of top candidates for validation. As the result, we identified TSLAC1 (Lab. name) for further detailed analyses of methylation and expression status using a larger panel of LADC tumors and cell lines. Frequent hypermethylation and silencing of TSLAC1 were observed in LAC tumors compared with paired non-tumorous lung tissues from early stage. Our results were also validated by a dataset obtained from the Cancer Genome Atlas (TCGA). Methylation-dependent silencing of TSLAC1 was confirmed by the 5-aza-2-deoxycytidine treatment-induced restoration of gene expression and promoter assay of the methylated region using LAC cell lines. In addition, a series of *in vitro* and *in vivo* functional analyses using LAC cell lines demonstrated that TSLAC1 suppresses anchorage-independent and dependent cell growth, suggesting that inactivation of this gene is a crucial early event for the development of LAC. Our results suggest that TSLAC1 is a novel TSG for LAC, which is inactivated by DNA hypermethylation regardless of smoking status from early stage of this disease.

2953W

Concerted transcriptional response to cancer drug treatment among genes involved in DNA methylation, demethylation, and folate-mediated one-carbon metabolism. *J. Krushkal[1], Y. Zhao[1], C. Hose[2], A. Monks[2], J.H. Doroshow[3], R. Simon[1].* 1) Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, MD; 2) Molecular Pharmacology Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 3) Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Dysregulation of DNA methylation and demethylation is abundant in cancer, and epigenetic pathways are increasingly being targeted in cancer treatment. Genetic components of the folate-mediated one-carbon metabolism pathway can affect DNA methylation and other vital cell functions, including DNA synthesis, amino acid biosynthesis, and cell growth. We examined how cancer drug treatment affects temporal changes in gene expression among epigenetic regulators of DNA methylation and demethylation, and one-carbon metabolism genes. We utilized data from the NCI-60 cell line panel, a popular resource from the U.S. National Cancer Institute, which includes 59 human cancer cell lines representing 9 cancer tissue types. We analyzed changes in gene expression after treatment with five antitumor agents, 5-azacytidine, doxorubicin, vorinostat, paclitaxel, and cisplatin. Each antitumor agent elicited concerted changes in gene expression of multiple pathway components across the cell lines. Expression changes of *FOLR2, SMUG1, GART, GADD45A, MBD1, MTR, MTHFD1,* and *CTH* were significantly correlated with chemosensitivity to some of the agents. The genes encoding DNA methyltransferases *DNMT1, DNMT3A,* and *DNMT3B,* epigenetic and DNA repair factors *MGMT, GADD45A,* and *MBD1,* and one-carbon metabolism pathway members *MTHFD1, TYMS, DHFR, MTR, MAT2A, SLC19A1, ATIC,* and *GART,* and a number of other important genes demonstrated concerted transcriptional response to individual antitumor agents. These transcriptional changes are likely to influence vital cellular functions of DNA methylation and demethylation, cellular growth, DNA biosynthesis, and DNA repair, and some of them may contribute to cytotoxic and apoptotic action of the cancer drugs. This concerted molecular response was observed in a time-dependent manner, which may provide future guidelines for temporal selection of genetic drug targets for combination drug therapy treatment regimens of cancer. Funded by NCI Contract No. HHSN261200800001E.

2954T

Characterization of microRNAs upregulated in metastatic lung adeno-carcinoma. *S.H.Y. Kung[1], K.S.S. Enfield[1], D.A. Rowbotham[1], C. Pastrello[2], B. de Carvalho Minatel[1], I. Jurisica[2], C. MacAulay[1], S. Lam[1], W.L. Lam[1].* 1) British Columbia Cancer Research Centre, Vancouver BC; 2) Princess Margaret Cancer Centre, Toronto ON.

   Introduction: Lung cancer is the leading cause of cancer-related death worldwide, where poor survival is attributed to late stage at diagnosis and lack of effective treatment options. Improved understanding of advanced stage lung cancer biology could lead to new therapeutic developments. MicroRNAs (miR-NAs) are known regulators of tumor cell invasion and migration. We identify *miR-106a* and *miR-106b* to be associated with metastasis in a clinical cohort of lung adenocarcinoma (AC) and assess their ability to regulate growth and metastasis. **Methods:** MiRNA expression was deduced from small RNA sequencing data derived from clinical lung AC specimens (60 localized, 27 with lymph node invasion) and paired non-malignant tissues. *MiR-106a* and *miR-106b* overexpression vectors and controls were stably transfected into stage I AC cell lines with low endogenous *miR-106* levels with an epithelial phenotype by lentiviral delivery. Migration and invasion was assessed by Boyden chamber assay, while cell proliferation was assessed by BrdU incorporation assay. Expression of epithelial-to-mesenchymal transition (EMT) markers and cell cycle proteins were assessed by Western Blot. DNA copy number, methylation, and Gene Set Enrichment Analysis were queried to identify putative mechanisms of *mir-106a/b* overexpression. Clinical associations in an external cohort were derived using publically available TCGA data. **Results:** *MiR-106a* and *miR-106b* were significantly overexpressed in lung AC with lymph node invasion. Overexpression of *miR-106a* and *miR-106b* significantly increased proliferation of lung AC cell lines, and was associated with decreased levels of predicted target, p21. AC cell lines displayed enhanced metastatic phenotypes *in vitro*, and were associated with increased mesenchymal and decreased epithelial markers, characteristic of EMT. Copy number and methylation status did not correlate with miRNA expression; however, upstream regulation by E2F is likely. In an external cohort, miR-106a/b and vimentin were associated with poor outcome. **Conclusions:** *MiR-106a* and *miR-106b* are overexpressed in metastatic lung AC and regulate the metastatic potential of cells by influencing EMT and cell proliferation. A deeper characterization of this observation may reveal therapeutic intervention points, or, with the development of miRNA therapeutics, *miR-106a/b* may be promising targets to prevent or treat metastatic disease. SHYK and KSSE contributed equally.

2955F

The Y-located TSPY and its X-homologue are co-activator and co-repressor respectively for androgen receptor functions in human cancers and diseases. *Y.-F.C. Lau, Y. Li, T. Kido.* Dept Medicine/ VAMC-111C5, Univ California, San Francisco, San Francisco, CA.

   Sexual dimorphisms are prevalent in human diseases and cancers. The male sex hormone, androgen, and its receptor, AR, could contribute to male-specific, e.g. prostate cancer, and sexually dimorphic cancers, e.g. liver cancer, and diseases, e.g. autism and Hirschsprung disease, in male-biased manners. Significantly, actions of Y chromosome genes could potentiate such hormonal effects on the oncogenic and pathogenic processes. The testis-specific protein Y-encoded (*TSPY*) is the gene for the gonadoblastoma locus on the Y chromosome (GBY). It has an X-located homologue, *TSPX*, which functions as a tumor suppressor in human cancers. *TSPY* and *TSPX* originated from the same ancestral gene and code for homologous proteins with a conserved protein-binding SET/NAP domain, but diverged significantly at their C-termini, in which TSPX harbors a large acidic domain, absent in TSPY. We show that both proteins interact with AR and its constitutively active variants (AR-Vs) lacking the ligand-binding domain. TSPY co-activates and TSPX co-represses AR/AR-V transactivation of target genes in ligand-dependent and independent manner respectively. ChIP analysis shows that TSPY and TSPX could co-localize with AR on the promoters and modulate the expression of endogenous AR target genes in prostate cancer cells. Domain mapping identifies the SET/NAP domain as their interactive domain with AR/AR-V, and the acidic tail of TSPX as the repressor domain. Transcriptome analysis shows that *TSPY* and *TSPX* exert contrasting properties on various canonical pathways, functions and diseases and upstream regulators, associated with onco-genesis and physiology. Hence, due to their evolutionary divergence, *TSPY* and *TSPX* function respectively as an oncogene and a tumor suppressor in androgen-responsive human cancers. Importantly, AR and AR-V bind to the *TSPY* promoter and up regulate its expression. Hence, *TSPY* and *AR/AR-V* form male-specific positive feedback loops that could amplify the male-biases in human oncogenesis. Further, *TSPX* also serves critical functions in neuro-evelopment and its mutations have been associated with various neural and cognitive diseases. The contrasting properties of *TSPY* suggest that, when expressed in somatic tissues, this Y-located gene could compete/disrupt *TSPX* functions and exert male-specific and/or androgen-sensitive effects in human development, physiology and disease processes, thereby contributing to the sexual dimorphisms in human development and diseases.

2956W

Investigating the role of *CIC* mutations in malignancy. *V.G. LeBlanc[1,2], S. Chittaranjan[1], M. Firme[1], S.Y. Chan[1], J. Song[1], M.H. Lee[1], S. Yip[3], M.A. Marra[1,4].* 1) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada; 2) Genome Science + Technology graduate program, University of British Columbia, Vancouver, BC, Canada; 3) Department of Pathology & Laboratory Medicine, UBC, Vancouver, BC, Canada; 4) Department of Medical Genetics, UBC, Vancouver, BC, Canada.

Type I low-grade gliomas (LGGs), characterized by 1p/19q co-deletions and *IDH1/2* mutations and mostly composed of oligodendrogliomas, show superior overall survival compared to other infiltrating gliomas; however, the molecular mechanisms underlying their characteristic slow growth and chemosensitivity remain poorly understood. Approximately 70% of Type I LGGs also harbour mutations in the *Capicua* (*CIC*) gene. CIC was originally identified as a tissue-specific transcriptional repressor in *Drosophila melanogaster*, functioning as an effector of the mitogen-activated protein kinase (MAPK) signalling pathway. These functions seem to be evolutionarily conserved and CIC appears to be involved in mammalian development, possibly playing a role in central nervous system development. CIC dysfunction has also been implicated in other cancer types, such as *CIC*-rearranged Ewing-like sarcomas, prostate cancer, and stomach adenocarcinomas (STADs), where *CIC* was recently found to be significantly mutated within the microsatellite instability (MSI) subtype. Nevertheless, CIC's role in human development and malignancy remains unclear. To address this discrepancy and to further elucidate CIC function, we developed *CIC* knockout cell line models and used transcriptome and proteome analyses to investigate the consequences of CIC loss. We also made use of publicly available data from The Cancer Genome Atlas (TCGA) and the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) for Type I LGGs and MSI subtype STADs to detect and explore transcriptomic and epigenomic differences between *CIC* wild type and *CIC* mutant tumour samples. We were able to identify known and novel high-confidence candidate targets of CIC transcriptional regulation. We found that, while CIC targets appear to be partially context-specific, the effects of CIC deficiency converge on the dysregulation of similar biological processes and pathways in different cancer types. In particular, we showed that CIC deficiency is associated with upregulated expression of up- and downstream members of the RAS/RAF/MEK/ERK signalling pathway. This suggests that loss of CIC may present a novel mechanism for activation of this oncogenic signalling pathway, which is affected in many cancer types. Overall, our results indicate an oncogenic role for *CIC* mutations and help uncover the mechanisms underlying their function in malignancy.

2957T

Elucidating the function of neomorphic *IDH* mutations in acute myeloid leukemia. *A. Lorzadeh[1], J. Parker[2], D. Knapp[3,4], M. Bilenky[1,2], M. Moksa[1], G. Paliouras[2], D. Hogge[4,5], M. Marra[2,3], C. Eaves[3,4,5], A. Karsan[2,5], M. Hirst[1,2].* 1) Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada; 2) Canada's Michael Smith Genome Science Centre, BCCA, Vancouver, BC, Canada; 3) 3Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Terry Fox Laboratory, BCCA, University of British Columbia, Vancouver, BC, Canada; 5) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

Introduction: Neomorphic isocitrate dehydrogenase (IDH) mutations are recurrent genetic lesions observed in Acute Myeloid Leukemia (AML) that drive the accumulation of the metabolite D-2-hydroxyglutarate (D2-HG). D2-HG is an inhibitor of 2-oxoglutarate-dependent dioxygenases (2-OGDD), a family of enzymes that include a class of epigenetic modifiers that are responsible for histone and DNA demethylation. In mouse models neomorphic IDH mutations have been shown to reversibly transform progenitor hematopoietic populations harboring FMS-like tyrosine kinase 3 internal tandem duplications (FLT3[ITD]). To investigate the role D2-HG might play in remodelling the epigenome of primary FLT3[ITD] AML blast cells we performed native ChIP-seq on cord blood derived CD34+ cells and primary AML cells harboring FLT3[ITD] in the presence and absence of neomorphic IDH mutations. **Methods:** We applied an optimized native ChIP-seq protocol to profile five histone modifications (histone3 lysine 4 tri-methylation (H3K4me3), H3K4me1, H3K27me3, H3K36me3, and H3K9me3). ChIP-seq was conducted on aliquots of ~10,000 CD34+ cells purified from cord blood by a 2-step Rosette-Sep-EasySep procedure to ≥95% purity and ~10,000 AML cells harboring the FLT3[ITD] allele and IDH2[R140Q] or IDH2[WT] alleles. The resulting libraries prepared from these cells were sequenced on an Illumina HiSeq platform in paired-end mode, analyzed using our ChIP-seq analysis tool FindER and integrated with existing matched RNA-seq datasets. **Results:** Pair-wise comparisons between AML-IDH2[WT]-FLT3[ITD] and AML-IDH2[R140Q]-FLT3[WT] revealed a global increase in H3K27me3 (a repressive histone modification) in AML-IDH2[R140Q]- FLT3[WT] compared to that of AML-IDH2[WT]-FLT3[ITD]. Among genomic regions with differentially modified nucleosomes were the HOXA and HOXB clusters that include the key AML driver HOXA9. Matched RNA-seq datasets confirmed transcriptional repression of genes with H3K27me3 gains in their promoter. Overall our results are consistent with a model where inhibition of 2-OGDD histone demethylases, such as the UTX (H3K27me3 demethylase) by 2D- HG lead localized gains of H3K27me3 and gene repression. **Conclusions:** Our results support a model where IDH neomorphic mutations lead to localized reprogramming of the epigenetic landscape of primary AML through the inhibition 2-OGDD histone demethylases.

2958F

Integrative genomics analysis identifies SIRPA as a key regulator of the EGFR pathway with both tumor suppressive and oncogenic properties. *E.A Marshall[1], R. Chari[2], J.Y. Kennett[1], S. Lam[1], W.L. Lam[1], L.A. Pikor[1].* 1) Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada; 2) Genetics, Harvard Medical School, Boston, USA.

   Background: The *Epidermal Growth Factor Receptor* (*EGFR*) signaling pathway is frequently deregulated in non-small cell lung cancer (NSCLC). Targeted therapies against EGFR effectively prolong survival. However, targeted therapies only benefit a subset of patients harboring EGFR alterations. Resistance inevitably develops, and additional genes downstream of *EGFR* are often abrogated in NSCLC. **Methods:** To investigate the complement of genetic alterations affecting the EGFR pathway in lung adenocarcinoma (AC), we performed a multi-dimensional integrative analysis of copy number, DNA methylation and gene expression profiles on 77 tumor and matched non-malignant tissues. Novel candidate genes were validated in external datasets and tissue microarray IHC was used to relate protein levels with clinical features. Signal-Regulatory Protein Alpha (SIRPA) functions were assessed by stable knockdown (KD) and overexpression (OE) in a panel of lung cancer cell lines *in vitro* and *in vivo*. **Results:** Our analysis revealed *SIRPA* to be underexpressed in 70% of tumors due to both copy loss and hypermethylation, ranking it in the 95th percentile of altered genes within the EGFR pathway. *SIRPA* negatively regulates receptor tyrosine kinsase activity and we observed frequent downregulation at both the mRNA and protein level in NSCLC tumors and cell lines which was correlated with EGFR mutations and an AC subtype. KD of SIRPA enhanced colony formation and wound healing while also simultaneously inducing a dramatic senescent phenotype, suggesting SIRPA may act as a barrier to tumorigenesis. We show this phenotype is dependent on upregulation of the CDK inhibitor p27 which hypophosphorylates RB leading to cell cycle blockade and reduced tumor growth in vivo. SIRPA OE abrogated senescence, promoting cell growth however it also promoted migration, suggesting SIRPA may also possess oncogenic properties. Western blotting and immunofluorescence confirmed that OE of SHP2 upon ectopic expression of SIRPA promotes migration through the inhibition of focal adhesions and that this phenotype is abrogated upon siRNA KD of SHP2. **Conclusion:** Our integrative analysis of the EGFR pathway revealed SIRPA to be frequently downregulated in lung AC. Through its regulation of multiple signalling pathways, SIRPA possesses both oncogenic and tumor suppressive properties, underscoring its importance in lung AC. To our knowledge, this is the first study to report a role for SIRPA in lung tumorigenesis.

2959W

The combined effect of methylation and copy number on gene expression across cancer types. *E.K. Moeller[1], O.C. Lingjaerde[1,2], A.L. Borresen-Dale[1], S. Nord[1].* 1) Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway; 2) Department of Computer Science, University of Oslo, Oslo, Norway.

   Regulation of gene expression is key to determine cell type and fate. An extra copy or loss of a gene or an entire chromosome will disrupt the normal DNA burden. Copy number alterations (CNAs) are frequently seen in many cancer types and it has been estimated that CNAs influence expression of ~40% of the genes in the breast cancer (BC) genome. Multiple regulatory factors of gene dosage are known, such as DNA methylation, histone modification, as well as post-transcriptional and translational regulation. In order to better understand what drives the variation in gene expression and the underlying biology of tumor cells, we propose to integrate the two most prominent features: CNAs and methylation. CNAs are a key driver of BC and differences in degree of methylation within BC subtypes have been reported. Therefore BC is a good model system for studying the integrative effect of methylation and/or CNAs on gene dosage in cancer. We have developed a novel method that captures methylation patterns per-gene rather than per-probe, which is the approach that has typically been applied in previous studies. The model was trained on all BC samples from TCGA (n=648) and results were validated in two independent cohorts. Our results show that CNAs have a greater impact on gene expression than methylation, both with respect to the number of genes they influence and their total effect. Overall, we found that methylation, CNAs or both, significantly influenced expression of 39%, 62.2% and 78% of all genes, respectively. Typically, focal aberrations showed the largest impact on expression, and interestingly, they harbored reduced methylation levels. This included hallmark alterations of BC, such as 8p11.23 and 17q12. To assess whether gene dosage regulation is uniform across cancer types, we applied the model to eight different cancers, including BC. Pan Cancer analysis revealed similar patterns with significant associations for a large number of genes, but their total impact on gene expression was limited. For most cancer types, CNAs rather than methylation showed a stronger influence on gene expression, and it was correlated to the genomic instability index (GII). However, the key regulator of gene dosage is cancer type dependent. For known focal aberrations in BC, an inverse correlation was observed between CNAs and methylation, enhancing the oncogenic effect of genes such as *HER2*.

2960T

Piwi-interacting RNAs display cancer-specific expression patterns that hold prognostic value in lung cancer. *K. Ng, B.C. Minatel, E.A. Marshall, K.S.S. Enfield, S. Lam, W.L. Lam, V.D. Martinez.* Integrative Oncology, BC Cancer Agency Research Centre, Vancouver, BC, Canada.

   Background: PIWI-interacting RNAs (piRNAs) are small (24-32bp) non-coding RNAs with a key role in epigenetic regulation of gene expression and maintenance of genomic stability in germ cells. Recent evidence suggests they are also expressed and functionally active in somatic tissues and cancer. Aberrant expression of individual piRNAs has been associated with clinical features in some cancer types; however, their involvement in lung cancer remains to be deciphered. Towards this end, we analyzed piRNA transcriptomes from human tumors and adjacent non-malignant tissues, and identify lung cancer specific expression patterns associated with clinical features and prognosis. **Methods:** We developed a custom small-RNA analysis pipeline to deduce expression of approximately 32K human piRNAs. As a baseline, we generated 6,378 piRNA transcriptomes (non-malignant/tumor tissue) from 11 organ sites. In lungs, we analyzed 1,082 tumors and 209 non-malignant samples from two different cohorts (The Cancer Genome Atlas/BC Cancer Agency). Using a network-based approach, weighted gene co-expression network (WGCNA), we evaluated clinical parameters of tumor aggressiveness (stage, number of mutations, nodal/distant metastasis) and outcome (overall/disease-free survival). Multi-piRNA survival signatures were identified using a Cox Proportional Hazard model. **Results:** Only a small proportion of piRNAs is somatically expressed, and displays highly organ-specific expression. Tumor piRNA expression profiles are markedly different from their non-malignant counterparts, and are correlated with clinical features. Specifically in lung cancer, we find piRNA expression differs between the subtypes adenocarcinoma and squamous cell carcinoma, and we identify piRNA expression networks associated with features of tumor aggressiveness and patient survival. **Conclusions:** We provide evidence of somatic, tissue-specific human piRNA expression. Aberrant expression patterns contribute to lung cancer subtype-specific biology. We discover piRNA-based signatures that identify aggressive lung tumors and have prognostic value. The unique expression patterns of piRNAs offer an opportunity to better understand cancer-specific biology as well as develop novel prognostic markers for clinical application.

2961F

Cell of origin shapes the molecular profile of tumors across cancer types. *S.H. Nordgard[1], L. Silwal-Pandit[1], E.K. Møller[1], O.C. Lingjærde[2], A. Langerød[1], A-L. Børresen-Dale[1], T. Sørlie[1].* 1) Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway; 2) Department of Informatics and Centre for Cancer Biomedicine, Biomedical Informatics, University of Oslo, Oslo, Norway.

   Cell of origin of many different tumor subtypes remains unknown. Basal-like tumors represent a unique molecular entity within breast cancer. Here we hypothesize that the distinctiveness of basal-like tumors may represent features that go beyond organ of origin to include squamous tumor of the cervix, head and neck, lung, bladder and esophagus. The commonality arises from similar cell of origin, and this may be reflected by shared methylation pattern. Taking advantage of the universality of the PAM50 breast cancer classifier, we clustered all 23 solid tumor types available through TCGA by their PAM50 expression profiles. This exercise was subsequently repeated using multiple other intrinsic gene-lists, ensuring robustness to our observations. Two supra clusters were produced; one with tumors clustering in an organ-specific manner including low proliferative and differentiated prostate, non basal-like breast, thyroid, kidney and pancreatic cancers, and a fraction of lung adenocarcinomas and brain tumors. The other supra cluster was composed of highly proliferative tumors where organ specificity was less prominent. The basal-like breast cancers formed, together with squamous carcinomas of the lung, head and neck, cervix, esophagus and bladder, a distinct cluster within this supra cluster (referred to as Pan Cancer basal-like), characterized by an epithelial signature and elevated expression of certain oncogenes and tumor suppressor genes such as *CDKN2A*. The remaining tumors included adenocarcinomas of the lung and digestive system and a subcluster of stem cell like tumors including amongst others, testicular germ cell cancers. Interestingly, clustering the same tumors by their corresponding PAM50 methylation profiles recapitulated the grouping seen by gene expression. The Pan Cancer Basal-like tumors continued to form a distinct entity, and this was amongst other characteristics, driven by the hypomethylation of keratins and the *PTTG1* oncogene. Our results support a "cell-of-origin" dominance on the molecular portraits of the tumors, and that the hypomethylation patterns of certain tumors may have limited influence on tumor progression, and instead, reflect the differentiation level of the cell of origin.

2962W

Lung cancer: One disease or many? *T. O'Brien[1,2], P. Jia[3], Z. Zhao[2,3].* 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA; 3) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

　　Molecular subtyping of cancers may facilitate the development of targeted therapies that treat each subtype like a single disease rather than one cancer type. However, many cancers are still only characterized by their specific tissue of origin. Lung cancer, for example, is classified into two main types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC has several subtypes, the two most common being lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). These classifications are mostly based on lung physiological and clinical characteristics, although there is increasing evidence of genetic and molecular differences as well. Previous work has identified unique somatic driver genes for each subtype and suggests more than one distinct disease. However, little work has been done to characterize unique germline susceptibility genes. We hypothesize that non-coding common variants act in a regulatory manner to control expression of genes associated with each lung cancer subtype. Thus, we explored the regulatory roles for SNPs associated with each subtype from genome-wide association studies (GWAS) of Europeans. We obtained SNPs associated with each subtype by $p < 10^{-4}$ including SNPs in linkage disequilibrium (LD) with $r^2 > 0.8$ from the 1000 Genomes Project. This expansion resulted in 1427, 1700, and 1059 SNPs for LUAD, LUSC, and SCLC, respectively. We identified SNPs within regulatory regions and their target genes using several lung tissue-specific expression quantitative trait loci (eQTL) and enhancer datasets. We found a total of 27, 60, and 19 germline regulated genes for LUAD, LUSC, and SCLC, respectively. Interestingly, only three genes (*CHRNA5*, *IDH3A*, and *RP11-650L12.2*) overlapped among all three subtypes. *CHRNA5* and *IDH3A* have prior associations with lung cancer, while the function of *RP11-650L12.2* has not been characterized. We found biological pathways enriched with genes from each subtype, such as regulation of actin cytoskeleton and focal adhesion that are both involved in the epithelial-mesenchymal transition (EMT), important in cancer metastasis, only in the most aggressive subtype, SCLC. These results, along with somatic profile differences, suggest these subtypes are unique diseases that only share a small set of germline regulated genes and the difference in the biological basis of each subtype may partially be explained by common variants altering expression of genes that are unique to each subtype.

2963T

Deregulation of long non-coding RNAs in lung adenocarcinoma highlight alternative regulatory mechanisms. *A. Sage, G. Stewart, K. Enfield, D. Rowbotham, R. Hubaux, S. Lam, W. Lam.* Integrative Oncology, BC Cancer Agency Research Centre, Vancouver, BC, Canada.

　　Background: Lung cancer remains the cause of the most cancer-related deaths each year, with a 5 year survival rate of less than 17%. The poor therapeutic outcome is largely due to a complex molecular background as well as late stage diagnosis, with most patients presenting unresectable local tumors, or metastatic disease. Targeted therapeutics have been developed against well-defined drivers of the lung adenocarcinoma (AC) subtype, but such treatment regimens are relevant only to the proportion of patients harbouring these genetic aberrations, emphasizing the need to explore alternative mechanisms of AC development. Natural antisense transcripts (NATs) are long non-coding RNA products expressed from the opposite strand of coding mRNAs. NATs can function in *cis* or *trans* to regulate the transcriptional activity of their cognate gene partner in either a positive or negative fashion. Here we take an unbiased approach to identify NATs deregulated in lung AC, and explore function of these genes with regards to their protein coding partner genes. Methods: We performed RNA-sequencing on a set of 36 lung AC and matched non-malignant lung tissues. A sign-rank test was used to identify NATs and partner genes with significantly altered expression between tumor and matched normal tissues. These findings were validated in an external dataset of lung AC tumors from The Cancer Genome Atlas (TCGA). Survival analysis was performed using a Cox Proportional hazard model, as well as the log-rank method. Results: Analysis of Illumina Hi-seq data from TCGA revealed the majority (79%) of deregulated sense-antisense partnerships observed in AC displayed concordant regulation. However, several cis-NAT pairs were identified including an antisense to *OIP5*, a lung cancer oncogene required for chromatin segregation. *OIP5 Anti Sense 1* (*OIP5-AS1*) was significantly underexpressed in AC, while in the same tumors we find the overlapping partner gene, *OIP5* mRNA, to be significantly overexpressed, suggesting that *OIP5-AS1* may negatively regulate its sense counterpart. In addition, we find that expression of both OIP5 and *OIP5-AS1* are significantly associated with 5-year survival. Conclusions: These findings suggest that deregulation of cancer associated genes through NATs may represent a novel mechanism regulating tumor phenotypes in lung AC.

2964F

Lemur Tyrosine Kinase-2, a potential novel target for prostate cancer therapy. *K. Shah, N. Bradbury.* Physiology & Biophysics, Rosalind Franklin University of Medicine & Science, North Chicago, IL.

Progression from early forms of prostate cancer to castration-resistant disease is associated with an increase in signal transduction activity. The majority of castration-resistance cancers persist in the expression of the androgen receptor (*AR*), as well as androgen-dependent genes. The *AR* is regulated not only by it associated steroid hormone but also by manifold regulatory and signaling molecules, including several kinases. We undertook an evaluation of the role of Lemur Tyrosine Kinase 2 (*LMTK2*) in modulating *AR* activity, as several Genome Wide Association Studies (GWAS) have shown a marked association of *LMTK2, TET2* and *FAM84B* gene expression with the prostate cancer risk. We confirm that not only is *LMTK2* mRNA reduced in prostate cancer tissue, but also *LMTK2* protein levels are markedly diminished. Furthermore, knockdown of *LMTK2* protein in prostate cancer cell lines significantly increased the transcription of androgen-responsive genes. Interestingly, loss of *LMTK2* led to an increase in populations of prostate cancer stem cell, indicative of increased tumourigenicity. Using multiple approaches, we also demonstrate that *LMTK2* interacts with the *AR* and that a loss of *LMTK2* in castrate resistant prostate cancer cell leads to an increase in total *AR* expression. Thus, indicating that a loss of *LMTK2* most probably stabilizes *AR* protein, increases *AR* phosphorylation and results in an increase in castrate-resistant prostate cancer cell proliferation and tumorigenicity. Takayama et al. recently showed that *AR* activity can mediate the expression of a *TET* family, *TET2*, which is correlated with prostate cancer progression. Hence, based on our data, there is a strong possibility that a loss of *LMTK2* can modulate *TET2* expression via androgen-AR axis. As such, our study identifies *LMTK2* as a novel regulator of *AR* and potentially as a novel therapeutic target.

2965W

Integrated analysis of genome-wide miRNAs and targeted gene expression in esophageal squamous cell carcinoma (ESCC). *H. Su[1,2], H. Yang[3], N. Hu[2], C. Wang[2], L. Wang[1,2], D. Ti[4], C. Giffen[5], A.M. Goldstein[2], M.P. Lee[3], P.R. Taylor[2].* 1) Leidos Biomedical Research, Inc. MD 21702, USA; 2) Genetic Epidemiology Branch, DCEG, NCI, MD 20892, USA; 3) Laboratory of Population Genetics, CCR, NCI, MD, 20892, USA; 4) Shanxi Cancer Hospital, Taiyuan, Shanxi 030013, P. R. China; 5) Information Management Services, Inc, Silver Spring, MD 20904, USA.

Esophageal squamous cell carcinoma (ESCC) occurs worldwide as the sixth leading cause of cancer mortality and the fourth most common new cancer in China. ESCC is an aggressive tumor with about less than 20% 5-year survival rate, which is fourth worst among all cancers in the USA. Studies have demonstrated that miRNAs modulate gene expression by binding to the 3' untranslated region (UTR) of target mRNA, causing either mRNA degradation or translation inhibition. DNA microarray and next-generation sequencing provide genome-wide scale techniques to identify miRNA signatures which can be potentially applied into cancer diagnosis, prognosis and even personalized medicine. Up to date, no genome-wide analysis of miRNA and mRNA expression in the same set of paired ESCC samples has been studied. Here, we conducted a study of miRNA and mRNA by using same 113 paired ESCC frozen samples to find that a) novel differential expressed miRNAs and mRNA involved in the pathogenesis of ESCC in the same sample set; b) miRNA significantly correlates with mRNA expressed within the same ESCC set; c) significant miRNA expressions associate with patients' clinical parameters including patient's survival as a potential prognostic biomarker.

2966T

Genome-wide DNA methylation assessment of "*BRCA1*-like" early-onset breast cancer: Data from the Australian Breast Cancer Family Registry. *E.M. Wong[1], C.M. Scott[1], J.H.E. Joo[1], N. O'Callaghan[1], J. Dowty[2], G.G. Giles[2,3], J.L. Hopper[2], M.C. Southey[1].* 1) Dept of Pathology, Univ Melbourne, Parkville, VIC, Australia; 2) Centre for Epidemiology and Biostatistics, Univ Melbourne, Parkville, VIC, Australia; 3) Cancer Epidemiology Centre, Cancer Council Victoria, VIC, Australia.

   Introduction: Breast cancers arising in women carrying a germline mutation in *BRCA1* are typically high-grade, early-onset, and have distinct morphological features. Approximately 75% of early-onset breast cancer cases of this subtype (often referred to as "*BRCA1*-like") have the same features but do not carry an identifiable germline *BRCA1* mutation. DNA methylation can mimic the effect of genetic mutation and result in gene silencing. **Methods:** In this study, we examined the role of methylation in "*BRCA1*-like" breast cancer by measuring genome-wide methylation in blood-derived DNA from women under the age of 40 years, participating in the Australian Breast Cancer Family Registry diagnosed with: i) "*BRCA1*-like" breast cancer ("*BRCA1*-like"; n=30); and ii) breast cancer without "*BRCA1*-like" morphological features (non "*BRCA1*-like"; n=30), and age-matched unaffected women (n=30) (controls). Corresponding tumor-derived DNA from 43 of the affected women was also assessed. **Results:** Methylation in blood-derived DNA was elevated across 17 consecutive marks in the *BRCA1* promoter region and decreased at several other genomic regions (including *TWIST2* and *CTBP1*) for seven (23%) women diagnosed with "*BRCA1*-like" breast cancer compared with women in the other groups. Corresponding tumor-derived DNA from these seven women had elevated methylation within the *BRCA1* and *SPHK2* promoter regions and as well as decreased methylation within *MYO10* and *HLA-E*. **Conclusion:** These methylation marks could be biomarkers of "*BRCA1*-like" breast cancer, and responsible in part for their distinctive morphological features, biology, and even risk of breast cancer. This could assist with prevention and targeted therapies for this cancer subtype.

2967F

Tracing Enhancer Networks using Epigenetic Traits (TENET) characterizes enhancer networks from heterogeneous methylomes. *S.K. Rhie[1], G. Yu[1], Y.G. Tak[1], L. Yao[1], H. Shen[2], G.A. Coetzee[2], P.W. Laird[2], P.J. Farnham[1].* 1) Department of Biochemistry and Molecular Biology and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Van Andel Research Institute, Grand Rapids, MI.

   A single genome can give rise to several hundred distinct cell types that are genetically identical but display different epigenetic marks at regulatory elements, leading to altered gene expression. There are two main types of regulatory elements involved in transcriptional activation, promoters and enhancers. Promoters are defined as a relatively small region surrounding a transcription start site of a gene, and are critical for basal transcription of that gene. Enhancers are regulatory elements, containing multiple transcription factors binding sites, which can be far upstream or downstream of the gene they regulate. In fact, the state most consistently linked to cellular identity is the 'active enhancer' state. Recent studies from the Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenome Mapping Consortium (REMC) have shown that more than ten thousand enhancers can be identified using epigenomic marks in a given cell line or tissue. However, it's not clear whether all of these enhancers are functional and which gene is connected to which enhancer. To facilitate understanding of enhancer networks, we have developed a method called Tracing Enhancer Networks using Epigenetic Traits (TENET), which identifies enhancers and genome-wide links between enhancers and genes using DNA methylation and gene expression data. Using profiles of tumor tissues from The Cancer Genome Atlas (TCGA) and next generation sequencing data in cell lines, we identified more than eighty thousand enhancers and further linked a subset of these enhancers to gene expression in prostate, breast and kidney tissues. Our studies revealed epigenetically regulated genes including key transcription factors that are involved in enhancer networks for specific cancer types. For example, we identified key transcription factors (e.g. *GATA3* in breast cancer and *HOXC6* in prostate cancer) linked to many tumor-specific enhancers, and further validated binding of these transcription factors to the enhancers using ChIP-seq data. Additionally, we detected a subset of prostate tumors mutated for *FOXA1* that has more tumor-specific enhancers, more enhancer to gene links, and higher Gleason scores. Thus, TENET can be used to characterize tumor-specific enhancer networks and mechanisms leading to the development of different cancer subgroups. TENET can also be applied to studies beyond cancer and can characterize enhancer networks for different types of case vs. control datasets.

2968W

Epigenome-wide tumor DNA methylation profiling identifies novel prognostic biomarkers of metastatic-lethal progression in men with clinically localized prostate cancer. *J.L. Stanford[1,2], S. Zhao[3], M.S. Geybels[1,4], A. Leonardson[1], R. Rubicz[1], S. Kolb[1], Q. Yan[5], B. Klotzle[6], M. Bibikova[6], A. Hurtado-Coll[7], D. Troyer[8], R. Lance[9], D.W. Lin[1,10], J.L. Wright[1,10], E.A. Ostrander[11], J. Fan[6], Z. Feng[5].* 1) Division of Public Health Sciences, Fred Hutchison Cancer Research Center, Seattle, Washington; 2) Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington; 3) National Institute of Environmental Health Sciences, Biostatistics & Computational Biology Branch, Research Triangle Park, North Carolina; 4) Department of Epidemiology, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands; 5) MD Anderson Cancer Center, Houston, Texas; 6) Illumina, Inc., San Diego, California; 7) Department of Urologic Sciences, University of British Columbia, and the Prostate Center, Vancouver General Hospital, Vancouver, Canada; 8) Departments of Pathology, Microbiology, and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk, Virginia; 9) Department of Urology, Eastern Virginia Medical School, Norfolk, Virginia; 10) Department of Urology, University of Washington School of Medicine, Seattle, Washington; 11) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland.

Purpose: Aside from Gleason sum few factors accurately identify the subset of prostate cancer (PCa) patients at high risk for metastatic progression. We hypothesized that epigenetic alterations could distinguish prostate tumors with life-threatening potential. **Experimental Design:** Epigenome-wide DNA methylation profiling was performed in surgically resected primary tumor tissues from a population-based (n = 430) and a replication (n = 80) cohort of PCa patients followed prospectively for at least five years. Metastasis was confirmed by positive bone scan, MRI, CT or biopsy, and death certificates confirmed cause of death. AUC, partial AUC (pAUC, 95% specificity), and P-value criteria were used to select differentially methylated CpG sites that robustly stratify patients with metastatic-lethal from non-recurrent tumors, and which were complementary to Gleason sum. **Results:** Forty-two biomarkers stratified patients with metastatic-lethal versus non-recurrent PCa in the discovery cohort, and eight of these CpGs replicated in the validation cohort based on a significant (P <0.05) AUC (range: 0.66-0.75) or pAUC (range: 0.007-0.009). The biomarkers that improved discrimination of patients with metastatic-lethal PCa include CpGs in five genes (*ALKBH5*, *ATP11A*, *FHAD1*, *KLHL8*, and *PI15)* and three intergenic regions. In the validation dataset the AUC for Gleason sum alone (0.82) significantly increased with the addition of four individual CpGs (range: 0.86-0.89; all P <0.05). **Conclusions:** Eight differentially methylated CpGs that distinguish patients with metastatic-lethal from non-recurrent tumors were validated. These novel epigenetic biomarkers warrant further investigation as they may improve prognostic classification of patients with clinically localized PCa and provide new insights on tumor aggressiveness.

2969T

TET2 loss modifies androgen signaling in prostate cancer. *M.L. Nickerson.* Laboratory of Translational Genomics, National Cancer Institute, Bethesda, MD.

We recently characterized the exomes of multiple tumors from a patient with metastatic prostate cancer (PCa) to identify alterations associated with disease. We detected a somatic alteration of the methylcytosine dioxygenase *ten-eleven translocation 2* (*TET2*), which is altered in 5-15% of myeloid, kidney, colon and prostate cancers. We show *TET2* is expressed in normal prostate tissue and reduced in a subset of tumors from the Cancer Genome Atlas (TCGA). Small interfering RNA (siRNA)-mediated *TET2* knockdown (KD) increases LNCaP and DU145 prostate cell proliferation, and LNCaP migration and wound healing, verifying *TET2* loss drives a cancer phenotype. By affinity chromatography and immunoprecipitation, endogenous TET2 bound the androgen receptor (AR) and AR-coactivator proteins in LNCaP cell extracts, and *TET2* KD increases prostate-specific antigen (*KLK3*/PSA) expression. Published data reveal TET2 binding sites and hydroxymethylcytosine proximal to *KLK3*. A gene co-expression network identified using TCGA prostate adenocarcinoma tumor RNA-sequencing data shows *TET2* expression is co-regulated with cancer genes associated with 2-oxoglutarate (2-OG) metabolism, including the lysine demethylase *KDM6A*, BRCA1-associated *BAP1*, and the citric acid cycle enzymes *IDH1/2*, *SDHA/*B, and *FH*. This suggests a role for TET2 as an energy sensor of 2-OG that may modify androgen-AR signaling based on the metabolic state of the cell. Decreased *TET2* mRNA expression in TCGA PCa tumors is strongly associated with reduced patient survival indicating reduced *TET2* expression in tumors maybe an informative biomarker of disease progression.

2970F

RNA polymerase II pausing regulates human gene expression including EGF signaling pathway. *J.M. Boden[1], V.G. Cheung[2,3,4].* 1) Cancer Biology Graduate Program; 2) Life Sciences Institute; 3) Howard Hughes Medical Institute; 4) Departments of Pediatrics and Human Genetics, University of Michigan, Ann Arbor, MI.

This project focuses on promoter-proximal RNA Polymerase II (RNA Pol II) pausing and its effects on global transcription and cellular functions. Pausing of RNA Pol II occurs 20-60 nucleotides downstream of transcription start sites through interactions with the pausing factors negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). In human cells, the expression levels of a few genes such as *MYC* and *JUNB* are known to be regulated through RNA Pol II pausing; however, little is known about the global effects of RNA Pol II pausing on human gene expression and cellular functions (Krumm, A. et al. 1992, Aida, M. et al. 2006). Here, we show that RNA Pol II pausing regulates thousands of human genes, including those in the epidermal growth factor (EGF) signaling pathway. First, we generated a genome-wide profile of paused RNA Pol II by PRO-seq (Kwak, H. et al. 2013) which identified proximal promoter pausing at over 7,800 human genes. We then confirmed the presence of RNA Pol II and the pausing factors NELF and DSIF at the promoter regions of these genes by ChIP-seq. To identify genes whose expression levels are regulated by the pausing factors, we knocked-down NELF and examined changes in pausing and gene expression. Following loss of NELF, we found that over 1,200 genes show a decrease in pausing as well as altered expression. Many of these genes are members of the EGF signaling pathway, including the EGF receptor (EGFR). Because EGFR is a driver of many human cancers, it has been the target of a number of successful therapies. However, patients commonly develop resistance to these treatments. A detailed understanding of transcriptional regulation of EGFR may enable development of strategies to overcome this resistance. To this end, we explored how EGF signaling is regulated by RNA Pol II pausing. We validated that EGFR transcript and protein levels increase upon loss of pausing. We then found that the increased EGFR expression results in an increased response to EGF stimulation as measured by the expression of *CDKN1A* and *CCND1*. In this presentation, we will show that pausing regulates EGFR expression and subsequently its function. We will present data on the effect of pausing on ERK kinase phosphorylation, its target gene expression and cell cycle progression.

2971W

Integrated multi'omics analysis identifies pathogenetic molecular phenotypes of lung cancer and COPD. *V.D Martinez[1], E.A. Vucic[1], K.L. Thu[1], E.A Marshall[1], K.S.S Enfield[1], B.C Minatel[1], S. Rahmati[2], M. Abovsky[2], I. Jurisica[2], C.E. MacAulay[1], S. Lam[1], R.T. Ng[3], W.L. Lam[1].* 1) Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada; 2) Princess Margaret Cancer Centre, Toronto, ON, Canada; 3) PROOF Centre of Excellence, St. Paul's Hospital, Vancouver, BC, Canada.

Background: Inflammatory processes are associated with an increased cancer risk in multiple organs. Chronic obstructive pulmonary disease (COPD) is an inflammatory lung disease associated with a 10-fold increased risk of lung cancer (LC), independent of smoking habits. We hypothesized that lung tumors arising from COPD and non-COPD patients exhibit molecular phenotypes reflective of distinct and specific biology underlying tumorigenesis in COPD patients. Here, we aim to identify disrupted functional gene networks associated with COPD, that contribute to the biology of COPD-related lung cancer. **Methods:** Genome-wide multi-omics profiling (copy number, DNA methylation, mRNA) was performed on tumor and normal lung tissues (n=146). Additionally, mRNA expression levels from non-cancerous, small (<2 mm) airways from COPD and non-COPD patients (n=267) were also analyzed. Integrated gene scores were generated based on magnitude and disruption of DNA and mRNA level change, on a per sample pair basis. Cluster analysis was carried out using the non-negative matrix factorization method. Module-based functional analyses were performed using the weighted gene co-expression network analysis (WGCNA) algorithm. Functional analyses were done using IPA and pathDIP algorithms. **Results:** Globally, COPD and non-COPD lung tumors cluster separately based on their gene integrated scores. COPD-related lung tumors are differentially enriched for gene sets associated with senescence (PITX2), lung malignancies (NR2F2, SP1, PPARG) and inflammation-associated fibrosis and cancer of the liver (HNF4A). A module-based co-expression network analysis identified XBP1/SEC31A/COG3 as the most significantly enriched gene regulatory network in non-cancerous COPD tissues (p = 9.01E-08), as well as highly altered at DNA and mRNA levels COPD-tumors. Interestingly, these networks have been associated with inflammatory liver disease, cirrhosis, proliferation of hepatocytes and liver cancer. **Conclusions:** COPD and non-COPD lung tumors harbor distinct biological processes. The identification of common regulatory network between non-cancerous COPD tissue and COPD-related lung cancer reflects the interplay of both diseases and reveals involvement of tumor-inducing mechanisms related to chronic inflammation in other organs. These and other gene networks altered in non-cancerous airway cells associated with LC in COPD patients may represent ideal targets for LC risk stratification or therapy.

2972T

Epigenetic signature of Gleason score and prostate cancer recurrence after radical prostatectomy. *M.S. Geybels[1,2], J.L. Wright[1,3], M. Bibikova[4], B. Klotze[4], J. Fan[4], S. Zhao[5], Z. Feng[6], E.A. Ostrander[7], D.W. Lin[1,3], P.S. Nelson[8,9,10], J.L. Stanford[1,11].* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, GROW School for Oncology and Developmental biology, Maastricht University, Maastricht, The Netherlands; 3) Department of Urology, University of Washington School of Medicine, Seattle, Washington; 4) Illumina, Inc., San Diego, California; 5) National Institute of Environmental Health Sciences, Biostatistics & Computational Biology Branch, Research Triangle Park, North Carolina; 6) Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas; 7) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland; 8) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington; 9) Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington; 10) Department of Medicine, University of Washington School of Medicine, Seattle, Washington; 11) Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington.

Background Identifying the subset of patients with localized prostate cancer at the highest risk of recurrence remains challenging and better prognostic markers are needed. DNA methylation patterns in tumors cells may predict prostate tumors' aggressiveness potential. The aim of this study was to identify an epigenetic signature of Gleason score (or tumor grade), the best predictor of prostate cancer prognosis, and evaluate its ability to predict prostate cancer recurrence after radical prostatectomy. **Methods** Genome-wide DNA methylation data from The Cancer Genome Atlas (TCGA; No. prostate cancer patients = 333) and the Elastic Net method were used to generate the tumor epigenetic signature. The signature was created by contrasting patients with high (8–10) versus low (≤6) Gleason score tumors. The signature was then tested for its ability to predict prostate cancer recurrence in another cohort of 431 patients with localized prostate cancer who had radical prostatectomy as primary treatment and were prospectively followed for recurrence (mean follow-up time = 8 years). Samples taken from the primary tumor were used for DNA methylation profiling. **Results** The epigenetic signature of Gleason score includes 52 differentially methylated CpG sites. In the validation cohort, the signature was associated with poorer recurrence-free survival (hazard ratio per 25 percent increase = 1.78; 95% confidence interval: 1.48, 2.16), which remained significant after adjusting for standard clinical-pathological parameters. Among patients with Gleason 7 tumors, which are heterogeneous and have a variable prognosis, the signature significantly improved the Area Under the Curve (AUC) for prostate cancer recurrence compared to clinical-pathological parameters alone (0.64 vs. 0.76, P = 1.34E-4). Results were similar for patients with Gleason 3+4 and those with 4+3 tumors. The signature includes CpGs in different genes, which are enriched for genes encoding cell cycle-related targets of E2F transcription factors (e.g., *PLK1*, *CDC25B*, *MKI67*, and *RRM2*). **Conclusions** This study shows evidence that DNA methylation patterns measured in prostate tumor cells are predictive of prostate cancer aggressiveness. The epigenetic signature may have clinical utility to improve prognostication particularly in patients with intermediate grade tumors that have a Gleason score of 7.

2973F

Internal smoking dose is associated with specific blood DNA methylation patterns and with different magnitude across race/ethnicity. *S.L. Park[1], Y.M. Patel[1], M. Tiirikainen[2], D.O. Stram[1], K. Siegmund[1], S.E. Murphy[3], L. Le Marchand[2].* 1) University of Southern California, Los Angeles, CA; 2) University of Hawaii Cancer Center, Honolulu, HI; 3) University of Minnesota, Minneapolis, MN.

Lung cancer is the most common cancer in the U.S. and leading cause of cancer-related death. While, cigarette smoking is the primary cause of this malignancy, risk differs across racial/ethnic groups. For the same number of cigarettes smoked, Native Hawaiians (NH) compared to whites, are at greater risk of lung cancer, while Japanese Americans (JA) are at lower risk of disease. DNA methylation of CpG sites as a result of cigarette smoking is one of the most common epigenetic modifications. Prior methylation studies have found that current smoking status was associated with differential methylation patterns from leukocytes in 26 genetic regions. However, the influence of internal smoking dose on the epigenome has not been investigated, nor has a study evaluated whether for the same smoking dose, peripheral blood DNA methylation patterns differ by race. Here, we have conducted an epigenome-wide association study for urinary nicotine equivalents (NE), which is an optimal marker of internal smoking dose, as opposed to smoking history or cotinine, since NE is not biased by self-report and reflects inter-individual variation in nicotine metabolism. This study includes 612 current smokers at time of specimen collection, from three racial/ethnic groups: whites, NH and JA. Genome-wide DNA methylation in blood leukocytes was measured using the Illumina 450K BeadChip array. Distribution of leukocyte cell-types (based on methylation data) were estimated for each sample (https://labs.genetics. ucla.edu/horvath/dnamage/). Average beta-values or the ratio of methylated probe intensities (i.e. intensity of the Methylated allele [M] / (intensity of the Unmethylated allele [U] + M)) was the dependent variable in linear regression models adjusting for age, sex, race (for pan-ethnic analysis) and cell-type distribution. We found that NE was associated with six differentially methylated probes (Bonferoni corrected $p < 1.48 \times 10^{-7}$) in or near the *FOXK2*, *PBX1*, *FNDC7*, *FUBP3*, and *BSND* genes, where higher levels of NE were associated with increasing beta-values in all six probes. These six probes also had statistically significant interactions by race, where the strongest associations were found in NH, suggesting that NH have greater differential DNA methylation patterns, compared to whites and JA (p-interaction$<8.8 \times 10^{-8}$), in relation to smoking dose. In conclusion, NE is associated with specific DNA methylation of leukocytes and these patterns differ in magnitude by race/ethnicity.

2974W

Microvesicular miRNAs regulate gene expression in local lung adenocarcinoma stromal cells. *J. Lawson[1,2], C. Dickman[1,2], S. MacLellan[1,2], R. Towle[1,2], C. Garnis[1,2].* 1) University of British Columbia, Vancouver, British Columbia, Canada; 2) BC Cancer Research Center, Department of Integrative Oncology, Vancouver, British Columbia, Canada.

   Introduction: Microvesicles are small vesicles released from all cell types which can be used as a form of cell to cell communication. Recently tumor derived microvesicles have been shown to play a key role in cancer development, growth, progression and angiogenesis. These microvesicles can be loaded with functional mRNAs, miRNAs and proteins which can be transferred from one cell to another. miRNAs within microvesicles differ in disease states compared to those derived from normal cells resulting in cancer microvesicles containing altered miRNA profiles. Tumor derived microvesicles have been known to enter neighboring cells including the surrounding stroma, and have been detected in virtually all types of biofluids. Our research shows that miRNAs transferred from lung adenocarcinoma cells through microvesicles can influence gene expression in endothelial cells and increase their ability to form new blood vessels. **Materials and Methods:** Using 5 lung adenocarcinoma cell lines H1395, H1437, H2073, H2228 and H2347 we isolated microvesicles using differential ultracentrifugation. RNA was extracted from the microvesicles as well as the cells from which they were derived. RNA was profiled for 742 miRNAs using the miRCURY LNA™ Universal RT miRNA PCR system to identify miRNAs that were enriched in the microvesicles. Isolated microvesicles were also co-cultured with endothelial cells (HMEC-1) to assess the ability of the cells to form capillary like structures. **Results:** An enrichment of a select set of miRNAs within the lung adenocarcinoma microvesicles were identifed. These miRNAs have previously been identified as tumor suppressors: miR-142-3p, miR-143-3p, miR-144-3p, miR-145-5p, miR-150-5p, miR-223-3p, miR-451a, miR-486-5p, miR-605-5p. When exosomes from miRNA over expressing adenocarcinoma lines were incubated with HMEC-1 cells an increase in their ability to form new blood vessels was observed. Several of these miRNAs were also found to be enriched in serum samples draining directly from lung adenocarcinoma tumors compared to arterial serum. **Significance:** Identifying key miRNAs in cancer-stromal signaling will allow for us to better understand this interaction, and potentially identify targets for novel therapeutics in order to better combat this disease. .

2975T

Parent of origin effect of hypomorphic mutations explains low penetrance in retinoblastoma not due to mosaicism of full mutations. *V. Imperatore[1], A.M. Pinto[1,2], S. Amitrano[2], S. Somma[1,2], M. Bruttini[1,2], E. Trevisson[6], T. Hadjistilianou[3], S. de Francesco[3], P. Toti[4], E. Gusson[5], M.A. Mencarelli[1,2], F. Mari[1,2], A. Renieri[1,2], F. Ariani[1,2].* 1) Medical Genetics, Department of Medical Biotechnology, University of Siena, Policlinico 'Santa Maria alle Scotte', SIENA, Italy; 2) Medical Genetics, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) Unit of Ophthalmology. And Retinoblastoma Referral Center, Department of Surgery, University of Siena, Policlinico 'Santa Maria alle Scotte', Siena, Italy; 4) Department of Human Pathology and Oncology, Section of Pathology, University of Siena, Policlinico 'Santa Maria alle Scotte', Siena, Italy; 5) Unit of Ophthalmology, Azienda Ospedaliera Universitaria Integrata, Verona, Italy; 6) Departement of Woman and Child Health, University of Padova Istituto di Ricerca Pediatrica, IRP, Città della Speranza, Padova, Italy.

   Incomplete penetrance in retinoblastoma (RB), one of the more extensively studied pediatric cancer, is still an unsolved question. In 2009, it has been demonstrated a ~ 3 fold excess of the *RB1* maternal canonical transcript due to an imprinting mechanism. As a consequence, maternally inherited hypomorphic mutations may retain sufficient suppressor activity to prevent tumor development. Here, we change the perspective of retinoblastoma field demonstrating a differential allelic expression of hypomorphic variants that significantly influence the phenotypic outcome in families at risk for RB onset. In one family, the father who presented with retinoma and the two children who developed retinoblastoma were all identified harboring an hypomorphic splicing variant (c.2663+2T>C). RT-PCR quantitative analysis showed a higher expression of the abnormally spliced product in the father along with an inversion of the ratio between the two differentially spliced isoforms. A parent-of-origin-effect was demonstrated for another hypomorphic mutation (p.Arg661Trp) able to induce RB in a child of a male carrier with an anamnestic history positive for osteosarcoma. For the first time, here we also identified somatic mosaicism as important factor that can impact the phenotype expression. Indeed, segregation analysis performed by high-depth NGS assay in other families, showed a ~ 50% of mutated molecules in the proband with retinoblastoma in comparison with a 25% in one parent with retinoma, accordingly with a somatic mosaicism. These findings underline the importance of performing high-sensitivity NGS for RB diagnosis and completely change the genetic and prenatal counselling and the clinical management of the families.

2976F

Mapping of differential allelic protein binding in eQTL SNPs associated with prostate cancer. *L. Tillmans[1], S. McDonnell[1], P. Zhang[2], M. Larson[1], S. Riska[1], Z. Fogarty[1], N. Larson[1], S. Baheti[1], D. Schaid[1], L. Wang[2], S. Thibodeau[1]. 1) Mayo Clinic, Rochester, MN; 2) Medical College of Wisconsin, Milwaukee, WI.*

In a previous study, we created a normal prostate tissue eQTL database using 471 tissue samples from patients with low Gleason grade prostate cancer, from which we identified 88 genes from 51 risk regions with eQTL signals. To identify possible functional SNPs for each of the risk-regions, we utilized a novel assay to test differential allele binding of each selected SNP to proteins in LnCap cells with and without androgen treatment. We chose 28 of our 51 eQTL positive risk regions for further study. For 24 of these regions, all significant eQTL SNPs that had r2 >0.5 with the risk SNP were examined along with a few from the remaining 4 regions (total number of SNPs tested was 1274). After quantifying the oligos that were bound to proteins by DNA sequencing, the ratio of oligo counts was determined and normalized with input ratios from a library of oligo mix. After excluding SNPs with low confidence, there were 683 high quality SNP sites (53%) representing 27 regions for further analysis. We found that low confidence SNP sites were correlated with low GC% (<40%). Significant allelic ratios were defined as less than 0.65 or greater than 1.5 (range 0.11-10.9). For androgen treated LnCap cell lines, 37% of the SNPs showed a significant ratio difference and untreated cell lines had 43%. We also compared the allelic ratio between the treated and untreated nuclear extracts and found 9% had significant ratio differences indicating a functional change under androgen stimulation. Finally, we looked at the oligos containing the SNP in programs such as MATCH ™ to find specific transcription binding proteins at these motifs. For example, the SNP rs113082846 on chromosome 10 shows a strong eQTL signal with the gene *NCOA4* [MIM: 601984] (p value 5.32 E-52). NCOA4 is an androgen receptor coactivator. In our binding analysis, the ratio of alternate allele C to reference G allele was 8.25 in androgen treated cell line nuclear extract and 10.09 in untreated extract. When the experimental oligo sequence with the C allele was entered in MATCH™, a binding site for ELK-1 (transcription activator) was reported with a core match of 1.000 and a matrix match of .988. No binding sites were reported for the G allele. Utilizing our experimental data along with bioinformatics tools provides a strategy for identifying candidate functional elements for prostate cancer susceptibility loci to help guide subsequent laboratory studies.

2977W

Onco-exaptation: Endogenous retroviral LTRs as promoters of oncogenic transcripts in cancer. *A. Babaian[1,2], Y. Zhang[1,2], L. Gagnier[1,2], M.T. Romanish[1,2], D.L. Mager[1,2]. 1) Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada; 2) Medical Genetics, University of British Columbia, Vancouver, BC, Canada.*

Exaptation is the incorporation of transposable elements (TEs) into functional, and in some cases necessary, genes or regulatory units over evolutionary time. We postulate that an analogous process occurs in oncogenesis, where TE-derived promoters generate "noisy" transcription and novel transcripts, which can be exploited and selected for to drive cancer transcriptome evolution. Such "onco-exaptation" of TE promoters occurs in Hodgkin Lymphoma (HL), resulting in expression of the oncogenes *CSF1R* and *IRF5*, but it is unclear how widespread this mechanism is. We hypothesize that epigenomic dysregulation in cancer leads to a genome-wide derepression of TE-initiated transcripts, some of which can play an oncogenic role. To begin to test this hypothesis, we developed a computational tool called 'LIONS' to analyze RNA-sequencing data for TE-initiated transcripts, specifically those driving the expression of coding genes or long non-coding RNAs. *LIONS* detects and quantifies TE-initiated transcripts through transcriptome assembly, applies a novel artificial neural network classifier to identify TE promoter events, and compares biological sets of data. Using this tool, we determined that the HL transcriptome has an overall increase in TE-initiated transcripts relative to non-cancerous B-cell controls. This increase is specifically driven by an increase in endogenous retroviral long terminal repeat (LTR) initiated transcripts in the HL transcriptome relative to the B-cell control (148.3 N =12 vs. 90.2 N=9 respectively, t-test P = 0.0024). The sets of TE-initiated transcripts across HL samples are more diverse compared to B-cells. However, certain TE-initiated transcripts are exceptionally recurrent and specific to HL, including the known LTR-driven oncogene transcripts of *CSF1R* and *IRF5,* as well as the enticing cases of *RALB* and *IL1R2* (reported to be greatly overexpressed in HL). A similar phenomenon of LTR derepression was observed in cells treated with the DNA demethylating agent 5-azacytidine, in cells in replicative senescence and in other cancers such as diffuse large-B-cell lymphoma and colorectal carcinoma, relative to their respective controls. We propose that this process of onco-exaptation is a distinct mechanism for oncogene activation and a model for studying exaptation and transcriptome evolution.

2978T

A novel isoform of IL-33 revealed by screening for retroelement activated genes in colorectal cancer. *F.E. Lock[1,2], Y. Zhang[1,2], L. Gagnier[1,2], A. Babaian[1,2], S. Kuah[1,2], D.L. Mager[1,2].* 1) Terry Fox Laboratory, British Columbia Cancer Agency, 675 West 10th Avenue, Vancouver, BC, V5Z1L3, Canada; 2) Department of Medical Genetics, University of British Columbia, 675 West 10th Avenue, Vancouver, BC, V5Z1L3, Canada.

Remnants of transposable elements (TEs) are abundant in the human genome. TEs, particularly long terminal repeats (LTRs) of endogenous retroviruses, contain enhancers and promoters and are capable of influencing genes. TEs can be released from epigenetic repression and become transcriptionally active in cancer, which can lead to activation of oncogenes, as described in two types of lymphoma. However, there are no reports of this mechanism in solid cancers. Here, we analyzed RNA-Seq data from 66 patients with colorectal cancer (CRC), compared to matched normal colon, to detect genes expressed through activation of TE promoters. We found an overall increase in LTR-initiated chimeric transcripts in CRC. Among these, we identified six such transcripts, where the gene has some known role in cancer, that are expressed in CRC but not normal colon, including an LTR-driven form of the ion transporter *SLCO1B3*, previously reported as cancer-specific. We confirmed expression of the six transcripts in CRC cell lines and further characterized one involving Interleukin 33 (LTR-*IL-33*) that is expressed in a subset of CRC samples through use of an ancient LTR promoter in an intron. The LTR promoter is silent in normal colon but its usage in CRC, combined with the native promoter, leads to an overall increase in *IL-33*. IL-33 is thought to play a role in solid tumors, including CRC, where it promotes polyposis and a tumorigenic microenvironment. IL-33 functions as a cell-free cytokine and also in the nucleus, where it acts as a transcriptional repressor through its N-terminal homeodomain-like region (amino acids 1-65). The LTR-*IL-33* transcript encodes a novel chimeric isoform of the protein, which is N-terminally truncated, missing the first 41aa, including highly conserved motifs. LTR-IL-33 is expressed in 3 out of 12 CRC cell lines tested, where its activity is linked to DNA hypomethylation. Transient LTR-IL-33 depletion using siRNA in these lines did not result in detectable changes in cell morphology or proliferation. LTR-IL-33 expression is regulated by cell confluence and is found in the nucleus of expressing cells, similar to native IL-33. Future studies will evaluate LTR-IL-33 function as a transcriptional regulator of the known native IL-33 target NF-κB p65, and assess LTR-IL-33 extracellular release and tumorigenicity of expressing cells. Taken together, these data demonstrate the significance of TEs as regulators of aberrant gene expression in colon cancer.

2979F

Polymorphisms of the *XRCC3* and *XPD* genes and risk of head and neck squamous cell carcinoma. *K. Chukka[1], Z. Vishnuvardhan[2], U. Radhakrishna[3].* 1) Biotechnology, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India; 2) Department of Botany and Microbiology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, India; 3) Department of Obstetrics and Gynecology, Oakland University- William Beaumont School of Medicine, Royal Oak, MI, USA.

The actual burden of Head and Neck Cancer (HNC) in India is much greater than previous literature suggested, as evidenced by the recent reports of *'Net-based Atlas of Cancer in India'*. Southeast Asia is likely to face a sharp increase of over 75% in the number of cancer deaths by 2020 as compared to 2000. DNA-repair genes play an important role in protecting individuals from cancer-causing agents. Single-Nucleotide Polymorphisms (SNPs) are the most common form of genetic variation found in the human population. The aim of this study was to evaluate the association between polymorphisms in *XPD* (Lys757Gln) and *XRCC3* (Thr245Met) with cases with head and neck squamous cell carcinoma (HNSCC). Sixty cases with HNSCC were selected from the Department of Radiotherapy, Government General Hospital, Guntur, AP, India. An age and sex-matched control group of healthy subjects (n=60) was used to compare the frequency of polymorphic variants. We studied the risk of HNSCC in relation to common non-synonymous SNPs in two DNA repair genes: *XPD* (Lys757Gln) and *XRCC3* (Thr245Met). DNA was extracted from peripheral blood. SNPs were genotyped by direct sequencing. Results: No differences were observed among the studied groups with regard to the genotype distribution of *XPD* (Lys757Gln), and *XRCC3* (Thr245Met). 79 % of the HNSCC cases were heterozygous and only 21% of the cases were homozygous. All the controls were heterozygous. Conclusion: *XPD* (Lys757Gln), and *XRCC3* (Thr245Met) may not be useful as biomarkers for HNSCC. The analysis of other genetic factors along with this parameter may provide the gateway for personal risk assessment and an opportunity for individualized therapy. We suggest that large and well-designed studies of common polymorphisms in DNA repair genes are needed. Such studies may benefit from analysis of polymorphisms as well as from the consideration of relevant exposures that may influence the likelihood of HNSCC when DNA repair capacity is reduced.

2980W

Autism linked to increased oncogene mutations but decreased cancer rate. *B. Darbro[1,2,3,4], R. Singh[1,5], M. Zimmerman[6], V. Mahajan[7,8], A. Bassuk[1,2,3,9,10,11,12]*. 1) Department of Pediatrics, Division of Medical Genetics, University of Iowa, Iowa City, Iowa, United States of America; 2) Interdisciplinary Program in Genetics, University of Iowa, Iowa City, Iowa, United States of America; 3) Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa, United States of America; 4) The Holden Comprehensive Cancer Center, University of Iowa, Iowa City, Iowa, United States of America; 5) Department of Pediatrics, Division of Pediatric Hematology/Oncology/BMT, University of Iowa, Iowa City, Iowa, United States of America; 6) Department of Biostatistics, University of Iowa College of Public Health, Iowa City, Iowa, United States of America; 7) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa, United States of America; 8) Department of Biology, University of Iowa, Iowa City, Iowa, United States of America; 9) Department of Pediatrics, Division of Neurology, University of Iowa, Iowa City, Iowa, United States of America; 10) Interdisciplinary Graduate Program in Molecular and Cellular Biology, University of Iowa, Iowa City, Iowa, United States of America; 11) Interdisciplinary Graduate Program in Neuroscience, University of Iowa, Iowa City, Iowa, United States of America; 12) University of Iowa eHealth and eNovation Center, University of Iowa, Iowa City, Iowa, United States of America.

Autism spectrum disorder (ASD) is one phenotypic aspect of many monogenic, hereditary cancer syndromes. Pleiotropic effects of cancer genes on the autism phenotype could lead to repurposing of oncology medications to treat this increasingly prevalent neurodevelopmental condition for which there is currently no treatment. To explore this hypothesis we sought to discover whether autistic patients more often have rare coding, single-nucleotide variants within tumor suppressor and oncogenes and whether autistic patients are more often diagnosed with neoplasms. Exome-sequencing data from the ARRA Autism Sequencing Collaboration was compared to that of a control cohort from the Exome Variant Server database revealing that rare, coding variants within oncogenes were enriched for in the ARRA ASD cohort ($p < 1.0 \times 10^{-8}$). In contrast, variants were not significantly enriched in tumor suppressor genes. Phenotypically, children and adults with ASD exhibited a protective effect against cancer, with a frequency of 1.3% vs. 3.9% ($p < 0.001$), but the protective effect decreased with age. The odds ratio of neoplasm for those with ASD relative to controls was 0.06 (95% CI: 0.02, 0.19; $p < 0.0001$) in the 0 to 14 age group; 0.35 (95% CI: 0.14, 0.87; $p = 0.024$) in the 15 to 29 age group; 0.41 (95% CI: 0.15, 1.17; $p = 0.095$) in the 30 to 54 age group; and 0.49 (95% CI: 0.14, 1.74; $p = 0.267$) in those 55 and older. Both males and females demonstrated the protective effect. These findings suggest that defects in cellular proliferation, and potentially senescence, might influence both autism and neoplasm, and already approved drugs targeting oncogenic pathways might also have therapeutic value for treating autism.

2981T

RNA-seq analysis of Tmed2[99J/WT] mice with NAFLD and liver tumor. *A. Djedid[1], J. Majewski[1], L. Majewska[2]*. 1) McGill University and Genome Quebec Innovation Center, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

Non-Alcoholic Fatty Liver Disease (NAFLD) is characterized by the accumulation of extra fat in liver cells not caused by alcohol. This disease can evolve in cancer and is becoming a major cause of hepatocellular carcinoma in the US. Recently, Dr. Majewska research laboratory identified a null allele of Transmembrane emp24 domain-containing protein or p24 (TMED2/p24) in an N-ethyl-N-nitrosourea (ENU) mutagenesis screen for recessive developmental abnormalities, in which 99J homozygous mice embryos present developmental delay, abnormal rostral-caudal elongation, randomized heart looping, absence of the labyrinth layer of the placenta, and death at mid-gestation. It has been reported that TMED2/p24 regulates protein-cargo selection and vesicle budding. TMED2/p24 is also required for liver homeostasis and liver development. In the current study, we investigate mechanisms underlying NAFLD and liver tumor pathogenesis in 99J heterozygous mice. We performed an integrated analysis on liver samples from four C3H/HeJ mice and four Tmed2[99J/WT] mice with NAFLD and tumor in order to identify genes that are differentially expressed between our controls and NAFLD/liver tumor samples, as well as fusions and splicing events. To achieve this, we built an in-house pipeline that uses latest efficient (open-source) RNASeq tools: STAR, HTSeq and DESeq for expression analysis, STAR-Fusion and GMAP-Fusion for fusion detections, then rMATS, MISO and Cufflinks for splicing event detections. David Pathway Analysis was used to determine dysregulated pathways in Tmed2[99J/WT] mice. In tumor samples, more than 750 genes were dysregulated. David analysis on these samples showed an up-regulation of genes implicated in cancer pathways. The Splicing analysis demonstrated a difference in RNA isoform expression between control and tumor samples of genes implicated in processes related to cancer such as cell death, immune system processes, protein metabolic processes, cell proliferation.

2982F

Generation of PCS (MVA) syndrome mutation knock-in mice using CRISPR/Cas9-mediated genome editing technology. *K. Hosoba, T. Miyamoto, S. Matsuura. Department of Genetics and Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan.*

Germline mutations of *BUB1B* gene encoding BUBR1, a central molecule of mitotic spindle assembly checkpoint, cause an autosomal recessive disorder PCS (MVA) syndrome ([MIM:176430]), which is characterized by constitutional aneuploidy, a high risk of childhood cancer and ciliopathy disease spectrum, including polycystic kidney and Dandy Walker anomaly. We previously reported that seven Japanese patients were compound heterozygotes for an exonic null mutation and an intergenic mutation located 44 kb upstream of the *BUB1B* transcription start site. It was reported that *BUB1B* knockout mice (*BUB1B-/-*) were embryonic lethal (Baker et al, 2004), while *BUB1B* hypomorphic mice (*BUB1B$^{H/H}$*) with 90% reduction in BUBR1 protein expression showed cellular senescence in most organs to induce accelerated aging and did not show cancer susceptibility. To validate the phenotypic difference between PCS (MVA) syndrome patients and *BUB1B$^{H/H}$* hypomorphic mice, here we generated PCS ;(;MVA); syndrome mutation knock-in mice. Co-microinjection of Cas9 mRNA, guide RNA for exon15 of *BUB1B*, and single stranded oligodeoxynucleotides (ssODN) with c.1833delT (F611fsX625) *BUB1B* mutation as a targeting donor into mouse fertilized eggs enabled to introduce the human mutation into mouse genome quickly. During this process, we also obtained a mouse carrying a 12-bp deletion (Δ608-611 LAST) in exon 15 of *BUB1B,* which was likely a hypomorphic mutation. These mutations were transmitted into germline. Currently we are crossing these mice to generate PCS (MVA) syndrome model mice.

2983W

Analysis of NF-kB activation compared to seric levels and genic expression of TNF-alfa in patients with sporadic colorectal cancer. *U. Santana[1,4], E. Torres-Anguiano[1,4], L. Bobadilla-Morales[1,2], B. González-Quezada[1], V. Maciel Gutiérrez[3], M. Centeno Flores[3], J. Corona Rivera[1], H. Pimentel Gutiérrez[2], C. Ortega de la Torre[2], A. Corona Rivera[1,2]. 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo", Instituto Dr. Enrique Corona Rivera", CUCS, U de G; 2) Unidad de Citogenética y Servicio de Colon y Recto, Nuevo Hospital Civil de Guadalajara, "Juan I. Menchaca; 3) Servicio de Colon y Recto, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca; 4) Doctorado en Ciencias en Biología Molecular en Medicina, CUCS, UdeG.*

The transcription factor NF-kB regulates several genes associated with cancer development, constitutive activation in colorectal cancer (CRC) patients has been reported. TNF- α is the main molecule associated with NF-kB activation. The aim of this study was to determine NF-kB activation status compared with *TNF-α* protein and mRNA levels in tumor tissue and normal adjacent tissue from the same patient with sporadic CRC, and analyze the mRNA levels of genes regulated by NFkB (*BIRC5, Bcl-2, c-Myc, TNF, PTGS1, NFKBIA, LOX, NFkB1, VEGF y CTNNB1*). Patients were classified into 2 groups: CRC advanced stages (CAS= III and IV, n = 11) and CRC early stages (CES= I and II, n = 9). Tumor and normal tissue were collected during surgery; we obtained cells for the NF-kB activation analysis, evaluating the p65 subunit phosphorylation by flow cytometry. The samples also underwent to RNA extraction for later cDNA synthesis. qRT-PCR were performed for *TNF-α, BIRC5, Bcl-2, c-Myc, TNF, PTGS1, NFKBIA, LOX, NFkB1, VEGF* and *CTNNB1* genes, we use the *GUS, ABL* and *ACTB* constitutive genes for assay validation. Comparative Ct method (ΔCt) was used to analyze the mRNA levels.Venous blood samples (5–10 mL) were collected in vacutainer tubes from CRC patients (CP) and controls (CTRL), serum was obtained and stored at −70°C until processed using ELISA. p65 imunostaining showed greater activation of NF-kB in tumor than in normal tissue, this were increased during CRC progression. *GUS* and *ACTB* were validated for CRC expression assays, we showed that *ABL* can be used in similar studies. All analyzed genes presented mRNA overexpression in tumor compared to normal tissue, although significant differences were observed just in *BIRC, PTGS. ALOX, TNF* and *NFKBI* genes, it was gradually increased during CRC progression. TNF- α serum levels were higher in CP compared to CTRL (11.18; 1.85 pg/ml), and CES presents elevated levels compared to CAS (15.3; 7.05 pg/ml). Our findings suggest that NF-kB is greatly activated in CRC tissue compared to normal tissue and it correlates with tumor progression. Simultaneously, NF-kB regulated genes exhibit overexpression in tumor compared to normal tissue;could be due to their activity in CRC processes. TNF-α serum levels were elevated in CRC patients, showing higher levels in CES, possibly due to inflammation.

2984T

Animal models support *TNS1* gene involvement in wild type GIST development. *N. Settas[1], F. Faucz[1], T. Silva[1], E. Rapizzi[2], M. Mannelli[2], S. Reincke[1], A. Sen[3], R. Cox[3], M. Miettienen[4], S. Lo[5], J.A. Carney[6], C.A. Stratakis[1].* 1) Section on Endocrinology & Genetics (SEGEN), NICHD, NIH, Bethesda, MD, USA; 2) Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy; 3) Department of Biochemistry and Molecular Biology, Uniformed Services University, Bethesda, MD, USA; 4) Department of Diagnostic Pathology, National Cancer Institute (NCI), NIH, Bethesda, MD, USA; 5) epartment of Biochemistry and Molecular Medicine University of California - Davis Sacramento, CA, USA; 6) Laboratory of Pathology, Mayo Clinic, Rochester, MN, USA.

Tensin 1 (*TNS1*) is a cytoplasmic protein that is located on the integrin-mediated cell-basement membrane junctions and is implicated in cytoskeletal organization, cell migration, and proliferation. It is known that *tns1* knockout (KO) mice develop polycystic kidney disease, while the female mice also exhibit reduced fertility. *Drosophila Melanogaster Tns1* KO ("blistery" fly) develop defective wing unfolding and subsequent blister formation. Genome-wide association studies have linked *TNS1* to lung disease, asthma and colorectal cancer, and TNS1 expression is reduced or absent in several prostate and breast cancer cell lines. Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal neoplasms of the gastrointestinal tract. Most GISTs carry activating mutations in the *KIT* and/or *PDGFRA* genes; some cases are attributed to germline mutations in succinate dehydrogenase subunit genes (SDHx). However, up to 10% of GISTs do not have mutations in the above genes and are known as wild type (WT). Patients with WT GISTs show an epigenetic down-regulation of *SDHC* gene and abnormal mitochondrial structure. In this study, we screened 24 patients with WT GISTs for *TNS1* gene mutations and we identified 12 variants predicted to be damaging after *in silico* analysis. Fifty percent (12/24) of the patients exhibited at least one *TNS1* variant. Six variants were novel, 4 of which were missense (p.Y327C, p.G1306R, p.G1227D, and p.R984W); in addition, we found an insertion (c.1978insCAGCAG) and a somatic frameshift mutation (p.P890Qfs*96). We also found 6 previously described variants (one aminoacid deletion, one splice site mutation and 4 missense variants) with low frequency in the general population. We studied the blistery flies and the *tns1* KO mice: we identified in both structural mitochondrial defects, which are consistent with those described in patients with WT GISTs. Additionally, as observed in patients, we noted significantly lower levels of SDHB expression in the blistery flies. The KO flies had also significantly delayed SDH activity. We conclude that *TNS1* genetic variants are frequent in WT GISTs; both mouse and fly models of *tns1* inactivation indicate that *TNS1* gene may be important in regulating mitochondrial SDH function.

2985F

Influence of *CYPA4*1B*, *CYP3A5*3*, *GSTT1* and *GSTM1* polymorphims with treatment-related chronic myeloid leukemia. *E.P. Silveira-Lacerda[1], K.A. Delmond[1], D.C. Abreu[1], T.M. Pinto[1], R.M. Golveia[1], A.P. Barbosa[2], R.S. Tavares[2].* 1) Genetics Department, Federal University of Goias, Goiânia, Goiás, Brazil; 2) Hematology Department. Federal University of Goias, Goiânia, Goiás, Brazil.

Imatinib Mesylate (IM) (Gleevec™ [STI571]; Novartis Pharmaceutical Corporation, East Hanover, NJ), is used in Brazil as a first- line drug for Chronic Myeloid Leukemia (CML) treatment. The IM act as potent and selective inhibitor of BCR–ABL tyrosine kinase activity, however cases of treatment failure or of suboptimal response are sometimes seen. The cytochrome-P450 (CYP) system is involved in the oxidative metabolism of IM and the major pathway is catalyzed by CYP3A4/5. Further, phase II enzymes Glutathione S-transferases (GSTs) are involved in the metabolism of a wide range of xenobiotics, including chemotherapeutic drugs. *CYP3A4*, *CYP3A5* and *GST*s are polymorphic genes, thus genetic variations these genes could influence the treatment response to IM. This is the first study of CML patients Goias (Brazil) residents in order to associate the genotype of patients for polymorphims in metabolic genes with cytogenetic and molecular responses during treatment with IM. The study was conducted in accordance with the Declaration of Helsinki and was approved by a recognized ethics committee (*Comitê de Ética em Pesquisa do Hospital das Clínicas da Universidade Federal de Goiás*). A total of 114 CML patients in use of IM from Hospital das Clínicas-UFG, Goiânia, Goiás, Brazil, were enrolled in this study. Response criteria were based on European Leukemia Net recommendations. Genotyping was carried out for *CYP3A4*1B* (-392A>G (rs2740574)), *CYP3A5*3* (6986A>G (rs776746)), *GSTT1*0* (null allele) and *GSTM1*0* (null allele) polymorphisms by q-PCR. Statistical analysis was carried out using GraphPad Prism5® software. The X_2-test or Fisher's exact test was used to compare response and specific genotypes (α=0,05). The results revealed no statistical difference between evaluated polymorphisms *CYP3A4*1B, CYP3A5*3, GSTT1*0* and *GSTM1*0* and responses to treatment with IM (*P*>0,05). The same result was seen when comparison was performed with polymorphisms grouped for each patient (*P*>0,05). These findings have shown that in the studied population the polymorphisms *CYP3A4*1B, CYP3A5*3, GSTT1*0* and *GSTM1*0* do not influence the performance efficiency of the drug (IM). Subsequent investigations to other genes involved in IM metabolizing pathway should be performed.   .

2986W

Effects of G6PD activity inhibition on the viability, migration and morphology of cervical cancer cells. *F. Zishui, J. Chengrui, F. Yi, C. RiXin, L. XiaoYing, Z. ZhiQiang, H. LuHao, C. XiaoDan, Li. HongYi, G. YiBin, J. WeiYing.* Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen Univer, GuangZhou, China.

Glucose-6-phosphate dehydrogenase (G6PD) deficiency has been revealed to be involved in the efficacy to anti-cancer therapy but the mechanism remains unclear. We aimed to investigate the anti-cancer mechanism of G6PD deficiency. In our study, Dehydroepiandrosterone (DHEA) and shRNA technology were used for inhibiting the activity of G6PD of cervical cancer cells. Flow cytometry was used to detect the apoptosis and reactive oxygen species (ROS) generation. Scanning electron microscopy was used to observe cell morphology. And the migration rate of cell was also observed using the imaging system of inverted living cell microscope. When G6PD was inhibited by DHEA or RNA interference, idioblasts were observed in HeLa cells. Simultaneously, G6PD deficiency resulted in decreased HeLa cells migration and proliferation ability but increased ROS generation inducing apoptosis. What's more, the inhibition of G6PD activity caused filopodia and reticulopodium of cells shorten and reduce as well as cells shrinkage. Our results indicated the anti-cervix cancer mechanism of G6PD deficiency may be involved with the decreased cancer cells proliferation caused by the increased ROS and decreased migration ability as a result of abnormal filopodia and reticulopodium of cells. Suppression of G6PD may be a promising strategy in developing novel therapeutic methods for cervical cancer.

2987T

Role of a p53 polymorphism in the development of nonfunctional pituitary adenomas. *G. Yagnik, A. Jahangiri, R. Chen, J. Wagner, M. Aghi.* Dept of Neurosurgery, University of California San Francisco, San Francisco, CA.

Non-secreting pituitary adenomas are typically benign, non-familial tumors with estimated prevalence of 17%, making them the most common primary brain tumor. Their evolution varies from indolent slowly growing tumors to tumors that, while histologically benign, often reach sufficient size to limit their therapeutic options and destroy patients' quality of life. Investigations of the molecular etiology of adenomas have failed to identify prevalent genetic changes and, while a role for p53 has been suggested, to date no study on pituitary adenomas has performed a comprehensive search for genetic aberrations in the TP53 gene. We therefore performed the first complete exonic sequencing on a cohort of 35 non-functioning pituitary adenomas with paired lymphocytes obtained from blood. While no somatic mutations were discovered between the blood and adenoma tissue, we did note a significant frequency of a previously reported polymorphism associated with other cancers, rs1042522, in our cohort. The functional effect of this variant, however, has thus far been conflicting and cell-type specific. Population analysis showed most ethnic groups had a C allele in codon 72 leading to a proline residue (P72). In contrast, our pituitary adenoma cohort showed a higher prevalence of the G nucleotide in the same location, leading to an arginine residue (R72). The allelic frequency of R72 in our cohort (75.7%) showed a statistically significant difference as compared to control populations matched to the self-reported ethnicity of our pituitary patients (34.6%) ($p<1.09 \times 10^{-9}$), a far greater association than demonstrated for this polymorphism in other cancers. To understand the potential implications of the P72 and R72 residues in pituitary adenoma, we performed functional analyses examining the downstream effects of both variants, including transcription, proliferation and migration assays, as well as a pulse-chase assay to determine changes in protein half-life. The results show that the R72 variant reduces expression of cell arrest gene p21 and increases adenoma cell proliferation with increased protein survival as compared to P72. Taken together, these data suggest that the R72 variant survives longer and can drive an increased level of proliferation through reduced expression of p21, which normally modulates cell growth. This increased proliferation in pituitary cells may be responsible for the growth of the adenoma to surgical proportions.

2988F

Mitochondrial matrix HtrA Proteases plays key role in tumorogenesis and regulates mitochondrial gene expression in colorectal and prostate cancer cell lines. *N. Fatima[1], M.W. Hameed[1,2].* 1) Panjwani Centre for Molecular Medicine And Drug Research, International Centre For Chemical And Biological Sciences, University Of Karachi, Karachi, Sindh, Pakistan; 2) Molecular Biology Lab, MaxPlank, Germany.

The High Temperature Requirement A (HtrA) proteins are serine proteases that plays a significant role in quality control of cellular proteins in variety of organisms. They are mitochondrial chaperones, has role in oncogenesis, neurodenerative disorders, and apoptosis. Loss of HtrA function accumulates unfolded proteins, disrupts mitochondrial respiration, generates reactive oxygen species and leads to the proliferation of tumor cells while over expression of this family of proteases triggers cell death. Human HtrA proteases are of four types, HtrA1, HtrA2, HtrA3 and HtrA4. HtrA2/Omi resides in the mitochondrial intermembrane space (IMS), its paralogs HtrA1, 3 and 4 are most likely targeted to the secretory pathway. In the current study, we have used colorectal cancer cell lines (HCT-116) and prostate cancer cell lines (PC3) to analyze if this ATP depended protease family (HtrA and other proteases of the interacting pathway) effects mitochondrial genome expression and its regulation in this cancer. Prostate cancer is the most common non-dermatologic cancer of males and is a leading cause of death in Pakistani males. The incidence rate of colorectal cancer is also increasing worldwide, having relatively high incidence in western countries than in Asia and Africa. It's the seventh most common cancer in women and ninth most common cancer in men of Pakistan. We have induced the expression of these proteases in these cell lines through heat mediated stress and drug induced ER stress and then monitored the transcription level of these proteases through Real Time PCR. Protein level of these proteases are observed through western blotting. Reactive oxygen levels were also observed after induction of these proteases. Moreover, the effect of these proteases on mitochondrial gene expression was studied and interesting results were obtained. Overexpression of these proteases correlates with increase in transcription of mitochondrial genome and therefore, we can regulate the mitochondrial genetic machinery by targeting these proteases.

2989W

Dissecting TRIM8 role in the pathogenesis of glioma. *S. Venuto[1], M. Monti[2], I. Appolloni[3], T. Mazza[5], P. Pucci[2], P. Malatesta[3,4], G. Merla[1], L. Micale[1].* 1) IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, FG, Italy; 2) CEINGE Advanced Biotechnology and Department of Organic Chemistry and Biochemistry, Federico II University, Napoli, Italy; 3) U.O. Trasferimento Genico, IRCCS-AOU San Martino-IST, Genoa, Italy; 4) IRCCS AOU San Martino – IST, University of Genova, Italy; 5) Bioinformatics Unit, IRCCS Casa Sollievo della Sofferenza, Istituto Mendel, Italy.

Gliomas are among the most lethal brain tumors with a still unclear molecular pathogenesis. This study investigates the role of E3 ubiquitin ligase TRIM8, a member of tripartite motif (TRIM) protein family, in the pathophysiology of glioma. Our preliminary data corroborate mounting evidence showing a down regulation of TRIM8 in glioma and its role in controlling cell growth. Many experimental evidences showed that some TRIMs through their E3 ubiquitin ligase activity are involved in oncogenic processes by proteolytic degradation of regulatory proteins, including those involved in cell proliferation and apoptosis. We reasoned that defects in TRIM8 E3 ligase activity might promote carcinogenesis and cancerous growth in glioma cells by contributing to oncogenes stabilization, oncosuppressors degradation, and tumor-related pathway deregulation. To assess this hypothesis we profiled Trim8-related protein and transcript signatures in Neural Stem Cells overexpressing Trim8 (NSC-Trim8), by Mass Spectrometry and RNAseq. Proteomic analysis identified 57 Trim8-related interactors and/or substrates including four members of the kinesin-like protein family, known to be involved in various kinds of mitotic spindle dynamics and drivers of invasion, proliferation and self-renewal in glioblastoma. Co-immunoprecipation confirmed Trim8-kinesin interaction. Global gene expression profile revealed the most enriched pathways as involved in axonal guidance signaling, synaptic long-term potentiation, Stat3 signaling and immune cell trafficking. Interestingly, a transcriptional deregulation of genes involved in mitotic spindle assembly and chromosome segregation was detected by RNASeq and validated by qPCR, in line with the proteomic data. Validation of transcriptomic data was carried out by PCR arrays kits that profile the expression of a panel of genes central to the Stat3 pathway and inflammation process. Several receptors, nuclear co-factors and co-activators of Stat proteins, Stat-inducibile genes and negative regulators of the pathway were found deregulated in NSC-Trim8 compared to the NSC control. A similar approach confirmed the deregulation of key secreted proteins central to the immune response in NSC–Trim8. Our expression and proteomics data give preliminary evidences that TRIM8 may participate in the carcinogenesis and progression of glioma through an aberrant regulation of biological pathways and protein networks known to be correlated with glioma pathogenesis.

2990T

Utilization and patient reported outcomes of clinical multi-gene panel testing for cancer susceptibility in the multi-center COGENT study. *A.R. Bradbury[1,2], L. Patrick-Miller[4], B.L. Egleston[6], S.M. Domchek[1,3], O.I. Olopade[4,5], M.J. Hall[7], M.B. Daly[7], L. Fleisher[10], G. Grana[9], P. Ganschow[8], D. Fetzer[7], The COGENT Research Team.* 1) Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA; 3) Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA; 4) Division of Hematology-Oncology, Department of Medicine, The University of Chicago, Chicago, IL; 5) Center for Clinical Cancer Genetics and Global Health, The University of Chicago, Chicago, IL; 6) Fox Chase Cancer Center, Temple University Health System, Biostatistics and Bioinformatics Facility, Philadelphia, PA; 7) Fox Chase Cancer Center, Temple University Health System, Department of Medical Genetics, Philadelphia, PA; 8) Department of Internal Medicine, The John H Stroger Jr Hospital of Cook County, Chicago, IL; 9) Division of Hematology-Oncology, MD Anderson Cancer Center at Cooper, Camden, NJ; 10) The Children's Hospital of Philadelphia Research Institute, Center for Injury and Research Prevention, Philadelphia, PA.

    Background: Multi-gene panels (MGP), including both high and moderate penetrance cancer susceptibility genes are currently being used despite questions regarding their clinical utility and no standard approach to genetic counseling. We have shown no difference in outcomes between telephone (TD) and in-person disclosure (IPD) of MGP results, but how frequent providers use MGP testing and how patient reported outcomes (PROS) differ from targeted testing has not been reported. **Methods:** We evaluated use of MGP testing and PROS in participants undergoing cancer genetic testing in the multi-center COGENT study (NCT01736345), a randomized study of TD versus IPD of genetic test results. In this "real world" study, genetic counselors offered targeted (i.e. *BRCA1/2*) or MGP considering clinical indication and insurance coverage. Participants completed PROS at enrollment (after pre-test counseling) and after result disclosure. **Results**: Since the inclusion of MGP testing in 2014, 600 participants with no prior genetic testing enrolled; 395 (66%) patients were offered MGP, 205 were offered *BRCA1/2* testing only. Offering MGP testing increased from 57% of patients in 2014 to 66% in 2015 (p=0.02), and varied by site (46-78%, p<0.01). In a regression, the most salient factors associated with being offered MGP testing were study site (p<0.01), not being of Ashkenazi heritage (OR 5.2, p<0.001), not having a known mutation in the family (OR 26.5, p<0.001), and patient request for IPD (OR 2.7, p=0.01). Patients offered MGP testing had lower genetic knowledge after pre-test counseling than those offered *BRCA1/2* testing (22.7 v. 23.6, p=0.001). 376 (95%) of patients proceeded with MGP testing. After adjusting for arm, patients receiving MGP had less increase in cancer worry (-0.07 v. +2.7, p=0.01), lower state (33.0 v. 36.1, p=0.01) and general anxiety (5.9 v. 6.6, p=0.04), higher satisfaction (40.2 v. 39.0, p=0.02) and lower post-test genetic knowledge (22.7 v. 23.7, p=0.001) than patients who received *BRCA1/2* results only. **Conclusions:** Use of MGP has increased over time. Affective outcomes are not worse after MGP, although genetic knowledge may be lower compared to *BRCA1/2* testing and some patients undergoing MGP testing express a preference for in-person disclosure. Longitudinal cognitive and behavioral outcomes are needed to better understand the clinical risks and benefits of using multi-gene cancer panels and guidelines for use of MGP testing in clinical cancer genetics.

2991F

Interest in web-based education and informed consent for return of individual genetic research results. *J.B. Gaieski[1], L. Patrick-Miller[4], B. Egleston[5], D. Fetzer[1], W.K. Chung[10,11], K. Maxwell[1], J. Churpek[9], K. Nathanson[8], O. Olopade[6], M.B. Terry[13], K. Karpink[1], S. Walser[1], A. Brandt[1], D. Clark[1], J.M. Long[1], J. Powers[1], D. McKenna[1], J. Bulafka[10], J. Martinez[10], C. Koval[12], M. Palese[7], D. Lech[7], C. Weipert[7], S. Domchek[1,3], A.R. Bradbury[1,2].* 1) Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 2) Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 3) Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 4) Division of Hematology-Oncology, Department of Medicine, The University of Chicago, Chicago, Illinois, USA; 5) Fox Chase Cancer Center, Temple University Health System, Biostatistics and Bioinformatics Facility, Philadelphia, Pennsylvania, USA; 6) Center for Clinical Cancer Genetics and Global Health, The University of Chicago, Chicago, Illinois, USA; 7) Biological Sciences Division, Section of Hematology/Oncology, Center for Clinical Cancer Genetics, University of Chicago, Chicago, Illinois, USA; 8) Department of Medicine, Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 9) Section of Hematology/Oncology, Department of Medicine, Comprehensive Cancer Research Center, University of Chicago, Chicago, Illinois, USA; 10) Department of Pediatrics, Columbia University Medical Center, New York, New York, USA; 11) Department of Medicine, Columbia University Medical Center, New York, New York, USA; 12) Department of Clinical Genetics, Columbia University Medical Center, New York, New York, USA; 13) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA.

    Background: Genetic counselors (GC) are frequently employed to conduct pre-disclosure education and return of individual research results (IRR). Given high costs and a limited GC workforce, alternative informed consent and delivery models are needed. **Methods**: In the multi-center RESPECT study, women with a personal or family history of breast cancer are offered the opportunity to learn multi-gene cancer susceptibility IRR. Participants have the choice to complete pre-disclosure education and informed consent with a genetic counselor or through a web-based intervention (WebEd). Descriptive statistics and logistic regressions were used to describe uptake of WebEd, website use characteristics, patient reported outcomes (PROS) and associated participant factors. **Results**: To date, 104 participants have completed pre-disclosure education and informed consent. 92 (88%) chose WebEd, 12 (12%) chose to speak with a GC because they did not have web access. Mean time to WebEd access was 10.3 days (SD 18, range 0-94). 35 (38%) completed required WebEd content and made a decision about receiving IRR on their first visit. Mean time to decision was 33.8 minutes (SD 22.5, range 4.3-111.8). 52 (57%) participants visited WebEd multiple times before making a decision. Mean number of visits was 3 (SD 2, range 2-12); mean total viewing time prior to making a decision was 63 minutes (SD 52, range 5-249). 7 (8%) participants asked to speak with a GC after using WedEd before making a decision. Being younger (OR 0.92, p=0.003) and having higher cancer specific distress at baseline (OR 1.05, p=0.05) were associated with making a decision on the first WebEd visit. Overall, 92 (88%) participants elected to receive IRR, 35 (100%) of those deciding at their first WebEd visit, 45 (79%) of those who visited WebEd multiple times and 12 (100%) among those who spoke with a GC. There were no significant differences in change in knowledge or distress from baseline to post WebEd between those who completed WebEd in one versus multiple visits. **Conclusions:** Web-based alternatives for genetic education and informed consent for return of IRR may be an acceptable and scalable alternative for many research participants. Further recruitment and collection of PROS after receipt of results is expected to inform the risks and benefits of using web-based alternatives for pre-disclosure education and informed consent for return of IRR.

2992W

Patient interest and risk communication preferences regarding a novel *BRCA1/2* **genetic modifier test.** *J.G. Hamilton[1], J.S. Westerman[1], M.C. Genoff[1], E. Shuk[1], J.L. Hay[1], K. Offit[2], M.E. Robson[2].* 1) Behavioral Sciences Service, Department of Psychiatry & Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.

Background: Recent research has identified a set of genetic variants that modify breast cancer risks among female *BRCA1/2* mutation carriers. By incorporating these genetic modifiers into a clinical genetic testing panel, healthcare providers may be able to more accurately predict *BRCA1/2* carriers' likelihood of developing breast cancer. This specificity would represent a substantial improvement in disease risk prediction for these patients, as currently breast cancer risk estimates by age 70 range from 44-78% for *BRCA1* carriers and from 31-56% for *BRCA2* carriers. To prepare for the future availability of genetic modifier testing, it is critical to understand the opinions of *BRCA1/2* mutation carriers about both the test and the ways that test results can be communicated. Method: We conducted in-person interviews with 30 female *BRCA1/2* mutation carriers about their perceptions, comprehension, and preferences regarding sample genetic modifier risk communication materials. Participants (aged 25-80 years; 87% Caucasian; mean numeracy score = 4.8 on a scale ranging from 1-6) were randomized to receive one of three sample genetic modifier test results (indicating low, moderate, or high breast cancer risk). Participants were shown results in four different formats (panel of genes, icon array providing social comparison information, verbal risk estimate, and graphical risk estimate) and asked to provide qualitative feedback based on a semi-structured interview guide. Results: Interest in obtaining genetic modifier testing was high, with 77% of participants indicating that they would be willing to pay out of pocket for the test. Many believed that genetic modifier test results would make their cancer risk management decisions easier, whereas a few were unsure and one indicated that it would make her decisions harder. The preferred format across the three breast cancer risk levels was the graphical estimate, alone or in combination (often with the verbal estimate). The least popular format was the icon array, which many participants deemed confusing. Conclusion: These results suggest that patients would be highly receptive to genetic modifier testing, and that the minimal amount of necessary information to include in genetic modifier test results would be graphical and verbal estimates of risk. These findings will inform future efforts to translate genetic modifier testing into care in a manner consistent with patients' preferences.

2993T

Systematic assessment and reporting of secondary findings in whole exome sequencing: The project of HOPE. *Y. Horiuchi[1,2], Y. Kiyozumi[1,4], H. Matsubayashi[2], S. Nishimura[2], K. Urakami[3], S. Oonami[3], M. Arai[1], M. Kusuhara[3], K. Yamaguchi[2,3].* 1) Dept of Psychiatry and Behavioral Sci., Tokyo Metropolitan Institute of Med. Sci., Tokyo, Japan; 2) Dept. of Genetic Counseling, Shizuoka Cancer Center, Shizuoka, Japan; 3) Research institute of Shizuoka Cancer Center, Shizuoka, Japan; 4) Graduate school of Ochanomizu University, Tokyo, Japan.

Recent advantage of whole genome sequencing (WGS) and whole exome sequencing (WES) techniques are increasingly used in both clinical and research contexts. On the other hand, new ethical issues regarding the disclosure of incidental or secondary findings, i.e., results that are unrelated to the original aims of the study may be expected to arise in the genome studies. In 2013, the American College of Medical Genetics (ACMG) announced recommendations for handling incidental findings in genome sequencing. However, the guidelines for the secondary findings and a system of the return of results for patients have not been well established yet in Japan. In this study, we systematically constructed the genetic counseling system from clinical genome study "Shizuoka Cancer Center (SCC) the Project HOPE (High-tech Omics-based Patient Evaluation)", as a basis for initial trials in Japan. The aim of Project HOPE is to evaluate biological characteristics of cancer and diathesis of each patient by multi-omics based analyses, which integrate genomics, transcriptomics, proteomics and metabolomics. We here report the prevalence of secondary findings and the system of the return of results to patient from the HOPE study. The disclosure of secondary findings for the patients were conducted carefully and systematically in SCC. To date, 1500 cases have been analyzed and approximately 3% of the total cases were reported to be associated with hereditary cancer syndromes and approximately 1% of the total cases were reported as based on the guidelines of the ACMG. Our systematic approach will contribute to a cancer genomic medicine and serve as a useful basis for genetic counseling system in Japan.

2994F

FamilyTalk: Developing a tool to help families communicate about colorectal cancer risk. *T. Hyams, D. Bowen.* University of Washington, SEATTLE, WA.

   Purpose: Relatives of people with colorectal cancer (CRC) are about twice as likely to also develop the disease. It is essential to educate both CRC patients and their families about their increased risk in order to increase screening behavior and reduce disease burden. The purpose of this poster is to present on the development of an innovative web-based tool kit called FamilyTalk with the purpose of increasing family communication about risk for developing CRC. The overarching goal is to increasing screening behavior through increasing communication in families with CRC. This poster will present on the development of FamilyTalk, findings from paper and electronic usability testing, and plans for future testing. Methods: FamilyTalk presents users with educational content about colorectal cancer and its genetic components including genetic screening, family history and treatment information. Family and provider communications sections help users navigate potentially complicated conversations about risk, and actions to be taken to reduce risk. FamilyTalk uses existing tools, such as the NCI colorectal cancer risk assessment tool (http://www.cancer.gov/colorectalcancerrisk/tool.aspx) which allows users to calculate their own risk for colorectal, and an interactive family history tool, built by this group, which allows participants to build on both familial and social relationships and create a family health history. Users can also invite members of their family and greater social network to utilize website features including allowing them to view and edit the family tree tool. We employed an 8-step process used previously in our website development research to translate our print intervention, included conceptual design, drafting web-appropriate informational and motivational content, prototyping materials, acceptability review with community members, detailed medical and genetics review, usability testing, system testing, and final modifications. Results: Colorectal families liked the paper version, but had multiple suggestions, including adding a better introduction, sections on genetics and family history, and clearer communication assistance. The web-based tool improved upon the linear book format with links and better section instructions. The processes outlined produced materials that satisfied diverse families and individuals with support. Results also include screenshots of the electronic version of FamilyTalk.

2995W

Beliefs and attitudes of Egyptian parents influencing participation in a pediatric biorepository. *R.M. Labib, O. Hassanian, M. Alaa, S. Ahmed, S. AbouEl-Naga.* Research, Children's Cancer Hospital, Cairo, Egypt.

   Background: Biobanks have become a powerful tool that can foster many types of research. The success of biobanks depends upon people's perception and willingness to donate their samples for future research. This is the first pediatric biorepository in the middle-east, hence, little is known about the beliefs and attitudes of parents towards their children participation in a research biorepository. **Aim:** To investigate the level of willingness to donate samples for research in an Egyptian Children's cancer hospital and understand factors influencing enrollment. **Materials and Methods:** A Standardized questionnaire was designed covering multiple items expected to affect the enrollment. This was conducted in-person and data collected included, demographics data, socioeconomic level, educational and religious constraints. Additionally, in the case of refusal, participants were asked about reasons for nonparticipation. **Results:** We succeeded to enroll 7,000 pediatric participants from November 2012 till March 2016. Yet, approximately 3.1% of patients have refused to participate and 0.3% have withdrawn. Three demographic factors were found having disparate trends in the parents decision making process to participate or not: father's education (p-value = 0.0001), mother's education (p-value = 0.0001) and father's age (p-value = 0.034). **Conclusion:** More intense awareness programs need to be designed specifically for parents with a higher level of education and who are of older age, as they tend to be more unwilling to participate in a research biorepository.

2996T

Pre-test genetic counseling as a requirement for germline hereditary cancer testing: What do patients do? *G.A. Lazarin, K. Sedgwick, D. Doyle, I.S. Haque, K. Ready.* Counsyl, South San Francisco, CA.

Background: To ensure informed consent and appropriate test utilization, some health insurers require individuals seeking certain hereditary cancer (HC) genetic testing to undergo pre-test genetic counseling (GC). The effect of this requirement on patient access to HC testing has not been fully studied.**Methods:** We evaluated BRCA1/2 or Lynch syndrome test orders to Counsyl for 327 members of two national health insurers to evaluate their risk status with respect to National Comprehensive Cancer Network (NCCN) guidelines for identification of individuals at high risk of an HC syndrome. NCCN was used as a proxy for insurance coverage criteria given similarities. Counsyl holds contract agreements with these payors. Patients are referred to a third-party fee-based GC service per the plan requirements. Patients were provided an estimated cost based on plan benefits and predicted risk category (high/low). All patients were offered a discounted self-pay option, regardless of insurance status. We examined patient decision making with respect to payment action, risk status, and GC action.**Results:** 262 patients (80%) met NCCN high-risk criteria based on information provided on the test requisition. In total, 29 patients (11%) canceled their tests after being informed of estimated test cost and the requirement for pre-test GC; 19 of these 29 were high-risk (7.3% of all high-risk patients). Of those continuing with testing, 110/230 (49%) high-risk and 44/65 (68%) low-risk individuals elected self-pay. 104/110 (95%) high-risk and 10/44 (77%) low-risk self payers did not seek pretest genetic counseling.**Conclusion:** 56% of patients meeting NCCN guidelines did not complete the pre-test GC requirement for insurance coverage, instead choosing to avoid the requirement through self-pay or canceling the test altogether. 7.3% of all high-risk patients canceled their testing. These data suggest the requirement for pre-test GC may not accomplish the stated goal of improving patient education, and may be seen as an impediment or unnecessary expense by patients.

2997F

Do patients receive recommended care following a diagnosis of Lynch Syndrome in an integrated health care system? *A.F. Rope[1], J.M. Zepp[1], K.R. Muessig[1], J.L. Schneider[1], K.M. Bergen[1], J.V. Davis[1], J.E. Hunter[1], S.K. Peterson[2], S. Syngal[3], L.S. Acheson[4], G.L. Wiesner[5], J.A. Reiss[1], K.A.B. Goddard[1].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Department of Behavioral Science, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Dana-Farber Cancer Institute, Boston, MA; 4) Departments of Family Medicine and Community Health, Reproductive Biology, and Oncology, Case Western Reserve University, Cleveland, OH; 5) Vanderbilt Hereditary Cancer Program, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN.

Individuals with Lynch Syndrome (LS) are at increased risk of developing cancer, primarily colorectal cancer (CRC) and endometrial cancer (EC). Because of the risk, patients with LS are recommended to receive screening for CRC via colonoscopy starting at an earlier age and at more frequent intervals. Women with LS may have risk reducing surgery to prevent EC, and additional screening and interventions may be recommended for other associated cancer risks. We investigated the use of recommended interventions for LS in a real-world integrated health care system. We also conducted interviews with 12 patients and 10 primary care providers (PCPs) with at least one patient with LS to better understand how care is managed, and barriers and facilitators to receiving recommended care. We identified 73 patients with LS via chart review at Kaiser Permanente Northwest (KPNW), representing only 6% of the expected number of cases based on the population prevalence of the condition. We used chart review to extract information on family history, LS diagnosis, cancer history, and related healthcare utilization. Most patients (83%) received their diagnosis of LS at KPNW, and 67% had a diagnosis of ≥1 cancer prior to their diagnosis of LS. Patients accumulated an average of 3.2 years (range, 0 – 14) of electronic medical record history since their diagnosis of LS. Patients received colonoscopy during 87% of the observed follow-up intervals; 42% of at-risk women had a hysterectomy, and 38% had an oophorectomy following diagnosis of LS. While most had an initial visit with the Genetics department, 70% of patients with sufficient follow-up never contacted Genetics again. In interviews, patients had good recall of their recommended care, and 1/3 reported it was primarily their responsibility to manage their care, while 2/3 reported it was a joint responsibility with their PCP. PCPs reported limited awareness of LS guidelines, and most often provided care for £2 patients with LS. PCPs often reported that it was primarily another specialty's responsibility to manage LS care or that they rely on joint responsibility with the patient. Both patients and PCPs desire greater support in managing patient care for LS patients. In this setting, patients are often able to obtain recommended care to reduce their cancer risks. The Genetics department has an important role in supporting coordination of care, and needs to find ways to communicate their useful roles to PCPs and patients.

2998W

"A change in perspective": Exploring the experiences of adolescents with hereditary tumor predisposition. *E. Weber[1,2], C. Shuman[1,2,3], J.D. Wasserman[4], M. Barrera[5,6], A.F. Patenaude[7], K. Fung[8], D. Chitayat[1,2,9], D. Malkin[6], H. Druker[1,3,6].* 1) Department of Molecular Genetics, University of Toronto, Toronto, Quebec, Canada; 2) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Genetic Counselling, The Hospital for Sick Children, Toronto, ON Canada; 4) Division of Endocrinology, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Psychology, The Hospital for Sick Children, Toronto, ON, Canada; 6) Division of Hematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA; 8) Department of Social Work, The Hospital for Sick Children, Toronto, ON, Canada; 9) Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada.

   Introduction: Hereditary tumor predisposition syndromes are being more frequently recognized in the context of pediatric neoplasms. As more adolescents are identified with hereditary tumor predisposition, insight from their experiences can guide the provision of health care including genetic counseling. Previous research indicates that disclosure of tumor susceptibility is a significant event in adolescents' lives. However, no previous research has explored the experiences and perceptions of adolescents following the identification of a hereditary tumor predisposition syndrome. Purpose: We explored the lived experiences of adolescents with hereditary tumor predisposition syndromes and their perceptions of living at risk. Methods: Participants were sampled purposively from the Cancer Genetics Program at the Hospital for Sick Children, and included adolescents with pediatric-onset hereditary tumor predisposition syndromes, irrespective of personal or family history of tumors. Seven adolescents (ages 14-17) with five different syndromes were interviewed by telephone or in person using a semi-structured interview guide. We explored the disclosure process, support systems, and the benefits and harms of knowledge of tumor susceptibility. Interviews were transcribed verbatim and analyzed via interpretive description. Results: Four major themes emerged from the data: 1) The benefits of knowing about tumor susceptibility outweigh the harms; 2) Tumor susceptibility leads to a change in life perspective and choices made with respect to the future; 3) Perceptions are shaped by the disclosure process within one's health and family contexts; 4) Self-concept is not defined by tumor risk. Conclusions: Adolescents recognize the challenges associated with awareness of tumor predisposition, but also identify positive aspects in their life experiences. Results of this study can guide pre- and post-test genetic counseling of adolescents for hereditary tumor predisposition syndromes, facilitating the incorporation of this genetic information into an adolescent's self-concept in an adaptive manner.

2999T

The cost and cost-trajectory of whole-genome analysis to guide treatment of patients with advanced cancers. *D. Weymann[1], J. Laskin[2], R. Roscoe[3], M. Marra[3,4], K. Schrader[4,5], S. Chia[2,6], S. Yip[7], W. Cheung[2,6], K. Gelmon[2,6], A. Karsan[2,3,7], D. Renouf[2,6], D.A. Regier[1,8].* 1) Canadian Centre for Applied Research in Cancer Control (ARCC), Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, Canada; 2) Division of Medical Oncology, BC Cancer Agency, Vancouver, British Columbia, Canada; 3) Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 4) Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; 5) Department of Molecular Oncology, BC Cancer Agency, Vancouver, British Columbia, Canada; 6) Department of Medicine, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; 7) Department of Pathology & Laboratory Medicine, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; 8) School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

   PURPOSE: Precision medicine is gaining momentum in oncology. While the costs of whole genome sequencing (WGS) have decreased over time, the feasibility of the widely hyped $1,000 per genome is uncertain. Few data on the real-world costs of applying WGS and analysis (WGA) have been published. This study estimated the costs of applying WGA to guide treatments for patients with advanced cancers and characterized how the costs of applying these next generation genetic technologies are expected to evolve over time. **METHODS:** The setting for this study is the BC Personalized OncoGenomics (POG) program at the BC Cancer Agency Genome Sciences Centre in Vancouver, Canada. Individual-level cost data was obtained for 301 patients who underwent POG-initiated WGA from July 2012 to December 2015. We estimated mean sequencing costs across patients using bootstrapping methods and applied time series analysis to determine whether sequencing costs changed over time. We specified autoregressive integrated moving average models with explanatory variables (ARIMAX) for WGA costs and WGS and transcriptome sequencing costs. We produced ten-year forecasts to determine when costs are expected to reach thresholds, such as $1,000 per patient. **RESULTS:** The mean cost of POG-initiated WGA was $34,991 per patient (95% CI: $34,153, $35,828). WGS and transcriptome sequencing drove the majority of expenditures, with a mean cost per patient of $19,400 (95% CI: $18,404, $20,395). Targeted gene panels cost an average of $1,530 per patient (95% CI: $1,443, $1,617). Over time, mean WGA costs decreased, driven mainly by decreases in the cost of WGS and transcriptome sequencing. However, mean costs of bioinformatics and validation increased over time. Forecasting models showed that WGS and transcriptome sequencing costs could reach $1,000 per patient in the next 6 years. Reaching significant thresholds for WGA costs could take longer. **CONCLUSION:** The cost of WGA decreased over the studied time horizon, but mean per patient expenditures needed to realize WGA remain significant. Despite these elevated costs, WGA will identify a greater number of clinically actionable findings than lower cost alternatives, such as gene panels. Future research exploring the trade-offs between costs and benefits of WGA-guided cancer care will be crucial to guide health policy.

3000F

The Angelina Effect: An ongoing impact on cancer referrals in Northwest London. *S.K. Best, M. Tsakok, A. Melville, M. Kosicka-Slawinska, A. Brady, B. Bernhard.* North West Thames Regional Genetics Service, Northwick Park Hospital, London North West Healthcare NHS Trust.

   Background: The North West Thames Regional Genetics Service (NWTRGS) serves a population of 3.6 million people in North West London and adjacent counties. Following the publicity of Angelina Jolie's *BRCA1* diagnosis and risk-reducing mastectomy in May 2013, the NWTRGS experienced an unprecedented increase in cancer referrals. NICE guidelines recommend referral to specialist genetic services for patients with moderate or high-risk of a genetic basis for cancer. Managing the sudden increase in referrals was challenging and required initiative to continue to provide adequate care. **Methods: Initial Audit** A retrospective study was conducted on all genetic cancer referrals in January and June 2013; before and after the Angelina Jolie *BRCA1* announcement. Data collected: referral date, cancer type, appointment date, risk category, diagnostic tests. Microsoft Excel analysis was performed. Findings were compared with NICE guidelines for referral to specialist services (2013).   **Intervention** The following changes were implemented following initial audit: Only high risk patients to be offered clinic appointments – lower risk patients to receive standardised letters explaining risk and screening advice. Increased triage workforce employed.   **Reaudit** A reaudit was performed on January and June 2014 cancer referrals as per initial audit methods to evaluate change in practice.   **Results: Referral patterns: Initial audit:** cancer referrals more than doubled (129 to 271) between January and June 2013. 82% of total cancer genetic referrals were for breast or ovarian cancer. **Reaudit:** The number of referrals reduced but remained higher than pre-announcement levels. Referrals for breast and ovarian cancer increased compared to 2013, representing 93% of total genetic cancer referrals.   **Consultations: Initial audit:** 85.5% of patients (moderate and high cancer risk) receiving clinic appointments. 40% of patients seen subsequently had genetic testing. **Reaudit:** Only high risk patients (53.5%) were offered clinic appointments. The remainder of referred patients received standardised letters. 89.5% of patients seen subsequently had genetic testing.   **Conclusions:** There has been a sustained surge in genetic cancer referrals following the Angelina Jolie *BRCA1* announcement in 2013. Increasing demands on genetic services has driven implementation of new triage systems and re-allocation of resources. Ongoing adjustment will be needed to manage high numbers of referrals associated with increasing public awareness.

3001W

Screening new diagnoses of endometrial cancer for Lynch syndrome to reduce morbidity and mortality: A systematic review. *H.C. Groom[1], M.A. Piper[1,2], J.L. Kuntz[1], C.S. Bennette[3], D.L. Veenstra[3], K.A.B. Goddard[1].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Hayes, Inc., Lansdale, PA; 3) Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA.

   Statement of purpose: We conducted a systematic review to assess the evidence for whether screening for Lynch syndrome (LS) among women newly diagnosed with endometrial cancer (EC) improves cancer morbidity and mortality outcomes for the patient or her family members, as the over-arching research question. **Methods**: An indirect evidence chain composed of 4 major questions was constructed to address the over-arching question. We conducted a systematic literature search from 2008 to 2016 for the over-arching question and to 2015 for all other questions. Two investigators independently reviewed 1086 abstracts and quality-rated 140 full-text articles against pre-specified inclusion criteria. Data was abstracted from included studies and randomly checked for accuracy. We also developed an economic evaluation. **Results**: No direct evidence addressed the overarching research question was identified. Studies in the indirect evidence chain indicate microsatellite instability and immunohistochemistry screening tests have diagnostic accuracy for LS in EC similar to LS in colorectal cancer (CRC). Sensitivities of 100 percent were reported for next generation sequencing to detect LS variants. Results of the economic model indicate that screening and followup Sanger sequencing of LS-related genes in screen-positive samples, compared to genome/exome sequencing, remains the cost-effective method for detecting LS even when considering the benefits of cascade testing among relatives. The distribution of EC versus CRC as the first cancer in LS patients who were diagnosed with both did not favor CRC, suggesting that EC is an important sentinel cancer in LS families. EC also appears to be the defining cancer in a small number of LS families. **Conclusions**: Screening new EC diagnoses could identify LS families in addition to those found by CRC screening alone. Indirect evidence suggests that screening new EC diagnoses can reduce CRC incidence and mortality, and EC (among female relatives) but not ovarian cancer incidence and burden. Improvements in uptake of genetic counseling are needed for screening to achieve full benefits; uptake of genetic testing and surveillance for cancer in LS-confirmed individuals could also be improved. Rare, serious harms are associated with CRC surveillance and with certain risk-reducing surgical options. This study was supported by funding from Leidos Biomedical Research, Inc. Prime Contract HHSN261200800001E issued by The National Cancer Institute.

3002T

Implementing a clinical partnership to provide telegenetics services at community sites without access to genetic counselors. *N.K. Reddy[1], E.M. Wood[1], D. Harris[1], S. Jideama[1], P. Higgins[1], R. Sawhney[2], P. Singh[2], K. Vidrine[2], H. Pinkston[2], J. Shevock[2], J. Jaeger[1], A. Bradbury[1]. 1) University of Pennsylvania, Philadelphia, PA 19104; 2) Bayhealth Medical Center, Dover, DE 19901.*

  Background: Many patients in community practices do not have access to cancer genetic services, creating potential disparities in care. Remote two-way video-conferencing (RVC) is a feasible alternative but scalability of this model remains unclear. **Methods:** The Penn Telegenetics Program provides cancer genetic counseling services using RVC to community hospitals without access to genetic providers. This program originated from an NIH-funded pilot-study investigating the feasibility of using RVC cancer genetic services at 3 community practices. Bayhealth Medical Center, one of the participating community sites, established a clinical contract to maintain continuity of telegenetic services. We present the data from the first 18 months of our clinical partnership. **Results:** 191 patients have been referred to the program by on-site physicians (n=105, 55%) and local community doctors (n=86, 45%). The majority of referrals are from oncologists (52%) and surgeons (32%). Genetic counselors obtained licensure in the site state, collaborating with the site physician who is the ordering provider. To date, 84 patients have been seen for pre-test (V1) RVC counseling; of the remaining 107 referrals not seen for V1, reasons include patient refusal of services, lost to follow-up or pending decision to receive services. 60 (71%) patients elected to have genetic testing and completed disclosure sessions (V2) by RVC (n=56) or phone (n=4). 5 (8%) had a positive result, 10 (17%) had a variant of unknown significance and 45 (75%) had negative results. 131 RVC sessions (including V1 and V2) were conducted. 109 (83%) sessions were completed successfully without disconnections, 19 (15%) were completed with difficulty (e. g. screen freezing, prolonged reduced audio/video quality) and 3 (2%) had technology failures prompting completion by phone. An online family history tool was added to increase efficiency, but some patients (25%) did not use the tool due to discomfort with computers or no internet access. The clinical contract covers the cost of the genetic counseling but testing is covered by patient insurance, consistent with usual care. **Conclusion:** Remote two-way video-conferencing services can be implemented through clinical contracts to provide access to genetic services at community sites without genetic counselors, providing a novel delivery model to enhance the dissemination of and reduce disparities in access to cancer genetic services.

3003F

Patient safety in genomic medicine: Learning from the experiences of cancer genetics professionals. *S.M. Fullerton[1], D.M. Korngiebel[2], J. Zech[3], A.M. Chapelle[3], D. Davies[1], J.D. Carline[2], W. Burke[1], T.H. Gallagher[3]. 1) Dept Bioethics & Humanities, University of Washington, Seattle, WA; 2) Dept Biomedical Informatics & Medical Education, University of Washington, Seattle, WA; 3) Dept Medicine, Div General Internal Medicine, University of Washington, Seattle, WA.*

  Introduction: Genetic and genomic testing poses a variety of concerns for patient safety, related to clinician training, practice variation, and rapidly evolving technologies and test standards. As genomic tests are increasingly integrated into medical care, anticipating and addressing such potential concerns will be crucial. Here, we report a qualitative investigation of the views of cancer genetics professionals aimed at identifying challenges to the use of cancer predisposition-related genetic testing in oncology, an area in which genomic medicine is already established and growing in significance. **Methods**: First, we conducted semi-structured interviews with 22 key informants (medical geneticists, genetic counselors, and medical oncologists) affiliated with diverse cancer practice settings in the US. Interviews were designed to explore informants' experiences with the use of genetic testing and to identify examples of testing-related care breakdowns. Next, we used the interview findings to develop case vignettes, which formed the basis for focus group discussions with 18 additional cancer genetics providers drawn from three different healthcare institutions in the Pacific Northwest. **Results**: Key informants offered numerous examples of challenges they and their patients had encountered in the course of pursuing cancer-related genetic testing. These rich first-person accounts informed the development of 8 vignettes, encompassing errors in genetics referral, test ordering, interpretation, and follow-up, as well as team-based communication about test results. Two of the vignettes, one focused on surgical referral for a patient with a Variant of Uncertain Significance (VUS), and one focused on clinical trial-related *BRCA1/2* testing in the absence of pre-test counseling, were brought to the focus groups for detailed consideration and discussion. Together, these results informed the development of a national survey, currently in the field, which aims to quantify the degree to which oncologists in academic and community settings grapple with the concerns identified by our preliminary sample of cancer genetics professionals. **Conclusions**: Our study has provided a description of recent practice challenges associated with cancer-related genetic and genomic testing. Additional research will be needed to identify and test suitable interventions aimed at promoting the safe incorporation of genetic and genomic tests into existing systems of healthcare.

3004W

Personal utility and economic benefit of returning incidental findings in patients with colorectal cancer/polyposis syndromes. *D. Regier[1,2], G. Jarvik[3], D. Veenstra[3].* 1) BC Cancer Agency, Vancouver, British Columbia, Canada; 2) School of Population and Public Health, University of British Columbia, Canada; 3) University of Washington, Seattle, WA.

   Background: A challenge with the application of next generation DNA sequencing technology is the possibility of uncovering incidental genomic findings. A paucity of evidence on patients' preferences and economic value for money has hindered health policy outlining the return of incidental findings. The study objective was to estimate the personal utility and net economic benefit for returning complex information derived from genomic incidental findings. **Methods:** A discrete-choice experiment evaluated preference-based personal utility and willingness to pay for incidental findings information. Study subjects were drawn from randomized controlled trial of exome sequencing versus usual care in patients with colorectal cancer/polyposis syndromes conducted in Seattle, Washington. Preference data were analyzed with a random parameters hurdle model. Decision analytic modeling was used to estimate the net-benefit of returning incidental findings information. **Results:** 130 respondents completed the preference questionnaire. Respondents valued receiving information on high-penetrance disorders, but expressed disutility for receiving information on low-penetrance disorders. Respondents indicated an average willingness to pay of $483 (95% CI, 359,695) to receive incidental findings information in a scenario where clinicians return high-penetrance, medically actionable disorders, but 6% of respondents were predicted to opt-out of receiving incidental findings information under these conditions. In terms of cost-effectiveness, incorporating WTP for genomic knowledge resulted in a net-benefit of $281 per patient. **Conclusions:** Most patients valued receiving incidental findings, but preferences depend on the type of finding and not all individuals want to receive incidental results. Our results indicate that returning incidental findings information may provide net economic value, but individual-level patient preference should be elicited in order to maximize societal benefit.

3005T

Personalizing cancer care: A policy perspective. *J. Zhang[1,2], G. Bertier[3,4], Y. Joly[3,4].* 1) Cancer Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 2) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; 3) Centre of Genomics and Policy, Montreal, Quebec,Canada; 4) Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

   Context: Precision medicine can be defined as a predictive, preventive, personalized and participative healthcare service delivery. Although it seemed like a futuristic ideal only a decade ago, recent developments in molecular biology and information technology make it possible for precision medicine to take place today, through the use of massive amounts of genetic, omics, clinical, environmental and lifestyle data. Cancer being one of the most prominent public health threat in developed countries, both the research community and governments have been investing significant time, money and efforts to make cancer precision medicine a reality. Cancer precision medicine (CPM) research is extremely promising. Indeed, genomic research results are providing us with a more thorough understanding of cancer. It is a complex, multifaceted disease, which challenges our established way to classify, counsel and treat patients. A number of these results are already actionable, and highly publicized examples demonstrate significant improvements of prevention, survival and patients quality of life in certain cancers. But a number of hurdles are still on the road to an optimal integration of standardized and evidence based use of CPM in healthcare systems. **Objective:** In this perspective, we describe key issues that public health policymakers will need to consider in the near future for CPM's promises to be fully realized. **Conclusions:** Policy makers have a duty to optimally integrate CPM research results to the clinic. Genomic medicine is one of the most promising avenues for cancer prevention, survival and treatment. Focusing on translating cancer genomic research results does not mean to disregard all other (non-genetic) aspects of cancer. Many areas still need research efforts, but these have to be made simultaneously with clinical applications.

3006F

A molecular autopsy program for sudden unexpected death: San Diego County study. *M. Rueda[1], E.G. Spencer[1], E.D. Muse[1], J.R. Lucas[2], E.J. Topol[1], A. Torkamani[1].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Medical Examiner Department, San Diego County, California, 92123, USA.

Many thousands of <45 year old individuals die suddenly and unexpectedly (SXD) in the USA. Depending on the underlying cause, SXD is divided into sudden cardiac death (SCD) including sudden infant death, or sudden death caused by non-cardiac failure such as pulmonary embolism, ruptured aortic aneurysm or infectious disease. Post-mortem examination of these cases yields negative results for ~10% of them, leaving relatives (often asymptomatic yet still at risk) with an inaccurate or ambiguous family health history. Genetic diagnosis is getting momentum in helping solving SXD mysteries. Analysis of literature shows that post-mortem genetic testing (a.k.a. molecular autopsy), even limited to gene panels of cardiac channelopathies, produced positive diagnosis in ~30% of the cases analyzed. According to our studies in the San Diego county, exome analysis of 25 cases referred by the Medical Examiner provided a likely cause of death (where a previously reported or expected pathogenic mutation was observed in a SCD-related gene) in 17% of them, a plausible cause in 25% of them (where a rare mutation of unknown significance was observed in a SCD gene), and a speculative cause in 25% of the cases (where a known pathogenic mutation was identified for a condition that potentially led to SXD). Despite the abundance of functional studies, in silico predictors of pathogenicity, and a plethora of databases of allele frequencies for variants, many of resulting variants remain of unknown significance (VUS). On top of that, most putative pathogenic mutations were inherited from living relatives and at population frequencies incompatible with full disease penetrance. These observations suggest that SXD risk may be influenced by incomplete penetrance, variable expressivity, and gene-environment interactions. We believe that a comprehensive molecular autopsy program would provide more accurate family health information to a wider spectrum of families. Consequently, we initiated a systematic and prospective, family-based, molecular autopsy study, jointly organized by a medical examiner office and an academic center. Our objective is to create a worldwide catalog of genetic variants associated to SXD to allow for large-scale genetic studies and empower the unrealized potential of prevention for SXD conditions.

3007W

Reaction to risk, decisional regret, and perception of locus control in central hubs of information exchange following disclosure of disclosing genetic risk for coronary heart disease: The MI-GENES study. *S. Brown, I. Kullo.* Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN.

In the myocardial infarction genes study, participants (n=203) aged 45-65 years at intermediate CHD risk based on conventional risk factors and not on statins were randomized to receive their conventional risk score (CRS) or also a genetic risk score (GRS) based on 28 variants, in meetings with a genetic counselor and then a physician. Individuals exhibiting high levels of information exchange via the internet and interpersonal communication were identified as central hubs. Whether disclosing genetic risk for coronary heart disease (CHD) to individuals identified as central hubs of information exchange in social networks influences their reaction to risk, decisional regret, and perception of locus control is unknown. Surveys to assess information exchange, as well as reaction to risk, decisional regret, and locus control were completed prior to, at, or following risk disclosure. We assessed whether reaction to risk, decisional regret, and locus control differed by GRS disclosure. Data were reported as mean with standard error or odds ratio with confidence interval; significance was determined by regression analysis. Upon risk disclosure, central nodes in the GRS group were more likely than GRS peripheral nodes to feel hopeful (OR 4.36, CI 1.46-15.52, p=0.01) and determined (OR 0.17, CI 0.03-0.67, p=0.01). Three months later, central nodes in the GRS group were less likely than GRS peripheral nodes to believe their CHD risk was primarily determined by chance (11.22±0.83 vs. 13.63±0.59, p=0.02). Central nodes in the GRS group (15.21±0.46 vs. 14.05±0.33, p=0.03) and the CRS group (15.3±0.59 vs. 13.74±0.31, p=0.02) were more likely than GRS and CRS peripheral nodes, respectively, to believe that their CHD risk was influenced by taking actions in response to doctors' recommendations. Among GRS participants, central nodes were less likely than peripheral nodes to regret their decision to undergo CHD genetic risk testing at six months after risk disclosure (5%±2 versus 14%±2, p=0.0016). There was no difference in reaction to risk or decisional regret between central and peripheral nodes in the CRS group. Individuals identified as central hubs reported higher locus of control, positive reactions to risk, and less decisional regret than their peers in the GRS group, and higher locus of control in the CRS group. These parameters may serve as psychosocial markers of successful health promotion in central hubs of social networks following disclosure of a GRS for CHD.

3008T

Motivation, perception, and treatment beliefs in the Myocardial Infarction Genes (MI-GENES) randomized clinical trial. *S. Kattel[1], T. Onyekwelu[2], S.A. Brown[1], H. Jouni[1], E. Austin[1], I.J. Kullo[1].* 1) Mayo Clinic, Rochester, MN; 2) Medical University of South Carolina, Charleston, SC.

   Background: Current risk estimation equations for coronary heart diseases (CHD) are based on conventional risk factors. Incorporating newly identified genetic markers for CHD may improve risk estimation and individualization of therapy. Little is known about individuals' motivation, perception, or treatment beliefs related to the use of genetic information in risk estimates for CHD. Methods: Participants age 40-65 years, not on statin therapy and at intermediate 10-year risk for CHD were randomized to receive their estimated 10-year risk for CHD as a conventional risk score (CRS) based on conventional risk factors alone, or a CRS and a genetic risk score (GRS) based on 28 CHD genetic risk variants. CHD risk was disclosed during a 30-min scripted session by a genetic counselor, followed by a physician visit for shared decision-making to determine the need for statin therapy initiation. After 6 months, participants who initially received only their CRS also received their GRS. Surveys on motivation and perception were completed at 3 months and treatment beliefs at 6 months following risk disclosure. Survey data were analyzed using Wilcoxon Signed-Rank, Chi-square, or Fischer's exact tests, or linear or logistic regression, as appropriate. Results: Overall, motivation to participate in the study was favorable and did not differ between the CRS and GRS groups (16.95 ± 0.82 vs. 17.58 ± 0.83, p=0.091), but participants who initially received their GRS indicated a greater desire to find ways to improve health as a reason for participation (OR: 0.53 (95%CI: 0.29,0.94), p=0.028). Correct perception about genetic testing did not vary between the two groups (15.29 ± 0.39 vs. 15.12 ± 0.40, p=0.835). However, participants who initially received their GRS were more inclined to recommend genetic testing to family and friends (9.95 ± 1.88 vs. 10.52 ± 2.17, p=0.023). Conclusions: In this study, motivation to participate and perception towards genetic testing were overall favorable. Genetic risk disclosure was associated with increased motivation to recommend genetic testing to family and friends. Perceptions about genetic testing were not affected by genetic risk disclosure.

3009F

A novel approach to screening for familial hypercholesterolemia. *H. Zierhut, M. Campbell, J. Humanski.* Genetics, Cell Biology, & Development, University of Minnesota Twin Cities, Minneapolis, MN.

   Heterozygous familial hypercholesterolemia (FH) is a common genetic condition affecting ~1 in 300 to 1 in 500 individuals worldwide. Yet FH is highly underdiagnosed with as few as 1% of individuals identified with the condition in the United States. The primary aim of this study was to test the feasibility of a public health screening program to identify individuals at high risk of FH through a novel screening approach at the Minnesota State Fair. A finger-prick, non-fasting lipid panel was obtained and a survey which consisted of 44 open and close ended questions divided into four sections: medical and family history of FH, opinions of cascade genetic testing, patient activation, and demographics was completed. A total of 971 participants met criteria and completed a cholesterol screen. In total, five individuals met either the Simon Broome Register or Dutch Lipid Clinic Network criteria for possible FH. Overall participants held positive opinions of genetic testing for FH and most participants responded that they would like to know if they had a genetic form of high cholesterol. The vast majority listed they had no barriers to communication of genetic testing information. However, the most common barrier listed was lack of communication skills. Participants' preferred mode of communication with the highest comfort with communicating in-person (n=591, 64%), followed by telephone (n=468, 51%), email (n=460, 50%), letter (n=230, 25%), link to website (n=220, 24%), or conversation through a physician or genetic counselor (n=209, 23%). Our results suggest that a public health screening program for FH is viable at the State Fair. Cascade screening for FH is acceptable by the public, but barriers to communication and faciliation of genetic testing will likely exist. We argue that further research is needed to expand this study to test the outcomes of a fully operational screening program for FH at the State Fair.

3010W

Cascade family screening following molecular autopsy for sudden cardiac death using targetted gene panels. *J.C.S. Dean[1], F. Cann[1], M. Corbett[2], D. O'Sullivan[1], S. Tennant[1], H. Hailey[1], P. Broadhurst[3], R. Rankin[4], J.H.K. Grieve[2].* 1) Medical Genetics, Aberdeen Royal Infirmary, Aberdeen, United Kingdom; 2) Forensic Medicine Unit, University of Aberdeen; 3) Department of Cardiology, Aberdeen Royal Infirmary; 4) Department of Pathology, Raigmore Hospital, Inverness.

Inherited cardiac conditions are thought to underlie between 27 and 44% of sudden arrhythmic death syndrome (SADS - sudden unexpected death with negative pathology and toxicology). This has led to a recommendation that a molecular autopsy should become part of the standard of care in such cases. Interpretation of genetic variants may be difficult where the phenotype is limited, although variant frequency data and family segregation studies may be helpful. 96 cases of sudden cardiac death were assessed through a multidisciplinary approach (46 with normal hearts, 50 with pathological evidence of cardiomyopathy) and DNA samples were screened using small gene panels comprising 6 channelopathy genes (KCNQ1, KCNH2, SCN5A, KCNE1, KCNE2 and RYR2), 4 Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) genes (PKP2, DSC2, DSP and DSG), 6 sarcomere genes (MYBPC3, MYH7, TPM1, MYL2, TNNI3, TNNT2) with or without LMNA, SCN5A and TTN depending on the phenotype (SADS, ARVC, dilated and hypertrophic cardiomyopathy respectively). 15% of SADS cases and 18% of cardiomyopathy cases had pathogenic variants in the genes tested. 3 families had variants previously reported as pathogenic which had no clinical phenotype in relatives, and were relatively common in the ExAC population – these cases were counted as negative. Amongst those with a positive molecular autopsy, 94% of families sought cascade screening, while only 41% of those with a negative molecular autopsy sought cardiological screening. In families who had a full evaluation, an inherited cardiac condition was identified in 33% of SADS families, and 56% of cardiomyopathy families. Molecular autopsy is useful in reaching a diagnosis after a sudden unexpected death, but variants must be interpreted in the context of family clinical investigations and population variant frequency data. A positive molecular autopsy encourages family cascade screening and may help prevent further morbidity and mortality.

3011T

Barriers and facilitators to optimal visit frequency in management of hypertension among African American and Latino patients in a safety-net clinic. *J. Lee[1,2], V. Fontil[2].* 1) San Francisco State University, 1600 Holloway ave. San Francisco, CA 94132; 2) University of California, 1001 Potrero Ave, San Francisco, CA 94110.

Despite improvements in management of hypertension (HTN) at the national level, rates of uncontrolled hypertension remain disproportionally high among African American and Latino patients. Visit frequency is a key process in HTN management that leads to improved blood pressure (BP) control. However, African Americans and Latino patients have the highest no-show rates in clinic, which may contribute to the racial disparity in BP control. We seek to examine cognitive and socioeconomic barriers to more frequent clinic visits in African American and Latino patients that contribute to higher no-show rates. We will conduct focus groups of African American, Latino, and White patients with uncontrolled HTN at the Zuckerberg San Francisco General Hospital's Richard Fine People's Clinic (RFPC) – a large safety-net clinic associated with an academic institution. Qualitative analysis will examine racial differences in patient barriers and facilitators to improving visit frequency for management of HTN at RFPC. We anticipate that patient identified challenges and preferences will vary between African Americans, Latinos and White patients at RFPC. A clearer understanding of patient identified facilitators and barriers will inform interventions and best practices to optimize visit frequency and ultimately improve hypertension control in safety-net clinics and in high-risk populations that are vulnerable to health disparities.

3012F

Cardiologists' responses to whole genome sequencing: Preliminary findings from the MedSeq Project. *K. Christensen[1], J.L. Vassy[1,2], A.L. Cirino[1], M.F. Murray[3], A.L. McGuire[4], R.C. Green[1,5,6] for the MedSeq Project.* 1) Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 2) VA Boston Healthcare System, Boston, MA; 3) Geisinger Health System, Danville, PA; 4) Baylor College of Medicine, Houston, TX; 5) Partners Healthcare Personalized Medicine, Cambridge, MA; 6) Broad Institute of MIT and Harvard, Cambridge, MA.

While whole genome sequencing (WGS) has great potential to improve patient care by identifying predispositions for disease and informing treatment decisions, policymakers and payers are apprehensive about the potential of WGS findings to initiate a cascade of follow-up testing and screening that could yield limited clinical benefits. Concerns are particularly strong in specialty settings where WGS is already being used for diagnostic purposes and physicians may receive secondary findings that are unrelated to their areas of expertise. As part of the MedSeq Project, 100 patients (mean age 56, 43% female, 88% non-Hispanic white) with diagnoses of dilated or hypertrophic cardiomyopathy were randomized to meet with one of seven cardiologists (who were also study participants) to review their family history of disease alone (FH arm) or in conjunction with a WGS report (WGS arm). In questionnaires completed immediately following disclosure sessions, cardiologists indicated that they ordered or recommended follow-up for one participant randomized to the FH arm and four participants randomized to the WGS arm (2% vs 8%, p=0.20). Follow-up in the FH arm included a referral to cancer genetics in response to a family history of pancreatic cancer. Follow-up in the WGS arm included referrals to cancer genetics and endocrinology in response to findings about monogenic disease, and two screening procedures in response to risk information about cardiometabolic traits. Cardiologists rated the current utility of WGS reports no higher than they rated the current utility of the family history analysis (4.0 vs 3.9 on 10-point scales, p=0.62), on average, but tended to believe that the WGS reports would have greater utility in the future (5.1 vs 4.0 on 10-point scales, p=0.016). Furthermore cardiologists stated that the WGS reports improved overall clinical care beyond what family history alone could have achieved for 33% of patients. Results suggest that cardiologists perceived some utility in WGS, even though findings were unlikely to motivate them to order follow-up tests. In conjunction with previous MedSeq Project data, these findings also suggest that cardiologists were less likely than primary care physicians to act on WGS findings.

3013W

MyGeneRank app: Empowering patients with their genetic risk. *N. Wineinger, B. Schrader, B. Molparia, A. Lippman, E. Spencer, M. Rueda, E. Muse, E. Topol, A. Torkamani.* Scripps Translational Science Institute, La Jolla, CA.

Can the results from genome-wide association studies (GWAS) of common disease be adopted to inform clinical practice and health decision-making? Researchers have begun to address this question by summarizing numerous signals across the genome to assess overall genetic risk, often as a genetic risk score (GRS). In many cases, they have shown that these scores can meaningfully differentiate individuals with and without disease. Coronary artery disease (CAD) has become a popular focus of GRS estimation because: 1) it remains the most common deadly disease; 2) has heritability between 40-60% while massive GWAS meta-analyses results made freely available to other researchers; 3) the Framingham risk score, the most popular clinical assessment of CAD risk, performs exceptionally bad at predicting CAD; and 4) a number of behavioral and pharmaceutical preventative interventions exist. CAD researchers have previously shown that a GRS constructed using >49,000 markers provides better differentiation between cases and controls than one using only genome-wide significant markers. Meanwhile, others have shown that individuals at the highest quintile of genetic risk get the most benefit from lipid-lowering medications (i.e., statins) whereas individuals at the lowest quintile get limited to no benefit. Given the clear clinical benefit, yet practical computation challenge in assessing CAD genetic risk, how can patients get this information? And if so, what health decisions will they make from it? We developed the MyGeneRank app available through Apple's ReserachKit. MyGeneRank is designed to provide individuals with their relative genetic risk based on estimates obtained from published GRS literature. The app currently supports a GRS for CAD based on ~49,000 markers, but more risk scores for other diseases are planned. To participate, individuals provide us with access to their genetic data, currently only available through the 23andMe API. Computations, including imputation, GRS calculation, and comparison to reference populations, are performed on our servers and are returned to the user via the app. This entire process takes only a few minutes. Individuals are also asked to complete a short survey, and six months later a follow-up survey. Our overall goal is to provide individuals access to information on their genetic risk for, as of now, CAD and observe how such access influences health decisions.

3014T

Creation and implementation of a BC Registry for familial hypercholesterolemia. *A. Shokoohi[1], L. Cermakova[2], I. Priecelova[2], J. Frohlich[1,2].* 1) University of British Columbia, Vancouver, BC; 2) Healthy Heart Program Prevention Clinic, St. Paul's Hospital, Vancouver, BC.

Cardiovascular disease (CVD) is now the leading cause of death worldwide. Atherosclerosis is the most common pathological change underlying CVD, with high cholesterol levels proven as a major risk factor for its development. Familial hypercholesterolemia (FH), a common autosomal dominant genetic disease, causes high cholesterol levels in the bloodstream, leading to premature CVD. With a prevalence of 1:500 in the population, approximately 20 million people worldwide and around 8000 in British Columbia are carriers of one of the genes for FH. Although more common than cystic fibrosis, vast majority of FH cases still remain undiagnosed, which implies that health professionals lack awareness of FH, its diagnostic features, and consequences. The purpose of this project is to create a registry with the diagnosis of FH in British Columbia to simplify education and treatment, ultimately reducing morbidity and mortality from CVD through early diagnosis and effective disease management. Patients with FH were identified based on their family history, physical examination, and laboratory results and stratified as definite, probable, or possible FH according to their final score on The Dutch Lipid Clinic Network FH criteria. Cascade screening was used to track the mode of inheritance of the disease to identify affected family members. Thus far, we have identified over 1000 individuals with FH based on chart review, with 693 consented to join the registry. Further research will aim to identify and educate FH individuals, their families, and their physicians regarding the appropriate management of hypercholesterolemia using lifestyle measures and medications.

3015F

Genetic counseling referrals for autism spectrum disorder: Review of a prenatal genetic counseling program. *D. Cutillo, L. Heckman, S. Floyd, A. Donnenfeld, D. Ramsey.* Genetic Counseling, Integrated Genetics, Philadelphia, PA.

Autism Spectrum Disorder (ASD) affects 1 in 68 children in the United States (U.S. Centers for Disease Control and Prevention 2012). The purpose of this retrospective study was to track genetic counseling referrals for family history of ASD at Integrated Genetics over a three year time period from 2013-2015. A total of 157,093 patients were seen for comprehensive prenatal/preconception genetic counseling between January 1, 2013 and December 31, 2015. 32,707 patients (20.8%) were referred for a family history of birth defects, intellectual disability and/or genetic disorders (included referrals for ASD). Genetic counseling patients referred with the specific indication of ASD by year were: 2013: 439 patients (3.8%), 2014: 502 patients (4.5%), and in 2015: 518 patients (5.2%). Of note, an additional 6993 patients not referred for ASD history, reported a first and/or second degree relative affected with ASD during the genetic counseling session. Risk assessment confirmed all patients referred for ASD (n=1459) were at increased risk for ASD in a current or future pregnancy. Fragile X carrier testing and prenatal microarray were among genetic tests offered to patients in this referral group. The acceptance rate for fragile X carrier testing was 24.4% (1/355 patients tested positive for a premutation). 11.8% of patients had testing previously (no abnormal results reported) and 63.8% of patients declined testing. Prenatal diagnosis with microarray was accepted by 21/1459 (1.4%), all with normal results. A review of this data indicates an increasing referral pattern of ASD among patients presenting for prenatal/preconception genetic counseling. Patients presenting for genetic counseling related to ASD are seeking information on etiologies, recurrence risk and prenatal testing options. Comprehensive genetic evaluation of affected individuals utilizing genetic testing, including microarray and whole exome sequencing is expected to identify a cause in up to 25% of individuals with ASD and is an important option to discuss during prenatal/preconception genetic counseling. As seen in this study, testing at risk family members with Fragile X carrier testing and offering prenatal microarray fails to clarify potential risks in most families with a history of ASD. Ongoing education specific to ASD is a critical component of a prenatal/preconception genetic counseling program to meet the needs of this patient population. .

3016W

Attitudes towards facial transplants among individuals with neurofibromatosis. *R.L. Glaser[1], D. Singman[1], J. McNairn[1], V. Iannone[2].* 1) Dept of Biology, Stevenson University, Stevenson, MD; 2) Dept of Psychology, Stevenson University, Stevenson, MD.

Neurofibromatosis (NF), an autosomal dominant disorder, has a phenotype ranging in severity from café-au-lait spots to multiple large tumors (neurofibromas). Surgery can remove neurofibromas, but they usually regrow. Full facial transplants, which have been used to treat NF since 2013, prevent neurofibromas from reoccurring by replacing the patient's facial tissue with donor tissue. This is the first study to assess the response of individuals with NF to facial transplants for cosmetic and functional purposes. A 26 item online survey, advertised on NFMidAtlantic.org between June-September 2015, assessed the reactions of individuals with NF to facial transplants for cosmetic and functional purposes. In addition, data on self-esteem, social behaviors, visibility and severity of NF, as well as demographic data such as type of NF, gender, and age were also collected. SPSS was used to analyze the survey data using independent sample two-tailed t-tests and Pearson's correlation coefficient. Overall, facial transplants for both cosmetic and functional purposes were not supported by our study population. The type of NF, visibility, and severity of the phenotype did not significantly affect the response to facial transplants. Even though our study population did not support facial transplants in general, females were significantly more likely than males to consider a cosmetic facial transplant for themselves (p=0.042) and for others (p=0.005). Lower self-esteem correlated with a greater consideration for cosmetic facial transplants for themselves (-0.341, p=0.01) and for others (-0.221, p=0.05). However, self-conscious individuals were significantly more likely to consider cosmetic facial transplants only for others (p=0.044), not themselves (p=0.429). Interestingly, the average self-esteem score of our study population was significantly lower than that of a control population (27.5 vs. 31.0, p<0.0001) and that of people with other facial deformities (27.50 vs. 29.61, p=0.025). Our results are reminiscent of the attitudes regarding cochlear implants in deaf study populations and attitudes regarding prenatal testing for achondroplasia in achondroplasia study populations.

3017T

Genetic counseling for male patients with myotonic dystrophy type 1 deciding whether to have a child. *M. Matsuo, M. Urano, K. Saito.* Institute of Medical Genetics, Tokyo Women's Medical University, Tokyo, Japan.

Myotonic dystrophy ;type 1 (DM1; OMIM#160900), with myotonia and progressive muscular weakness, is the most common adult muscular dystrophy. DM1 is due to a CTG repeat expansion in the 3' UTR of the *DMPK* gene. The mean intergenerational CTG repeat expansion tends to be minimal when transmitted by the father but high when transmitted by the mother. Invasive prenatal genetic testing (GT) and preimplantation genetic diagnosis (PGD) are permitted only when the fetus may have a severe disorder according to "Opinions Concerning Prenatal Genetic Testing" and "Opinions Concerning Pre-implantation Diagnosis" from the Japan Society of Obstetrics and Gynecology. Considering the parental gender difference in severity, prenatal GT and PGD are under discussion for paternal transmission of DM1 in Japan. We herein report our genetic counseling experience with two male DM1 patients. Case 1, an asymptomatic Japanese male in his thirties whose father had DM1, requested pre-symptomatic GT before marriage to decide about conceiving a child. GT revealed an expanded allele with about 100 CTG repeats. He contemplated having prenatal GT or PGD abroad and whether or not to marry just after his GT. He married and fostered a child 8 years later. Case 2, an asymptomatic Japanese male in his forties, had three siblings with DM1. He requested pre-symptomatic GT to decide whether to conceive a child. The GT revealed that he had an expanded allele of about 100 CTG repeats. To date, he has had no children. Discussion It is difficult to determine fetal phenotypic severity by prenatal GT or PGD in pregnant women with DM1. Thus, testing is of limited value for allelic confirmation. Mildly affected fetuses are comparable to those with paternally transmitted disease and severely affected fetuses are subjects for prenatal GT. Severely affected congenital DM1 fetuses are comparable to those with maternally transmitted disease and are not suitable for prenatal GT. Allowances for prenatal GT and PGD need to be reassessed in relation to fetal severity by ethics committees of the Japanese Society of Obstetrics and Gynecology and / or genetics-related societies Our society regards parents giving birth to and raising a child as fully qualified adults. "Blood relation" and "pedigree" are traditional Japanese values, with relatives and the public expecting a couple reproduce. Thus, being "childless" and "fostering a child" are not alternatives in Japan. We need a mature society with diverse family values.

3018F

Long-term impact of Huntington's presymptomatic genetic testing: Interviews with at-risk individuals 20-30 years after testing. *R.L. Dvoskin[1], J.M. Bollinger[1], A. McCague[1,2], K.M. Stuttgen[2], B. Shpritz[3], J. Brandt[3,4], D.J.H. Mathews[1].* 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD.

   To learn about the long-term impacts of genetic testing, we are conducting a follow-up study of participants from the presymptomatic testing program (PTP) of Johns Hopkins' Baltimore Huntington's Disease Project, one of the two initial programs in the U.S. We are conducting semi-structured interviews with at-risk probands who enrolled in the PTP between 1986 and 1996. Target enrollment is 20 people with an expanded HD repeat, 20 with normal repeats, and 10 who enrolled in PTP but dropped out before testing or disclosure. We anticipate completing enrollment in Summer 2016. We ask participants how their testing experience and knowledge of the result has influenced their life over the past 20-30 years, including decisions about career, children, and relationships. Nearly all participants reported positive feelings about their decision to enter the PTP and about their testing experience. Participants found the pre-test counseling extremely helpful. Some found the number of counseling sessions excessive; nevertheless, most mentioned the dangers of receiving one's HD risk status without sufficient preparation. To date, all (N=27) said if they were to do it over they would still be tested. Many believed knowing their risk status had a major impact on their life. For those with expanded repeats, positive long-term impacts included the relief of knowing and the ability to make informed decisions and plans regarding: long-term care and health insurance; reproduction; building social or family support; and how to spend their presymptomatic years. Negative impacts included fear of employment discrimination and guilt or concern for children who are now themselves at risk. For people with a normal repeat, positive effects included relief and the ability to live without uncertainty or symptomatizing; empowerment to leave an unhappy union or to pursue a new relationship; the ability to have children without fear of passing on HD risk; and motivation to pursue career goals. People varied in how they reacted to, processed, and communicated about their result. There were many instances of differing attitudes among at-risk siblings, which complicated family communication. Furthermore, many participants reported family trauma beyond HD, including alcohol abuse, physical abuse, and mental illness. These factors highlight the importance of considering an individual's family history, mental health, and social environment to be able to provide appropriate counseling around testing.

3019W

Identifying and counseling patients amenable to mutation-specific therapies in Duchenne Muscular Dystrophy (DMD): Knowledge of resources will fuel genetic counselors' impact. *E. O'Rourke[1], L. Bogue[2], D. Martin[1], M. Pastore[3], A. Smith[4].* 1) Sarepta Therapeutics, Inc., Cambridge, MA; 2) Parent Project Muscular Dystrophy, Hackensack, NJ; 3) Nationwide Children's Hospital, Columbus, OH; 4) University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh, Pittsburgh, PA.

   Mutation specific therapies are currently in clinical trials for treating DMD, a progressive, life-limiting X-linked neuromuscular disorder caused by mutations in the *DMD* gene (Xp21.2) that result in a lack of functional dystrophin. Current approaches include nonsense mutation suppression and exon skipping. The goal of nonsense suppression is to read through the premature stop codon to produce functional dystrophin, whereas the goal of exon skipping is to restore the mRNA reading frame to generate internally deleted, functional dystrophin. Considering both approaches, more than 60% of DMD cases are potentially amenable.   Determining patient eligibility for mutation-specific clinical trials involves barriers that genetic counselors are uniquely positioned to help solve. Access to genetic testing is one such barrier. In a survey of 27 U.S. physicians who manage over 2,000 DMD patients, there was significant variability in the percentage of patients genotyped (~25% - 100%). Awareness of financially viable testing options like the Decode Duchenne program, which provides free genetic testing to eligible individuals, is essential.   Accurate interpretation of genetic test results represents a second barrier between patients and mutation-specific approaches. While nonsense mutations are categorically amenable to nonsense suppression, other *DMD* mutations vary in their amenability to exon-skipping therapy based on size and location in the gene. To counsel patients accurately, clinicians must first be able to discern if a mutation is appropriate for exon skipping and second, whether the specific mutation is amenable to a particular exon skipping therapeutic.   With this level of complexity, access to interpretation tools and guidance is paramount for genetic counselors to advise both patients and other providers. We present and review the *DMD* gene exon map, the Leiden DMD database (http://www/dmd.nl), the Decode Duchenne program website (www.duchenneconnect.org/decode), and the University of Utah mutation tables (http://www.genome.utah.edu/DMD/mutation_tables.cgi) as tools that counselors may use to help determine DMD patient eligibility to exon-skipping.

3020T

Perspectives on genetic testing and return of results from the first cohort of presymptomatically tested individuals at-risk for HD. *K.M. Stuttgen[1,2], R.L. Dvoskin[1], J.M. Bollinger[1], A. McCague[1,2], B. Shpritz[3], J. Brandt[3], D.J.H. Mathews[1].* 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine, Baltimore, MD.

For decades, Huntington's disease (HD) has served as a model for how we think about genetic testing, and its benefits and harms for at-risk individuals and their families. In 1983, HD was the first genetic disease mapped using DNA polymorphisms. Shortly thereafter, presymptomatic genetic testing for HD began in the context of two clinical trials. The presymptomatic testing program (PTP) of the Baltimore Huntington's Disease Project at Johns Hopkins University began in 1986 and enrolled 180 individuals. Experiences with this cohort influenced collective thinking about issues related to genetic testing. Challenges around these issues are still widely discussed and debated. This study is obtaining opinions on and attitudes toward genetic testing from people who enrolled in the Johns Hopkins PTP 20-30 years ago. One-hour semi-structured interviews are being conducted with 20 people found to carry an expanded HD repeat, 20 with normal repeats, and 10 who dropped out of the PTP before disclosure of test results. As part of the interview, participants are asked their opinions on the importance of autonomy in the decision to be tested, whether a formal testing protocol is necessary, whether online direct-to-consumer genetic testing for HD would ever be acceptable, and whether incidental findings (including the presence of HD or other genetic risk factors) should be returned in the context of whole exome/genome sequencing. Results to date suggest that participants believe that the decision to undergo testing should be autonomous, though a small number believe everyone at risk for HD should know his/her genetic status. Additionally, most participants believe that a formal protocol is necessary when undergoing genetic testing for HD. A small number felt that the protocol was helpful but that ultimately a patient has the right to know his/her genetic information without having to go through a formal protocol. Thus far, most participants have stated that DTC genetic testing related to serious disease is unacceptable. The majority of participants believe that those undergoing testing should be asked before testing whether s/he wants to be informed of any incidental or secondary findings. A small number of individuals believe patients should be told of all incidental findings so that they can 1) prepare themselves for the onset of a serious disease, such as HD, and 2) make lifestyle changes in an effort to prevent or mitigate their risk for diseases such as diabetes.

3021F

Well, good luck with "that": Reactions to learning of increased genetic risk for Alzheimer's Disease. *D.T. Zallen.* Dept Sci & Tech in Society, Virginia Tech, Blacksburg, VA. and Virginia Tech Carilion School of Medicine, Roanoke, VA.

The e4 allele at the *apolipoprotein E (APOE)* locus is a marker for increased risk for the common form of Alzheimer's disease (AD2 [MIM 104310]). Major professional organizations have recommended against this genetic test for asymptomatic individuals because of concerns about its effect on those who learn of higher risk (due to the presence of one or two e4 alleles) when there are no validated interventions to postpone or prevent this disease. Clinical trials carried out by the NIH REVEAL study reported that psychological distress from learning of higher risk is temporary and minor. In that study, in line with ethical guidelines, substantial genetic counseling was given to prospective participants. Also people were excluded if they failed to meet psychological criteria. It thus remains unclear if these findings apply to an unselected population of individuals who have such testing. Given the growing demand, it is important to understand the range of reactions associated with a higher-risk genetic-test result. We carried out 20 semi-structured research interviews with individuals who had already learned of their higher-risk status following *APOE* testing. They had undergone genetic testing to estimate their risk for Alzheimer's disease, or had learned of their risk inadvertently via genealogy testing or testing for other health problems. They were recruited through the ApoE4.info education and support group. Qualitative analysis of the interviews showed three main findings: (1) For about half of the participants, learning of their higher risk came as a shock and produced a period of painful, sometimes incapacitating, psychological upset. They were distressed about the failure of providers to prepare them for the information that the test would provide. (2) While nearly all felt that they had benefitted in the long term from life-style changes made after obtaining this genetic information, there was regret that they received only dire prognoses from their doctors and were left largely on their own to decipher medical literature and determine what steps they could take. (3) There was a great deal of concern expressed for those who, unlike themselves, had little support and may still be struggling with their initial shock. Based on these findings, there is definitely a responsibility to prepare people *prior* to any genetic testing for Alzheimer's risk and to actively assist them *afterward* in how best to deal with the results.

3022W

What is the economic impact of whole genome sequencing in child health? A cost-consequence analysis in autism spectrum disorder. *W.J. Ungar[1], K. Tsiplova[1,2], R.M. Zur[1,2], C.R. Marshall[3], D.J. Stavropoulos[3], S. Pereira[2], D. Merico[2], E.J. Young[3], S. Scherer[2]. 1) Child Health Evaluative Sciences, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada; 3) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada.*

   Purpose: Whole genome sequencing (WGS) is at the cusp of translation - poised to exert a profound effect on diagnosis in child health, with implications for health systems, decision-makers, clinicians, patients and families. While chromosomal microarray (CMA) is the first-tier genetic test in autism spectrum disorder (ASD), WGS offers promise in improving diagnostic yield but at a higher cost. The optimal deployment of WGS depends on its added value but deployment decisions are hampered by the absence of comprehensive estimates of associated costs. The objectives were to estimate the microcosts of WGS in ASD and to assess the incremental cost-effectiveness of alternative genetic test strategies. **Methods:** Patient-level and program costs for a clinical WGS service were estimated over 5 years from an institutional perspective. Labour, equipment, depreciation, supplies, follow-up testing, bioinformatics, interpretation, report writing and overhead costs for Illumina HiSeq2500 and HiSeqX platforms were estimated. Future costs were discounted at 3%. Parameter uncertainty was tested with probabilistic sensitivity analysis. An incremental cost-consequence analysis examined incremental costs per unit improvement in diagnostic yield for alternative genetic testing strategies that included CMA and whole exome sequencing (WES). **Results:** The cost per sample was US$2847 (95% CI 2745, 2946) on Illuminia HiSeq X and $4682 (4468, 4894) on Illumina HiSeq 2500 compared to $596 (572, 620) for CMA. Reagent supplies accounted for the largest proportion. Five-year program costs for 300 tests/year were $3.96 million (3.82, 4.10). The cost per additional ASD patient with a positive genetic diagnosis was $24,397 when substituting CMA with CMA+WES and $50,570 when substituting CMA with WGS HiSeq X. **Conclusions:** Although test improvements and newer platforms have resulted in lower laboratory costs, WGS remains costly. Incremental costs to replace older with newer technology exceeded $20,000 to detect an additional pathologic variant in ASD. Further improvements in diagnostic yield will increase the cost-effectiveness of WGS. Additional research is required to assess the impact of WGS in the ASD care pathway to determine effects on outcomes. This study provides comprehensive costs for use in economic evaluations of WGS in ASD and allows for a cost model that can be easily adapted to other pediatric conditions.

3023T

Ring14 International: A worldwide network to fight disorders of the 14th chromosome. *M. Crimi[1], Y. DeWoody[2], I. Deguisne[3], S. Hiwaitzi[4], E. Marin[5], J. Shutte[6], S. Azzali[1]. 1) Scientific Coordination office, Ring14 International, Milan(Italy); 2) Ring14 USA; 3) Ring14 France; 4) Ring14 England; 5) Ring14 Spain; 6) Ring14 Netherlands.*

   Chromosome 14 syndromes are rare genetic disorders. In some few cases, one of the two 14 chromosomes assumes the form of a ring resulting in the "Ring14 syndrome" (omim #616606), characterized by early onset refractory epilepsy, motor & mental retardation, autism and a multitude of other physical challenges. In other 14th chromosome syndromes, the chromatid pair loses or transfers genetic material sharing usually a clinical picture similar with "Ring14". Ring14 Association, founded in Italy in 2002, has recently evolved into Ring14 International (R14) to networking the local chapters worldwide located with the mission is to advance biomedical research towards best care and possible cure. R14 creates and coordinates the action of the independent national chapters, which gather individual families from single countries. R14 helps and assists families, divulges information and oversees the most efficient administration of internal resources. Basically, R14 is the only reference point for hundreds of families around the world and communicates in five languages. In addition to providing support to its families, R14 also promotes (and directly funds) biomedical research initiatives for diseases due to chromosome-14 rearrangements. For R14, the funding strategy has a pivotal role and really drive us to open a path towards new ways of research and clinical support. All funding decisions are subjected to peer-review by a panel of experts and any research initiative is always internationally managed. Most relevant results to-date: 1) 6 research projects directly funded (11 papers already published with the R14 support) on the molecular mechanisms underlying Ring14 and to creating in-vivo & in-vitro models; 2) Clinical projects supported, with a focus on the autistic traits and language impairments; 3) Biobanking project developed: hundreds of biological samples has been already collected (and available) in a biobank; 4) Database released: clinical records are digitalized and online searchable. Moreover, R14 collaborates with other international partners (such as Eurordis and RD-Connect) and plans yearly conferences with researchers and families. In the next future, R14 aims also to coordinate a large-scale project and to open new chapters, which will benefit of our full support for the start-up phase. In conclusion, we believe that the advocacy work being done by a patient association can be expanded beyond just improving the visibility of an underdiagnosed orphan disorder.

3024F

Is knowledge power? Family-centered outcomes from genomic sequencing in children with intellectual disability. *K.B. Brothers[1,2], K.M. East[3], W.V. Kelley[3], C.A. Rich[1], K.M. Bowling[3], E.M. Bebin[4], E.J. Lose[5], S. Simmons[4], J.A. Myers[1], G. Barsh[3,6], C. Finnila[3], D. Gray[3], S. Hiatt[3], N. Lamb[3], J. Whittle[3], M. Amaral[3], M. Thompson[3], R.M. Myers[3], M.A. Rothstein[2], G.M. Cooper[3].* 1) Department of Pediatrics, University of Louisville, Louisville, KY; 2) Institute for Bioethics, Health Policy, and Law, University of Louisville, Louisville, KY; 3) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 4) Department of Neurology, University of Alabama at Birmingham, Birmingham, AL; 5) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 6) Department of Genetics, Stanford University, Stanford, CA.

Background The HudsonAlpha CSER Study is a translational study of the utility of whole genome sequencing for identifying a genetic cause in children or adults with intellectual disability. To date we have enrolled 332 families, with sequencing and analysis complete for 312 families. Due to ongoing scheduling of return of results visits, data collection, and data entry, data from 186 mothers are available for this analysis. Results Findings from these analyses are summarized in Table 1. Pathogenic Finding mothers (N=50) and VUS mothers (N=26) were significantly more likely to recommend sequencing (p<0.001) than No Diagnostic Finding mothers (N=110). Pathogenic Finding mothers also reported significantly higher utility from sequencing than No Diagnostic Finding mothers (p<0.001). VUS mothers reported utility intermediate between the other two groups, although differences between VUS and the other two groups were non-significant. Perceived severity and perceived control was measured at baseline and post return of results. Changes in perceived severity and perceived control from pre- to post-test measurements were small, and were not significantly different among the three result groups. Post-test anxiety also did not differ among the three groups. Conclusions Mothers who receive a diagnostic finding for their child's illness find greater utility in sequencing and are more likely to recommend sequencing than those who do not receive a diagnosis, even if the returned result is explained as a VUS. However, receiving a pathogenic finding or VUS is no more likely to increase perceived control than receiving no diagnostic finding. Levels of maternal anxiety are similar regardless of the outcome of sequencing. Ongoing interviews with parents will help contribute to understanding about the utility families derive from diagnostic sequencing results.

Family-Centered Outcomes Based on Diagnostic Finding Category .

	Pathogenic Finding (N=28)	VUS (N=17)	No Diagnostic Finding (N=63)	p values
Recommend Sequencing	6.56	6.85	5.94	<0.001
Sequencing Utility	5.94	4.36	3.39	<0.001
State Anxiety	38.84	35.62	35.09	0.266
Change in Perceived Severity	0.122	-0.384	-0.018	0.817
Change in Perceived Control	-0.36	-0.461	-0.227	0.239

3025W

TIDE Complex Diagnostic Clinic: An innovative model of multidisciplinary care for complex intellectual disability. *S. Race[1], R. Salvarinova[1], G. Horvath[1], A. Ghani[5], P. Eydoux[6], H. Vallance[6], W. Gibson[3], A. Kalenchuk[2], M. Van Allen[3], M. Connolly[1], S. Stockler[1], C. van Karnebeek[1,4,5].* 1) Department of Pediatrics, BC Children's Hospital, Vancouver, BC, Canada; 2) Department of Psychiatry, BC Children's Hospital, Vancouver, BC, Canada; 3) Department of Medical Genetics, BC Children's Hospital, Vancouver, BC, Canada; 4) Center for Molecular Medicine and Therapeutics; 5) Child and Family Research Institute, University of British Columbia, Vancouver Canada; 6) Dept of Pathology and Laboratory Medicine University of British Columbia, Vancouver, Canada.

Background: When the underlying cause of intellectual disability (ID) is identified patients and their families benefit because it can lead to improved genetic counselling, management and adaptation. However, the diagnostic odysseys for these patients are often long and expensive because of the etiologic heterogeneity of ID. Methods: The multidisciplinary Complex Diagnostic Clinic (CDC) was established in an effort to expedite the diagnostic process for ID patients with unexplained complex systemic and/or neurologic features. Patients are evaluated by 3-4 specialists in one morning. The specialists then meet for multidisciplinary rounds. **Results:** 95 children were evaluated during 27 clinics held over four and a half years. The ages of the patients ranged from 6 months-18 years. 45 were male and 50 were female. All the children had previously been seen by 2-12 specialist services. To date, the diagnostic yield of the CDC includes 32 confirmed and 17 working diagnoses (34% and 18% of children respectively). The diagnosis in patients was identified by: clinical exam in 7 chromosome micro-array in 3 single gene tests in 4 gene panel in 6 clinical whole exome sequencing in 5; and research exome in 7. These numbers will likely continue to increase as pending tests come in. In all cases, detailed phenotyping with clinical history and physical exam was essential for guiding the specific test and/or interpreting genomic data. In some cases the diagnosis enabled more individualized management with targeted intervention or prevention. Surveys showed positive feedback from both families and medical professionals. Chart reviews revealed cost-savings due to fewer trips to hospital and fewer tests via more streamlined evaluations **Conclusions**: The CDC represents an innovative model of multidisciplinary diagnostic care for complex intellectual disability. Careful phenotype along with close communication and collaboration among the specialists enhances the interpretation of relevant clinical, biochemical and genomic data. New diagnoses continue to be made possible due to the collective knowledge of the group, with improved satisfaction of patients, family and health care providers. Finally, diagnosis directly impacted management and treatment beyond genetic counseling and prognosis in 6 patients.

3026T

The impact of genetic counseling on knowledge and beliefs of an indigenous Filipino community with a high prevalence of otitis media due to an *A2ML1* **variant.** *E.M. Cutiongco-de la Paz[1,2], M. Pedro[3], T. Yarza[4], S.M. Lagrana[3], A. Amoranto[1], M.T. Domine[1], E.J. Jover[1], S.M. Leal[5], C.M. Chiong[3,4], R.L.P. Santos-Cortez[5].* 1) University of the Philippines Manila-National Institutes of Health (UPM-NIH), Pedro Gil St., Manila 1000, Philippines; 2) Philippine Genome Center, University of the Philippines, Diliman, Quezon City 1101, Philippines; 3) Philippine National Ear Institute, UPM-NIH, Taft Ave., Manila 1000, Philippines; 4) Newborn Hearing Screening Reference Center, UPM-NIH, Pedro Gil St., Manila 1000, Philippines; 5) Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, 1 Baylor Plaza 700D, Houston, Texas 77030, USA.

   Previously we identified an *A2ML1* duplication variant that confers a four-fold risk for otitis media and induces middle ear microbial differences within an indigenous Filipino community with a ~50% prevalence of otitis media. As fulfillment of the agreement during free and prior informed consent for genetic research, we shared our exome and microbiome findings via a community meeting. Before the meeting, 13 out of 30 (43.3%) heads of households were interviewed regarding their knowledge and beliefs about otitis media. After the meeting, the heads of 24 households (80%) participated in genetic counseling and were asked the same pre-meeting interview questions. Majority (67%) of participants did not graduate from elementary school. Among those who were interviewed twice, after the community sharing the proportion of individuals who believed that otitis media has a heritable component doubled (from 36% to 77%; p=0.04) and those who believed that better personal hygiene (79%) and immunization (71%) can help alleviate otitis media burden increased to 100%. Among all respondents, the belief that it is normal for children to have ear discharge and hearing loss occurred in 42% and 21%, respectively, even after counseling. Those who believed that poor hygiene plays a role in the development of otitis media were younger (p=0.01), were less likely to be early adopters of *A2ML1* testing (p=0.04), submitted more DNA samples at later stages of the study (p=0.0004), and have a lower proportion of *A2ML1* variant carriers per household (p=0.048). Younger respondents also tend to believe that only children can have otitis media (p=0.02). Heads of households with a greater number and proportion of *A2ML1* variant carriers were more likely to believe that otitis media is heritable (p=0.03). When asked what they do when there is ear discharge, 83% reported cleaning their own ears usually with cotton swabs, however only 46% would seek clinical consult. Families with more members with chronic otitis media would clean their own ears (p=0.04), while participants who agreed to pre-counseling interview were more likely to seek clinical consult (p=0.02). Our experiences in sharing genomic information with a marginalized community highlight the challenges of community-based counseling within the context of poor health access, high disease prevalence, high population frequency of a disease-causal variant, and a complex trait with strong gene-environment interaction. .

3027F

More efficient diagnosis of patients with hearing and vision loss: Use of collaborative clinics and molecular genetics. *C. Sloan-Heggen[1,2], A. Ko[3], W. Pfeifer[4], D. Kolbe[1], C. Nishimura[1], K. Frees[1], A.E. Black-Ziegelbein[1], K. Booth[1], C. Campbell[1], S. Ephraim[1], P. Ranum[1], S. Shibata[1], D. Wang[1], A. Weaver[1], H. Azaiez[1], A. Drack[4], R. Smith[1,2].* 1) Otolaryngology, University of Iowa, Iowa City, IA; 2) Department of Molecular Physiology & Biophysics, University of Iowa College of Medicine, Iowa City, Iowa; 3) Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC; 4) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA.

   Hearing and vision are both integral to human interaction with our environment and each other. A syndrome affecting both of these senses can severely impact a person and leave them isolated. Diagnosis of such syndromes can be difficult because they are often nebulous with variable expressivity. Patients may be labeled with an incorrect or incomplete diagnosis for years, missing out on appropriate monitoring or testing and potential delay of symptom identification. To evaluate the benefit of a collaborative diagnostic process, we performed a retrospective study of patients referred to the Multidisciplinary Genetic Ophthalmology-Otolaryngology Clinic at the University of Iowa Children's Hospital. Individuals were seen by both an otolaryngologist and ophthalmologist at a single appointment. Patients were carefully evaluated clinically, diagnostically, and molecularly. Fourteen families (17 individuals) have been evaluated collaboratively to determine the most likely diagnosis. Eleven families (79%) were provided a likely molecular diagnosis (via targeted genomic enrichment and massively parallel sequencing or comparative genomic hybridization) consistent with the presenting phenotype; three families remain under molecular investigation but have likely clinical diagnoses. Importantly, as a result of this multidisciplinary clinic, nine patients received a refined diagnosis. In summary, close collaboration between multiple medical specialties, as well as integration of careful clinical and comprehensive molecular analyses, provides the best avenue through which an accurate diagnosis can be reached and future clinical direction and follow-up can be guided. In this cohort, the combined clinic has allowed for early visual monitoring for individuals with Usher syndrome and refined complex diagnoses like Waardenburg syndrome, Cowchock syndrome, and Baraitser-Winter syndrome.

3028W

Treatment research reporting trends for genetic eye diseases and patient sharing on Twitter: Implications for clinical communication. *T. Bubela, W. Luth, S. Kowal.* School of Public Health, University of Alberta, Edmonton, Alberta, Canada.

   Purpose Reports on highly experimental treatments such as bionic eye, gene therapy, and stem cell research give hope to patients with currently untreatable genetic eye diseases. News media is known to sensationalize translational research, leaving patients with incomplete information about therapeutic potential and research timelines. Research teams craft press releases to present findings; however, easy online information access and sharing allows for repeated reinterpretation of these findings as they move from press releases to newspapers, on to advocacy organizations, and between patients on social media. Our research on information pathways illustrates how reports of research progress change as they move through information outlets and among individual patients. Information accuracy and quality morphs, affecting communication requirements for clinicians and genetic councillors helping patients choose disease management options within current clinical realities. Thus, we also identify challenges and pose solutions for clinical communications about ocular treatment research. **Methods** We conducted a media analysis of press releases and traditional news media in Canada, the USA, and the UK to compare what stories were most heavily reported in which regions and how these stories were presented. Subsequently, we analyzed if and how ocular disease advocacy and support groups passed the same research news on to online users. Finally, we conducted an analysis of a decade of twitter conversations by patients around treatment research to identify which sources and stories patients most share and the ways in which patients discuss novel treatment research. **Results** News media and press releases: 1) overstate benefits and often omit risks of experimental treatments for ocular diseases; 2) conflate research with treatment or cure; and 3) misrepresent research timelines. Some advocacy websites present overly optimistic representations of research realities while others offer more balanced reports. Our twitter analysis shows the types and quality of sources patients and individuals follow and share. **Significance** Our research informs how clinicians can prepare to respond to overly positive reporting and who they should suggest that patients follow for treatment research information. We suggest using optimism towards the future of genetic eye disease research tempered by evidence to make clear the potential and timelines associated with therapies still in clinical trial phases.

3029T

Pharmacogenomics of age-related macular degeneration: A systematic review of the evidence. *A.M. Issa[1], V.S. Chaudhari[1], P. Pednekar[1], W.J. Foster[2].* 1) Personalized Medicine & Targeted Therapeutics, University of the Sciences in Philadelphia, Philadelphia, PA; 2) Dept of Ophthalmology and Dept of Bioengineering, Temple University, Philadelphia, PA.

   Age-related macular degeneration (ARMD) is the leading cause of blindness among older adults. Although the etiology of ARMD is not well understood, it is thought to include both environmental and genetic influences. There is an increased interest in understanding the genetics of ARMD, and particularly in the association of genotypes with therapeutic interventions for ARMD. We systematically assessed the evidence about genetic variants associated with ARMD, and the association of standard treatment options for ARMD with specific genotypes. Methods: A systematic review and quantitative assessment of the English- language literature on the genetics of ARMD between Jan 1996 – Nov 2015 was done. We performed searches using multiple databases including PubMed, Science Direct, Ovid Medline, and the Cochrane Collaboration. Two reviewers independently reviewed articles according to a priori established eligibility criteria using a standard screening form. Randomized controlled trials non-randomized trials, cohort and case control studies were included. Conference reports, commentaries, communications, descriptive and editorial studies were excluded. Results: We initially found 1477 articles related to ARMD genetics. Of these, 733 were original studies showing an association between ARMD and particular genetic variants or genotypes. Thirty one articles met eligibility criteria and showed an association between specific genotypes and specific interventions for ARMD. Of 64 different genetic variants, complement factor H (CFH) polymorphisms were suggested to play a role in ARMD susceptibilities in 124 studies. The genetic variant age-related maculopathy susceptibility 2 (ARMS2) was found to be associated with ARMD in 64 studies; high temperature requirement factor A1 (HTRA1) was suggested to be play a role as an ARMD susceptibility in 48 articles, and the VEGFA gene was associated with ARMD in 34 studies. CFH and ARMS2 was found to be most frequently associated with current standard therapeutic interventions, particularly with anti-VEGF. Conclusions:Current evidence suggests that CFH and ARMS2 appear to contribute the most to ARMD susceptibility. CFH and ARMS2 appear to be most closely associated with response to anti-VEGF therapeutic interventions. Further work remains to be done to explore further the associations between the identified genotypes and pharmacogenomic effects of ARMD interventions.   .

3030F

Coenzyme Q10 and pro-inflammatory markers in children with Down syndrome: Clinical and biochemical aspects. *H. El-Bassyouni[1], M. Zaki[2], E. Youness[3], J. Hussein[4], A. Tosson[5].* 1) Clinical Genetics, National Research Centre, Cairo, Cairo, Egypt; 2) Anthropometry Department, National Research Centre, Cairo, Cairo, Egypt; 3) Clinical Biochemistry Department, National Research Centre, Cairo, Cairo, Egypt; 4) Biochemistry Department, National Research Centre, Cairo, Cairo, Egypt; 5) Pediatrics Department, Abul Rish hospital, Cairo University.

Objective: Evidence of oxidative stress was reported in individuals with Down syndrome. There is a growing interest in the contribution of the immune system in Down syndrome. The aim of this study is to evaluate the Coenzyme Q10 (CoQ10) and selected pro-inflammatory markers such as interleukin 6 (IL-6) and tumor necrosis factor α (TNFα) in children with Down syndrome. Methods: Eighty six children (5-8 years of age) were enrolled in this case control study from 2 public institutions. At time of sampling the patients and controls suffered from no acute or chronic illnesses and received no therapy or supplements. The levels of IL-6, TNFα, CoQ10, fasting blood glucose and intelligent quotient were measured. Results: Forty three young Down syndrome children and forty three controls were included over a period of 8 months (January-August 2014). Compared with the control group, the Down syndrome patients showed significant increase in IL-6 and TNFα (P = 0.002), while CoQ10 was significantly decreased (P = 0.002). Also, body mass index and fasting blood glucose were significantly increased in patients. There was a significantly positive correlation between CoQ10 and intelligent quotient levels, as well as between Il-6 and TNFα. Conclusion: IL-6 and TNFα levels in young children with Down syndrome may be used as biomarkers reflecting the neurodegenerative process in them. Coenzyme Q10 might have a role as a good supplement in young children with Down syndrome to ameliorate the neurological symptoms.

3031W

Creating a medically actionable genetic screening panel for healthy individuals. *E.V. Haverfield, E.D. Esplin, B. Johnson, S. Gandomi, A. Lynch, S. Aradhya, R. Nussbaum.* Invitae Corporation, San Francisco, CA.

Background The 2013 ACMG policy statement recommended 56 genes for the return of incidental findings but noted that this list will, and should, evolve. Catalyzed by accelerating advances in clinical genomics, we generated an expanded gene list where pathogenic variants would be considered clinically actionable. **Methods** We reviewed the ACMG56 list, the ACMG Working Group process, and expanded gene lists published by multiple genomics groups. An expert panel of genetic counselors and medical and clinical geneticists reviewed the clinical actionability of individual genes beyond the ACMG56 using criteria such as penetrance, mode of inheritance, and the availability of published medical management recommendations. **Results** Our novel list of 124 clinically actionable genes includes the ACMG56 plus 17 conditions (24 genes) with increased risk for a cancer-related phenotype, three conditions (38 genes) with increased risk for a cardiovascular-related phenotype, and two conditions (6 genes) with increased risk for other medically actionable disorders, all of which have published guidelines for medical management. **Conclusions** The ACMG policy statement addresses pathogenic variants discovered by diagnostic whole-exome or whole-genome sequencing (WES or WGS). WES/WGS is increasingly available to healthy individuals seeking to proactively inform their healthcare. The high cost and mostly uninterpretable results of these broad tests are obstacles for integration into routine healthcare. A focused gene panel restricted to clinically actionable variants presents an opportunity for healthy patients to partner with their healthcare providers for preventive genetic testing with significant potential to inform personalized medical care. This can significantly impact current clinical genetics practices and necessitate the development of new models for clinical genetic counseling. With decreasing costs of and growing interest in this type of information, the transition from what are considered incidental findings to primary findings represents a novel opportunity for genetic information to be introduced into routine medical practice, which in turn can lead to increased clinical utilization for the preventive care of patients.

3032T

Metabolic diet app suite for inborn errors of amino acid metabolism. *G. Ho[1], K. Ueda[1], R. Houben[3], J. Joa[3], A. Geizen[1], B. Cheng[1], C. van Karnebeek[1,2].* 1) Department of Biochemical Disease, BC Children's Hospital, Vancouver, Canada; 2) Department of Pediatrics, Centre for Molecular Medicine & Therapeutics, Child & Family Research Institute, University of British Columbia; Vancouver, Canada; 3) Health2Media, Vancouver, Canada.

Background Inborn errors of metabolism although individually rare, form the largest group of monogenic disorders. The mutations in the genes result in a loss or gain of function in the protein produced, usually an enzyme or a transporter. This class of genetic diseases are particularly amenable through treatment targeting pathophysiology, often a lifelong dietary restriction combined with supplements and medications. *Objectives* To assist diet treated inborn errors of metabolism (IEM) patients with daily management of their diet. By developing a metabolic diet app suite, we are able to help patients with tracking their nutrient intake and adhering to the specific prescribed diet. *Design and Methods* Nutrient information is based on the Genetic Metabolic Dietitians International (GMDI) MetabolicPro™ database. The application suite was tested by biochemical geneticists, registered dietitians and parents of children with Pyridoxine Dependent Epilepsy and Phenylketonuria. Feedback was submitted through an online survey as well as through the diet app feedback function. The collected initial pilot tester feedback was positive and their suggestions were used to further improve the App suite. *Results* The metabolic diet app suite is a free, user-friendly, online app available on 2 platforms (computer and mobile device) via www.mdapp.org. Fifteen metabolic disorder specific apps and one protein tracking app were created, each containing nutrient information on over 100,000 different food products. The Diet Apps offer functions such as: secured personal user login/password, goal setting, daily intake tracking, food content checks, adding foods and homemade recipes, exportable daily food diary log, and developer feedback. General disorder information is also provided for each diet app.*Conclusion* The metabolic diet apps provide an individualized and reliable source of nutrient information, to aide families affected with IEMs to track and plan their meals. With the metabolic diet app, we hope to improve patient adherence, quality of life and personalization of care. Due to its specific nature, the metabolic diet app is a tool, and therefore not intended to be a replacement for medical-metabolic nutritional professional advice. Future research should evaluate its impact on patient adherence, metabolic control, quality of life and health-related outcomes.

3033F

A taxonomy of medical uncertainties in clinical genome sequencing. *B. Biesecker[1], P. Han[2], B. Bernhardt[3], R. Green[4], S. Joffe[5], B. Koenig[6], I. Krantz[7], K. Umstead[1], L. Biesecker[8].* 1) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 2) Center for Outcomes Research and Evaluation Maine Medical Center, Portland, ME; 3) Division of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Partners Personalized Medicine, Harvard Medical School, Broad Institute of MIT and Harvard, Boston, MA; 5) Division of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadephia PA; 6) Institute for Health and Aging, University of California, San Francisco, CA; 7) Division of Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 8) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, Bethesda, MD.

Introduction: Clinical exome and genome sequencing (CEGS) are introducing new opportunities and challenges. Simultaneously these technologies are generating uncertainties of unprecedented scale that laboratories, clinicians, and patients need to manage. Building on an existing taxonomy of uncertainty in medicine, we developed a taxonomy for uncertainties in the use of CEGS in health care. Methods: Interviews with genomics experts were conducted to expand an initial genome taxonomy into its multiple dimensions. Findings were coded and analyzed and incorporated into a revised taxonomy. Multiple iterations were made after additional expert review. In parallel an interactive taxonomy website was developed to disseminate the taxonomy to researchers to engage them in its further refinement. Results: The proposed taxonomy divides uncertainty along three dimensions: source, issue, and locus, and further discriminates the uncertainties into five layers with multiple domains. We illustrate how the taxonomy can be applied to findings from CEGS and used to guide stakeholders through interpretation and implementation of variant results. Conclusion: The proposed taxonomy of uncertainties allows for better understanding of the unknown dimensions of the clinical implementation of CEGS. It also fosters standardization of terms related to uncertainties for research design, clinical guidelines and discussion of uncertainties of CEGS in publications.

3034W

Risk assessment and counselling training for non-genetics healthcare professionals: Competency-based education for the NHS sickle cell & thalassemia screening programme. *A. Sajoo[1], L. Oni[2], C. Peart[3], J. Walters[4].* 1) Department of Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, UK; 2) Brent Sickle Cell and Thalassemia Centre, Central Middlesex Hospital, London, UK; 3) South East London Sickle Cell and Thalassemia Centre, London, UK; 4) Faculty of Nursing & Midwifery, Kings College, London, UK.

In 2010, the National Screening Committee of the National Health Service (UK) put its PEGASUS hemoglobinopathy counsellors training course up for tender offering 25 funded places annually. Post-registration professional training is moving to Institutions of Higher Education. Background: A blended learning initiative was developed with two, two-day episodes separated by four weeks self-directed learning. Topics included National programme specifications, review of Medelian genetics, a survey of hemoglobinopathies and interpretation of lab reports. Culture, ethnicity, communication and counselling methods were discussed in the context of carrier identification, risk assessment, reproductive options and decision-making. The teaching team consisted of specialist nurses, academic and laboratory staff as well as a genetic counsellor. Audit, laboratory testing, reporting compliance and research issues were covered by guest lecturers. Outcomes included proficiency in genetic risk assessment, effective communication of lab results and reproductive options, as well as attaining genetic competencies. Learners agreed that their reactions to new material mimicked their patients' anxieties when receiving test results in pregnancy. Results: There was an average 80% evaluation response rate with 30-40 respondents each year. Comments are categorised into content, delivery and assessment. Teaching recessive genetics using the Punnett square and via 'mating games' were mentioned as being most effective teaching strategies. Particiants remarked on the interdisciplinary dynamic which enriched case studies. Those new in post appreciated the 'just in time' learning. Discussion: Evaluations demonstrated that learners began to feel confident integrating family history with laboratory results in the last two weeks of the course. Roleplay allowed learners to practice their skills. Barriers to learning included difficulty grasping complex genetic concepts and working out risks for rarer Hb variants. Scaffolding content, self-directed learning and networking with clinical colleagues and lab staff helped surmount barriers. Annual evaluation demonstrates that the curriculum meets the needs of highly skilled NHS staff. Learners have requested a longer course, update days and more time spent on rare variants. In response, course content has shifted, but no additional time is to be funded. Summary: Learners taking this course meet their nursing and midwifery-related genetic competencies.

3035T

A collaborative genetic and genomic education pilot for health insurer staff: Development and implementation of a 13-episode webinar series. *R. Wildin[1], S. Belinson[2], H. Brown[2], V. Pratt[3], D. Flannery[4], D. Messersmith[1], Inter-Society Coordinating Committee Insurer Education Working Group.* 1) Division of Policy, Communications, and Education, Nat'l Human Genome Research Inst., Bethesda, MD; 2) Blue Cross Blue Shield Association, Chicago, IL; 3) Indiana University, Dept. Medical and Molecular Genetics, Indianapolis, IN; 4) American College of Medical Genetics, Bethesda, MD.

Rationale: Equipping healthcare providers to incorporate genetics and genomics into clinical practice is a universally identified need, yet efforts have not addressed the requirements of insurer medical directors and staff. The National Human Genome Research Institute's (NHGRI) Inter-Society Coordinating Committee for Provider Education (ISCC) collaborated with the Blue Cross Blue Shield Association (BCBSA) to produce a webinar series targeted to this group. The goal was for insurers to understand genetic testing strategies, interpretations, outcomes, and utility for patient care, and how their use enhances sound health decision making for their insured. **Methods:** The ISCC's Insurer Education Working Group (WG) developed a list of topics essential to understand advances in genetics and genomics in clinical practice, and a case-guided lecture framework. Presenters were subject-matter experts identified by the WG members. BCBSA piloted the activity and assisted presenters in optimizing content and presentation format. Each month, new live webinar content was presented to pre-registered BCBSA staff through their learning system, with all sessions recorded. Primary outcomes included enrollment and attendance counts, standard learning evaluations, and participation in interactive post-presentation discussions. User access and view statistics for recordings posted publicly on https://www.genome.gov/27563343/ provide a secondary, continuing outcome measure. **Results:** Thirteen hour-long webinars were presented between June 2015 and June 2016. Individual session content included foundational genetics knowledge (language of genetics, choosing a test, what can be diagnosed), clinical pathology (validity and utility, CPT coding), clinical applications of next generation sequencing (hereditary cancer, non-cancer, targeted cancer therapy, genome sequencing), regulation, and ethical testing and counseling. On average, at least 60 staff attended the webinars. Evaluations and qualitative discussions indicated general interest in and appreciation for the content among attendees, along with desire for more definitive data supporting clinical utility. **Conclusions:** Insurer medical staff had, perhaps for the first time, access to organized lectures addressing advances in, uses of, and challenges for genetics and genomics in healthcare. Continued strong attendance throughout the year suggests novel information was imparted and sufficient value was found to propose expansion.

3036F

Familial Creutzfeldt-Jakob Disease: Case report and consequences of post mortem testing without appropriate genetic counseling. *K. Clift[1], K. Guthrie[1], E. Klee[2], N. Boczek[2], M. Cousin[2], P. Blackburn[1], P. Atwal[1].* 1) Center for Individualized Medicine, Mayo Clinic, Jacksonville, FL; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN.

Here we present a case of an asymptomatic 53-year-old woman who sought genetic testing for Familial Creutzfeldt-Jakob Disease (fCJD [MIM 123400]) after inadvertently learning that her mother had fCJD. The patient's mother had a sudden onset of memory problems and rapidly deteriorating mental faculties in her late 70's, which lead to difficulties ambulating, progressive non-fluent aphasia, dysphagia and death within ~1 year of symptom onset. The cause of death was reported as "rapid onset dementia." The patient's family, unhappy with the vague diagnosis, aggressively pursued causation and submitted frozen brain tissue from the mother to the National Prion Disease Surveillance Center where testing revealed a previously described 5-octapeptide repeat insertion (5-OPRI) in the prion protein gene (PRNP [MIM: 176640]) that is known to cause fCJD. The family had additional questions about the implications of this result and thus independently sought out genetic counseling. While rare, fCJD is likely underdiagnosed due to clinical heterogeneity, rapid onset, early non-specific symptomatology, and overlap in the differential diagnosis of Alzheimer's disease and Lewy body dementias. When fCJD is identified, a multidisciplinary approach to return of results that includes the affected patient's provider, genetics professionals, and mental health professionals is key to the care of the family. We present an example case which discusses the psychosocial issues encountered and the role of genetic counseling in presymptomatic testing for incurable neurodegenerative conditions. Ordering physicians should be aware of the basic issues surrounding presymptomatic genetic testing and identify local genetic counseling resources for their patients.

3037W

Utilization of telehealth in a translational genome-wide sequencing clinic in British Columbia. *A.M. Elliott[1,2,3], C. du Souich[1,2,3], S. Adam[1], J. Gillis[4], C. van Karnebeek[3,4], A. Lehman[1,2,3], J.M. Friedman[1,2,3].* 1) Dept of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Provincial Medical Genetics Program, BC Women's Hospital and Health Centre, Vancouver, BC, Canada; 3) Child and Family Research Institute, Vancouver, BC, Canada; 4) Division of Biochemical Diseases, Dept of Pediatrics, BC Children's Hospital, Vancouver, BC, Canada.

Background: Telehealth (TH) involves the use of information and communication technology to deliver health services to patients over distance. The time and cost of travel can constitute a barrier for some families to access centralized tertiary healthcare services. The use of TH in genetic counseling for common indications (e.g. advanced maternal age, cancer) has been reported; however, its use with genome-wide sequencing has not been well characterized. The CAUSES Research clinic, a three-year initiative to sequence 500 British Columbia (BC) pediatric patients with suspected genetic disorders, provides patient evaluation, genetic counseling, exome sequencing (ES), clinical interpretation of the genomic data and follow up. All families receive both pre- and post test genetic counseling that includes discussion of complex issues related to ES and obtaining informed consent. Patients who live in remote areas and have been recently evaluated by a clinical geneticist are offered the option of TH. *Methods:* We analyzed the number and type of TH encounters within the CAUSES clinic process. Additional variables studied included indications for ES and regions of TH uptake within BC. *Findings:* TH has been utilized at multiple stages of the CAUSES clinic process. Of the first 145 families, 22 (16%) were seen via TH for their pre-test genetic counseling. Clinical indications for ES included: 17 patients with intellectual disability +/- other anomalies (e.g. dysmorphisms, seizures), 2 with multiple congenital anomalies, and 3 with complex disorders (e.g. myopathy). Seven of the 8 major regions of BC utilized TH with 31% of families residing on Vancouver Island, where access to the BC Children's Hospital is via air or ferry only. TH was also utilized to provide post-test genetic counseling (results) to families and for clinical presentations with referring physicians who practice in remote areas to discuss annotated variant lists. *Conclusions:* TH is well utilized in the CAUSES clinic, for pre- and post test genetic counseling as well as for clinical presentations/consultations with referring clinicians. TH is an important consideration when implementing ES in research and clinical programs, allowing access to this technology with informed genetic counseling for families in remote areas. Additional advantages include decreased cost and travel for patients and clinicians and reduced time away from family and community.

3038T

Scaling genetic counseling: A human-centered approach to creating an electronic GC platform (eGC). *E. Levin[1,2].* 1) Icahn School of Medicine at Mount Sinai 1 Gustave L. Levy Place New York, NY 10029-5674; 2) Helix Inc 1 CIRCLE STAR WAY FL 2 San Carlos, CA 94070.

  BACKGROUND: There are not enough genetic counselors (GCs) to meet the expanding genetic testing market demand, increasing wait times, and potentially negatively impacting both the patient and genetic counselor experience. To address this challenge, the Icahn School of Medicine at Mount Sinai sponsored an interaction design project to lead a 3-phase project to discover opportunities for developing a scalable platform that will help genetic counselors, patients, labs, and ordering providers and enhance the overall genetic/genomic testing experience. **DESIGN**: The project applied a qualitative, human-centered design approach to assess the current challenges and opportunities around prenatal carrier screening. Phase 1: Collaborate with key stakeholders to create concept sketches. Phase 2: Conduct deep, in-person user interviews with patients and/or couples who had gone through prenatal carrier screening in the previous 12 months as well as genetic counselors in both the clinic and lab. Phase 3: Design user experience architecture and key workflows based on iterative user feedback. **OUTCOMES**: Four main areas were identified for deep study: Automate and enhance workforce efficiencies; centralize data aggregation and documentation; enable on-demand information access and transparency; and improve user experience for all stakeholders. **CONCLUSIONS**: GCs are overloaded with logistics and repetitive tasks that steal time from actual counseling. Patients bounce back and forth across stakeholders almost at every step, creating an anxiety-inducing process in which they do not feel in control. As a result, the design team focused on creating two "experiences". The Patient Experience is a scalable, institution-agnostic self-service experience curated by the GCs. The Genetic Counseling Experience, is a primary tool to help stage the patient's journey and streamline operational workflows. Designs were created, tested, and iterated with both stakeholder groups to yield a detailed framework for software execution.

3039F

PhenoDB - A web-based next generation sequencing educational platform for medical and graduate students, residents, clinicians and counselors. *S. Liao[1], H. Wick[2], F. Schiettecatte[3], A. Hamosh[2], D. Valle[2], J. Bodurtha[2], N. Sobreira[2].* 1) Johns Hopkins University School of Medicine, Baltimore, MD 21205; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205; 3) FS Consulting, Salem, MA.

  Next Generation Sequencing (NGS) has become a major diagnostic tool throughout medicine. Its use has allowed for the molecular diagnosis of many inherited phenotypes and malignancies, identification of pharmacogenetic susceptibilities, and discovery of individual genetic variation that confers risk for common complex traits. Thus, it is critical for physicians in all areas of medicine to be familiar with NGS technology, indications, data analysis and interpretation, genetic counseling, and ethical considerations. Here we present an online, problem-based, teaching tool designed for medical and graduate students, residents, clinicians and counselors. It includes a didactic part and hands-on learning exercises that will allow the users to: i) understand the differences, applications and consequences of single gene, genes panels, WES and whole genome sequencing (WGS) in clinic; ii) become familiar with laboratory and analytic techniques used to generate NGS data; iii) become familiar with various databases and programs used to evaluate sequence data and interpret variants; iv) become familiar with variant classification criteria and incidental findings and how these results are reported; and, v) appreciate the need for pre- and post-test genetic counseling for sequence-based tests. The teaching tool was developed as part of PhenoDB, a web-based tool for storage and analysis of phenotypic and genomic data, freely available at www.phenodb.org (Sobreira at al., 2015). The hands-on learning exercises include 15 clinical cases created by adding specific published disease variants to HapMap WES data. The clinical cases are relevant to different medical specialities and the WES data can be analyzed using the PhenoDB Variant Analysis Tool. Each clinical case includes a list of questions that addresses one or more of the objectives listed above and that can also be used as an assessment tool. The didactic part is under development and includes a series of one hour classes specific for each objective that will be stored and made continuously available for access in the PhenoDB teaching tool together with the clinical cases. As of today, the PhenoDB teaching tool has been used to teach graduate and medical students, post-docs, medical genetics residents and counselors with positive evaluations. We plan on making this resource freely available on line for further dissemination and to enhance basic and applied understanding and use of genetic information in healthcare.

3040W

A community genomics initiative using blood banks for recruitment enables access to a diverse, altruistic population. *E. Ramos[1], M. McGinniss[1], E. Thorpe[1], L. Fosler[1], K. Schahl[2], D. Eklund[3], H. Bixenman[3], D. Morton[3], D. Wellis[3].* 1) Illumina, Inc., San Diego, CA; 2) InformedDNA, St. Petersburg, FL; 3) San Diego Blood Bank, San Diego, CA.

   Multiple studies are assessing the future of personal genome sequencing but the majority of participants fall into a narrow demographic profile – wealthy, highly educated and Caucasian. This lack of diversity is a barrier to assessing issues that may arise when implementing clinical whole genome sequencing (cWGS) in a population as diverse as that of the United States. Conversely, blood bank donor populations mimic U.S. population diversity. San Diego Blood Bank (SDBB), in collaboration with Illumina, initiated an IRB-approved pilot project to determine if blood banks can be used to recruit a diverse group of participants for genomics research.   SDBB sent a 13-question survey via email to approximately 150,000 active, inactive, and deferred donors. Donors were asked if they would be interested in participating in a research study that would follow people's health over time to understand risk and disease and what type of data they would be willing to provide. 7,177 unique responses were collected and 71.7% stated that they would provide genetic information. 70 individuals, demographically matched to the San Diego population, were enrolled in the study. The cohort's self-reported ethnicity was 27% Caucasian, 26% Hispanic/Latino, 13% Black/African-American, 11% Asian and the remainder other/two or more races. 67% (41/61) reported a household income of <$100,000.   Participants had cWGS and received interpreted results for about 1200 Mendelian disorders and a pharmacogenomics panel. The SDBB cohort was compared to a 2016 U.S. UYG cohort with a homogeneous "typical" profile. The average number of pathogenic/likely pathogenic variants per participant was similar between the SDBB and U.S. cohorts (2.17 vs 2.69). The average number of variants of unknown significance (VUS) was higher for the subset of the SDBB cohort that did not self-report as Caucasian only when compared to the U.S. UYG cohort (1150 vs 1046). Two SDBB participants (2.9%) had a likely pathogenic variant in one of the 56 "ACMG genes", which is consistent with other studies.   The U.S. Precision Medicine Initiative has committed to recruiting a diverse cohort of participants. This pilot demonstrates that blood banks are trusted entities that can be leveraged to recruit large numbers of diverse volunteers willing to participate in genomic research. Additionally, results suggest that while a higher VUS rate may be seen with increasing diversity, participants are equally likely to benefit from testing.

3041T

The Brazilian Initiative on Precision Medicine (BIPMed): The first publically available genomic database in Latin America. *C.S. Rocha[1], B.S. Carvalho[2], I. Lopes-Cendes[1] for the BIPMed Collaborative Group and the Brazilian Institute of Neuroscience and Neurotechnology.* 1) Department of Medical Genetics, School of Medical Sciences - Unicamp, Campinas, São Paulo, Brazil; 2) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing - Unicamp, Campinas, São Paulo, Brazil.

   Introduction: BIPMed is an initiative of five Research Innovation and Dissemination Centers established in the state of Sao Paulo, Brazil. Precision Medicine has emerged recently as a concept in which scientific knowledge and technology will come together to provide the basis for the 21st century medicine. It involves translational research, genomics and personalized medicine to propose a new level of data integration in order to improve health care. The BIPMed genomic database is the first product released by our initiative to fulfill the increasing need for publically available genetic and genomic information about the Brazilian population; however our aim is that it becomes a reference for data sharing for different populations in Latin America. BIPMed is also part of the Beacon Project, which is a data discovery engine supported by the Global Alliance for Genomics and Health (GA4GH). **Materials and Methods:** BIPMed genomic database is currently based on a software platform, the Leiden Open Variation Database and it was implemented as a fully web-based gene sequence variation database. The design of the database follows the recommendations of the Human Genome Variation Society and the principles and guidelines of the GA4GH for the ethical and responsible sharing of genomic and clinical information. In addition, two other databases were designed to store and display information from SNP-arrays and RNA-sequencing data. Furthermore, a platform able to integrate different genomic information is under development. **Results:** Currently, the database contains variants detected using whole exome sequence as well as SNP-genotyping from reference population ascertained based on their geographic origin in Brazil. In this dataset we identified over ten million variants in 20842 genes. There were 209 variants which were unique to the population studied. In addition, our genomic database has attracted world-wide attention and it has been accessed by an average of 165 hits daily from countries all over the globe. **Discussion and conclusion:** We expect the database to grow fast and include data-sets from disease cohorts as well as other types of –omics data. This platform, the first of its kind in Latin America, is intended to be used by clinicians and scientists worldwide, to obtain and share information about various aspects of genomic medicine and human health, as well as to support dissemination and training. Supported by: FAPESP, BRAZIL.

3042F

Lowering barriers to accessing genomic information: Enabling a search function within the ClinGen Resource. *M.S. Williams[1], C.L. Overby[2,4], W.S. Rubinstein[3,4], D.R. Maglott[3,4], T.H. Nelson[5], A. Milosavljevic[6], C.L. Martin[5], S.R. Goehringer[7], R.R. Freimuth[7], G. Del Fiol[8], B.S.E. Heale[9].* 1) Dir, Genomic Med Inst, Geisinger Health System, Danville, PA; 2) Johns Hopkins School of Medicine, Baltimore, MD; 3) The National Center for Biotechnology Information, Bethesda, MD; 4) National Institutes of Health, Bethesda, MD; 5) Geisinger Health System, Autism & Developmental Medicine Institute, Danville, PA; 6) Baylor College of Medicine, Houston, TX; 7) Mayo Clinic, Rochester, MN; 8) University of Utah, Salt Lake City, Utah; 9) Intermountain Healthcare, Salt Lake City, Utah.

    Statement of Purpose: The NIH-funded Clinical Genome Resource (ClinGen) is dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. The Electronic Health Record Working Group (EHRWG) ensures that the ClinGen resource is designed to be accessible to providers and patients through EHR and related systems. Here we describe the functionality and design choices for the ClinGen website (clinicalgenome.org) that allows searching across all ClinGen resources as well as selected external content collections. The technical aspects of the search function have been described elsewhere[1]. **Methods Used:** An initial resource list was compiled by the chair of the EHRWG and circulated to the ClinGen steering committee that identified additional resources. The expanded list was presented to attendees of the ClinGen session at the 2014 ASHG annual meeting to ensure completeness and assist prioritization. Once the resources were assembled on the webpage, the EHRWG enabled the search of resources by disorder, by medication or by gene name. To facilitate search term selection, the search interface assists users as they type by presenting terms and synonyms obtained from the resources. Functionality and design choices for the resource were presented to the ClinGen Steering Committee, and to attendees of the ClinGen session at the 2016 ACMG annual meeting. ACMG meeting attendees were encouraged to interact with a live version of the resource during the meeting. Over 30 individuals provided informal feedback which was compiled, summarized and documented. Web navigation statistics are being collected to provide additional usage data. **Summary Results:** Input from groups listed above was used to optimize the functionality and design choices. ClinGen native resources were added when available and optimized to return relevant annotated content for disease, gene and medication. Information was provided to non-ClinGen resources to facilitate return of content in response to a single query. Since 5/27/2015 over 5,000 searches have been performed. 36 different resources have been accessed following the search for a variety of topics. Future work will involve formal usability testing with a diverse group of end users and improve interfaces with EHRs. [1]Heale BSE, et al. Integrating Genomic Resources with Electronic Health Records using the HL7 Infobutton Standard. Appl Clin Inform (under review).

3043W

Development of Korean genome reference material. *J.N. Kim[1], I.C. Yang[3], Y.K. Bae[3], J. Sung[2].* 1) Seoul National University, Seoul, Seoul, South Korea; 2) Complex Disease and Genome Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul, Korea; 3) Korea Research Institute of Standards and Science, Center for Bio-Analysis, Daejeon, Korea.

    The Asian Reference Genome Project aims to complement current reference genome (hg38) by adding Asian-specific alternative sequences. Although the hg38 is widely used in the field, demand and importance for ethnically specific reference genome sequences has been recognized. It was reported that this could be evidence for the emergence of novel sequences, resulted from integration of ethnically specific genomic information. In this project, therefore, we are developing Korean genome reference material to enable comparative characterization of Korean-specific sequences. To this end, the study is constructing an Asian-specific reference sequence based on deep sequencing of one individual (AK1), 3 trios and 2 identical twin pairs. One of the research products from the study will be the Korean genome reference material (gRM), which is similar to the reference material developed by the NIST (National Institute of Standards and Technology). A reference genomic DNA with known sequence will serve as a "correct answer" for different analytic platforms and bioinformatics tools, to evaluate the validity of diverse genome technologies. For development of the gRM, DNA after homogenizing was analyzed by multiple sequencing platforms including short read whole genome deep sequencing by Illumina (X10, depth x90), long read analysis by PacBio,(up to x100), and virtual long read analysis using 10X platform. The presentation summarizes the current status of gRM development. This study is open to collaborations at any level.

3044T

The Genomic Consultation: A clinical service designed to enrich patient selection for exome sequencing. *A. Lehman[1], C. du Souich[1], S. Adam[1], J. Gillis[2], C. Van Karnebeek[2], J.M. Friedman[1], A. Elliott[1].* 1) Dept Medical Genetics, Univ British Columbia, Children's and Women's Health Centre, Vancouver, BC, Canada; 2) Dept Pediatrics, Univ British Columbia, BC Children's Hospital, Vancouver, BC, Canada.

Background: Exome sequencing (ES) is not routinely available in British Columbia. Physicians must submit an application to the provincial insurer, the Medical Services Plan (MSP), and funding is granted in restricted situations. Recently, access to ES increased via establishment of the CAUSES research clinic (a 3-year initiative to perform parent-child trio ES for 500 BC pediatric patients), which prompted the development of a clinical Genomic Consultation Service (GCS) to provide genomic advice to physicians considering ES for their patients. Inclusion and exclusion criteria for ES were identified, and education sessions for referring physicians were conducted prior to GCS and CAUSES clinic launch. A panel to review GCS referrals was comprised of geneticists, a pediatric subspecialist, and genetic counselors. Medical records of patients are reviewed weekly, and genomic advice (e.g. targeted testing, genomic panel, ES, referral to Medical Genetics for further evaluation) is provided in a letter. The GCS was established in April 2015. Over 400 referrals were reviewed in the first year. *Findings*: Analysis of the first 350 GCS patients revealed the most common indication was intellectual disability (ID). Many patients had ID in combination with other anomalies. Median age was 7 years (average 7.8 years). Medical genetics was the most common referring discipline (58%), followed by biochemical diseases (19%) and neurology (12%). Others included hematology/oncology, general pediatrics, endocrinology, cardiology and neuropsychiatry. Outcomes: 67% of referrals were appropriate for ES and subsequently invited to participate in the CAUSES clinic; 12% were declined due to lack of suitability and a targeted clinical genomic test was recommended for 9%. The remainder received other advice, including referral to Medical Genetics for further evaluation or whole genome sequencing. *Conclusions*: The GCS enriches selection of patients suitable for ES and provides physicians with patient-specific genomic advice, which can support applications for insured funding. This service enhances clinical implementation of ES, a test perceived as costly to the health care system.

3045F

Patients and their families and friends as developers of medical treatments/devices. *V. Francisco[1], S.A. Oliveira[1], P. Oliveira[2].* 1) Instituto de Medicina Molecular, Lisboa, Portugal, Portugal; 2) Católica-Lisbon School of Business and Economics.

This study looks at the sources of health care innovations and finds that patients of chronic diseases and their non-professional caregivers have developed a very significant number of non-drug medical innovations that have proven valuable in dealing with their diseases, namely by improving their quality of life. Many of these innovations were also evaluated as novel by expert medical evaluators. In some cases, patients even saved their own lives. These innovations usually occur behind closed doors and might never be knowledge or used by anyone else. However, if successful solutions and knowledge were shared with other patients with similar need, it could improve the lives of many others. One way to intervene is to reduce the diffusion costs and develop a centralized inventory of patient developed solutions. With this in mind, Patient Innovation – https://patient-innovation.com – was developed as a nonprofit international, multilingual and open platform, designed to allow patients and caregivers to show and share the innovative solutions they developed to fight their diseases, as well as to foster collaboration among patients, caregivers and others. This experimental platform is aimed at increasing both the rate of patient innovations and its diffusion. Prestigious institutions and reputable individuals, including several Nobel Laureates, distinguished scholars and patient associations from around the world have endorsed it. From February 2014 to September 2015 the Patient Innovation platform collected and curated about 400 innovative solutions developed by patients and caregivers from 30 countries with the USA, Portugal, England, Australia and Brazil as the main contributors. In this poster we will report the developments of the project and the implications for health-care innovation practice and policy.

3046W

Workshop in applied genomic medicine: The outcome. *A.D. Gilbert[1], C.R. Marshall[1,5,6], P.N. Ray[1,3,4,5,6], S.W. Scherer[1,3,4,5], R.D. Cohn[1,2,3,4], S. Bowdin[1,2].* 1) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada; 4) Departments of Paediatrics and Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hosptial for Sick Children, Toronto, ON, Canada.

   Background: The introduction of clinical genome and exome sequencing (CGES) has placed clinical genetics healthcare professionals (CGHP) at the forefront of genomic medicine. These diagnostic modalities are changing not only the scope of medicine, but also the overall scope of practice for the CGHP. Designing relevant educational material and providing venues for geneticists and other health care providers to acquire the skills necessary to integrate CGES into patient care is of paramount importance. A team from SickKids, The Centre for Applied Genomics (TCAG), and the University of Toronto designed and hosted a workshop that addressed the evolving educational needs. **Goals:** *At the end of this workshop, participants will be able to:* integrate CGES into their clinical practice, assess phenotypic and bioinformatics data, debate counseling and ethical considerations, and describe translational and research methodologies. **Method:** SickKids Hospital hosted a 2.5 day applied genomic medicine case-based workshop combining a didactic component and interactive breakout sessions, focusing on the key components of performing clinical genomic sequencing and returning results. Experts from Canada, the US, and Ireland comprised the team of 19 instructors and facilitators. The target audience was primarily clinical genetic providers, trainees and bioinformaticians from across North America who are using CGES. To ensure the participation of both geneticists in training and geneticists in underprivileged nations, scholarships were available. **Results:** 39 genetic professionals from across North America registered for the workshop. Scholarships were awarded to all the trainees that applied. Data was obtained from a pre and post question set that was distributed and completed on the first and last day of the workshop. In addition, a feedback survey was electronically distributed to the attendees. 37 (95%) commenced the feedback survey, with 20 (54%) completing it. **Conclusion:** Based on the overwhelming positive feedback, this workshop was successful and achieved the educational goals set forth by the organisers. Further, 100% of the survey respondents agreed or strongly agreed that the workshop was excellent, and 100% would unequivocally encourage a colleague to attend in the future.

3047T

Enhancement of undergraduate education using a collaborative model to engage K-12 students through genetics literacy, human genetics and GMO investigation. *T.N. Turley-Stoulig[1], T.M. Tinney[2], O. Bell-Hanegan[1,2], K. Bokun[1], M. Celestine[2], R. Corbett[1], L. Davis[1], J. Frosch[1], J. Jacoby[1], I. Juarez[1], J. Khanal[1], A. McIntyre[2], M. McLaurin[1], K. Phan[1], K. Provost[1,2], S. Walker[1].* 1) Dept of Biological Sciences, Southeastern Louisiana Univ, Hammond, LA; 2) Northshore Technical Community College, Mandeville, LA.

   In the rapidly advancing field of genetics and the technologically fast-paced world of modern science, there is a monumental need to interest students in genetics, and STEM in general, early on not only for future study in science-related fields, but also to educate them so they may better understand advances being made as young adults growing up in such an era. The purpose of this study was to increase genetics-based knowledge and understanding, increase awareness of genetics technologies and research and encourage study in genetics and STEM disciplines in general for a range of students at progressive levels. We used an innovative, multi-dimensional approach involving a university/community college partnership coupled with outreach that spans the fields of genetically modified organisms (GMOs) and human genetics by bringing laboratory techniques of both disciplines into the classrooms of high school students, exposing them to genetics-based materials they would otherwise not encounter. Students were educated in basic genetics through a series of primer modules and trained in DNA isolation, PCR and GMO analysis through a series of practical modules. Instruction incorporated a variety of modalities, including animations and hands-on techniques. The study spanned the academic year and involved a ladder of learning whereby a cohort of trained undergraduate students from Southeastern Louisiana University became the teachers of a student cohort from Northshore Technical Community College. In turn, the undergraduate cohorts then team-taught high school students both in their high school classrooms and on the Southeastern campus. The introductory modules were sequenced as Primer Modules I and II, covering basic genetics concepts, human genetic disease, GMOs, and modern laboratory techniques. Practical Modules I and II provided laboratory extensions for conducting GMO experiments using techniques discussed in the Primer Modules. At the high school level, results showed an increase in student understanding of genetics concepts and a stronger interest in the study of cells and/or genes in the future. Similar results were seen for both undergraduate cohorts. We conclude that this approach improves the educational experience of community college and university undergraduate students as well as high school students and bridges the three groups in an invaluable learning experience. We plan to extend this study to evaluate the effectiveness of these methods further.

3048F

Supporting genetics in primary care: Investigating how theory can inform professional education. *B.J. Wilson[1], R. Islam[2], J.J. Francis[3], J.M. Grimshaw[1,2,4], J.A. Permaul[5], J.E. Allanson[6], S. Blaine[5,7], I.D. Graham[1,2], W.S. Meschino[8,9], C.R. Ramsay[10], J.C. Carroll[5,11].* 1) School of Epidemiology, Public Health and Preventive Medicine, Univ Ottawa, Ottawa, ON, Canada; 2) Centre for Practice Changing Research, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON Canada; 3) School of Health Sciences, City University London, London, UK; 4) Department of Medicine, University of Ottawa, Ottawa, ON, Canada; 5) Department of Family Medicine, Mount Sinai Hospital, Toronto, ON, Canada; 6) Department of Paediatrics, University of Ottawa, Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 7) STAR Family Health Team, Stratford, ON, Canada; 8) Genetics Program, North York General Hospital, Toronto, ON, Canada; 9) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 10) Health Services Research Unit, University of Aberdeen, Aberdeen, UK; 11) Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada.

Background Many studies have shown that barriers exist to the integration of genetic case finding into primary care. To obtain deeper insight into professional education strategies, we examined the determinants of three behaviors related to using breast cancer genetics referral guidelines effectively. Methods The behaviors of interest were: *taking a family history*, *making a risk assessment*, and *making a referral decision*. These mapped on to practice guidelines available to participants at the time of the study. Using vignettes of primary care consultations with hypothetical patients as a cue, we used the Theory of Planned Behavior to develop a survey instrument to capture data on intention for each behavior of interest (as a proxy outcome) and its predictors (attitude, subjective norm and perceived behavioral control), and mailed it to a sample of Canadian family physicians (FPs). Correlation and regression analyses explored the relationships between predictor and dependent variables. Results The response rate was 96/125 (77%). The predictor variables explained 38-83% of the variance in intention across the three behaviors. For the *family history* behavior, the data suggested that, although FPs appeared consistently supportive of family history taking in inherited cancer risk assessment, a proportion were in fact skeptical that this should be part of their own practice and, for some, low intentions were associated with lack of confidence. For the *risk assessment* behavior, we found a broad set of reactions, from respondents extremely positive in all respects, to those who clearly did not intend to carry out this behavior, saw no value in it, felt little pressure from colleagues to do it, and/or would not feel confident in doing it anyway. For the *referral decision* behavior, we found overall high intentions, but negative as well as very positive evaluations across all domains. The most striking was for perceived behavioral control. Our observations suggest potential utility in tailoring educational interventions specifically to address different barriers to using genetics in practice. For example, techniques such as social comparison or modelling may address skepticism that a behavior should be expected of a FP (subjective norm), whereas graded tasks or behavioral rehearsal may support low confidence about skills (perceived behavioral control). Conclusion Theory-based needs assessment may offer an effective approach to tailoring genetics education for FPs.

3049W

Discordance in selected designee for return and the legal recipient of genomic findings in the event of participant death. *L.M. Amendola[1], M. Horike-Pyne[1], S.B. Trinidad[2], S.M. Fullerton[2], B.J. Evans[3], W. Burke[1,2], G.P. Jarvik[1,4].* 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Bioethics and Humanities, University of Washington; 3) Center for Biotechnology & Law, University of Houston Law Center; 4) Department of Genome Sciences, University of Washington.

Introduction: The Office for Civil Rights in the U.S. Department of Health and Human Services has provided guidance that HIPAA-covered entities are allowed to disclose a deceased person's health records to: (1) a personal representative whom the decedent appointed, prior to his or her death and (2) members of the deceased person's family, even if the deceased person did not appoint them as a personal representative, unless the deceased person had expressed a preference not to disclose his or her information to them. **Methods**: The University of Washington NEXT Medicine Study of the Clinical Sequencing Exploratory Research (CSER) program previously reported participant preferences regarding designating an individual to receive their genomic results in the event of their death, including whether they want the results to be shared, and if so, with what person (PMID:26479557). Participants are also asked whether this designee is the executor of their will or estate. Participants were undergoing genetic testing clinically indicated by colorectal cancer risk. **Results**: Of 59 participants (31 males; 52%), 18 (31%) did not have a will or estate plan. Of those 41 with an estate plan, 28 (68%) had the same designee for their genomic data as their estate executor. Of these 28 participants, 27 designated their spouse/partner and 1 a family member. Thirteen of the 41 (32%) participants with an estate plan had designated someone other than their executor to receive their genomic results; 2 designated their spouse/partner and 11 a family member. Participants who designated a family member rather than a spouse/partner for receipt of their genomic results in the event of their death were more likely to have selected a different individual estate executor (Fisher's exact p <0.001). **Conclusions**: The participant is best positioned to specify whether and with whom they would like their genomic data shared in the case of their death. This work highlights that the individual deemed most appropriate by the participant to receive and disseminate genomic results may not be the same person they have chosen to administer their estate and thus may not have legal access within the current laws. These data were collected from patients for whom a genetic cancer test had been clinically ordered, thus, they support provision of the opportunity to choose such a designee on both research and clinical consent forms for genomic tests.

3050T

Adolescent decision-making regarding secondary findings in whole genome sequencing. *R. Byrne[1,7], R. Hayeems[2,6], M. Kaufman[3], K. Sappleton[4], D. Chitayat[1,5], C. Shuman[1,6,7], N. Monfared[1,6].* 1) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Division of Adolescent Medicine, Department of Paediatrics, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 4) Center for Innovation and Excellence in Child & Family Centered Care, The Hospital for Sick Children, Toronto, ON, Canada; 5) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, ON, Canada; 6) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 7) Molecular Genetics, University of Toronto, Toronto, ON, Canada.

Background: Genomic testing is widely used to diagnose adolescent patients with suspected genetic conditions. Such testing is preceded by an educational component to support informed decision-making. Recent research suggests that adolescents benefit from participation in medical decision-making regarding their health care. However, little is known about adolescents' understanding of the risks and benefits related to secondary findings (SFs) found in genomic sequencing. Identifying and understanding the factors that contribute to adolescents' decision-making about SFs will inform new approaches to education and consent for adolescents undergoing genomic sequencing. Objective: To explore factors that are important to adolescents in their decision-making process related to learning about SFs in genomic research. Study Design: Using semi-structured interviews, we explored the experiences of adolescents, ages 12-19, in their decision to learn about SFs in the context of the whole genome sequencing (WGS) in a research setting. Adolescents were previously offered WGS to determine the molecular etiology of their chronic medical condition. Interview questions were used to assess adolescents' understanding of SFs, awareness of the potential risks and benefits and the factors that adolescents consider when deciding to learn about SFs in genomic research. Results: Eight adolescents, ranging from 12-19 years of age, of whom 7/8 opted to learn about SFs were recruited. Thematic analysis of the transcripts using constant comparative analysis was conducted. To date, three sets of factors have emerged as influential with respect to decision-making about secondary findings: familial factors, including family history and parental discussion; personal health history; and perceived value of knowledge. Conclusion: Preliminary results from the study suggest that adolescents' personal health and family relationships impact their decision to learn about SF. This indicates the importance of exploring these factors with adolescents in the pre-test consent process for genomic sequencing.

3051F

NF1 & me: **Study of the impacts of neurofibromatosis type 1 on the lives of affected adolescents.** *F. Duplain-Laferrière[1], R. Drouin[2], C. Bouffard[1].* 1) Pediatrics, Universite de Sherbrooke, Sherbrooke, Quebec, Canada; 2) Pediatrics, Laval University, Quebec City, Quebec, Canada.

Adolescence is a critical moment wherein the learning of social life, love and sexuality, as well as academic skills largely determine the conditions of life as an adult. **Neurofibromatosis Type 1 (NF1)** is a genetic condition that may entail significant medical, aesthetic, psychosocial and learning-related problems. This period can be more difficult for those having the disease. There is a high risk of stigmatization, rejection and depression. *NF1 & Me* is a research project that focus on the perceptions, experiences and concerns of Canadian adolescents aged 15 to 19 years with a diagnosis of NF1. We want to: 1) determine their level of knowledge and what they would like or need to know regarding the clinical and genetic aspects of NF1, 2) know their concerns about its impact on their social life (school, work, company) and private life (sexuality, love, reproduction) and 3) Appraise the role and importance of intergenerational dialogue in their lives. **Methodology :** Asynchronous online text-based interview, in order to let all the necessary time for the participants to answer, in anonymity, without the consent of their parents. **Results :** Almost all the adolescents are well informed on the origin of NF1 and it's mode of inheritance. However, it is not easy to find resources when they have questions. They all mentioned negatives impacts of NF1 on their lives. The results also reveal the link between the interpersonal and intergenerational relationships with health and well-being. There is connection between the interpretation adolescents make of their parents' attitudes towards NF1 and the perception they have of themselves, their quality of life and their future. Those whose parents avoid talking about NF1 seem to have a more negative perception of NF1 and their prospects of success. NF1 adolescents who openly discuss their disorder with at least one parent, show healthy social interactions and seem to be more resourceful to take control of their lives. **Conclusion :** *NF1 & Me* will generate rapidly transferrable and unique results on the medical and psychosocial aspects of NF1 and the intergenerational and interpersonal communication. It leads to innovative approaches to help adolescents with NF1 and their families and educating them as well as health professionals to optimize their means of intervention and communication. *NF1 & Me* has the potential to be useful in the context of other genetic diseases, handicaps, psychological disorders and social problems.

3052W

Review of return of results policy, as reflected in clinical trial informed consent documents. *A. Franca[1], D.J.H. Mathews[2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Johns Hopkins Berman Institute of Bioethics, Johns Hopkins Unitversity, Baltimore, MD.

The recent launch of the Precision Medicine Initiative, the move toward learning health systems, and the changing role of patients in research necessitates that we reconsider the responsibility of an investigator to communicate research results to participants. This reconsideration has been taking place in earnest for several years, through theoretical and empirical work on the return of results and incidental findings (IFs), resulting in several recommendations and guidances from scholars and organizations. In an effort to further inform this debate, we sought to understand the current landscape of practice regarding the return of results, as assessed through an analysis of informed consent documents. We began with a survey of institutional guidelines and/or policies (G/P) regulating informed consent at the top ten NIH-funded U.S. research institutions (USRI). Of these institutions, 8/10 had a G/P mentioning the return of individual results (IR). About half of these (4/10) were specific to genetic results. Five institutions had a G/P regarding IFs. Of these five, one was specific to imaging results, three were specific to genetic results, and one covered both genetic and overall IFs. We next analyzed 39 publicly available informed consent forms (ICFs) from clinical trials published in the New England Journal of Medicine (NEJM) between January 2014 and March 2016. Of the 39 ICFs, 35 addressed the return of research results. Of these 35 ICFs, 22 stated that they would return IR. Of these 22, 13 included specific exceptions regarding the return of results. A second set of ICFs derived from genetics-focused research projects were also analyzed, and included 10 ICFs; 7 that were procured by emailing corresponding authors of clinical trials performed at USRI and published in NEJM between January 2014 and March 2016 with the word "gene" in the title or abstract, and three from three of the five Centers for Mendelian Genomics. Seven of the 10 ICFs mentioned the return of IR. Of these, 6 stated that IR would be returned, and all 6 included specific exceptions. Our findings indicate that there is a tremendous amount of variation in policies, guidelines and practice with regard to the return of IR. We recommend additional guidance and outreach to provide uniformity among institutions in order to decrease variation and ensure research subjects at minimum understand what results will and will not be returned from any trial in which they are being asked to participate.

3053T

Online education and e-consent for preventive genomic screening: The *GeneScreen* **pilots.** *G. Henderson[1], R.J. Cadigan[1], K. Kuczynski[1], D. Skinner[1], C. Rini[1], R. Butterfield[1], A. Prince[1], M. Van Riper[1], M. Roche[1], K. Foreman[1], J. O'Daniel[1], D. Reuland[1], K. Muessig[2], J. Evans[1], K. Goddard[2].* 1) University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Center for Health Research, Kaiser Permanente NW, Portland, OR.

Programs that screen adults for rare, medically actionable conditions will likely use web-based education and consent. Yet, there is little evidence for the effectiveness of e-consent to convey study features, and harms and benefits of screening in healthy populations. In *GeneScreen*, we investigate issues raised by such screening, focusing on 11 medically actionable conditions. We piloted recruitment at a hospital-based primary care practice (n=425) and a research biobank (n=650). Individuals were mailed a brochure and letter directing them to the *GeneScreen* website. The website provided a summary of 14 study features relevant for informed consent, which could be expanded to display more information. *GeneScreen* "joiners" were surveyed about decision-making and understanding, and a subset were interviewed in more detail. Here we explore predictors of ease of deciding to participate and understanding of the main features of the study.   Of 1076 who were approached, 242 joined *GeneScreen* and completed the survey (23% response rate). 74% were from the biobank. The average age was 59 (range 24-95); and most were female (69%) and non-Hispanic White (83%). Median income was $75,000 - $99,999. Two-thirds had at least a college degree. The majority had good health literacy: only 12% reported difficulty reading medical information. Genetic self-efficacy was high (Mean 1.92 on a scale of 1-6, with 1 showing higher efficacy). Two-thirds of participants (67%) did not expand any of the 14 consent summary items; 11% expanded one, and 22% expanded two or more. On a 6-item scale measuring understanding of the *GeneScreen* study, the mean score was 4.52 (range 0-6). Most (81%) found the decision to join extremely or very easy. Telephone interviews (n=44) found that over two-thirds made the decision to join after reading the printed recruitment materials, before going online.   In multivariable analysis, difficulty in deciding to join was significantly associated with higher education, income, and greater number of consent summary items expanded, and negatively associated with genetic self-efficacy. Those who found the decision easy did not expand the consent summaries, perhaps having decided early in the process. Expanding more consent summaries was not associated with study knowledge; the only significant predictors were income and education. We discuss implications of these findings in the context of e-consent for genetic screening.

3054F

Chartering precision medicine: Addressing health disparities through effective engagement partnerships. *R. Isasi[1,2], M. Pericak-Vance[2,1].* 1) Dpt. Human Genetics, Miller School of Medicine, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics.

Precision Medicine (PM) aims at improving the understanding of complex biological, environmental and lifestyle factors. It promises to revolutionize medicine by improving disease prediction, prevention, diagnosis and treatment. Increasingly, research is moving away from the traditional unidirectional model to one in which patients and other key stakeholders are fully engaged. Such engagement produces research that is more patient-centered, useful, and trustworthy and ultimately leads to greater use and uptake of research results by patient and the broader community. In the wake of PM, the problematic of health disparities has taken renewed importance. Against the backdrop of those purported benefits PM could bring, are concerns surrounding the fact that PM could actually end exacerbating existing disparities in health care, if its benefits are not equitable distributed. Key to the success of PM is to take into account the rich diversity of the USA population to ensure that the benefits of these advances accrue to all. The disproportional representation of ancestral-racial-ethnic minorities and socially disadvantaged populations in clinical translation efforts constitutes an important roadblock for the optimal implementation of PM. For PM to be effective as a meaningful stakeholders' partnership endeavour, novel regulatory, governance and ethical approaches would need to be conceptualized and implemented. Embracing a dynamic, culturally competent, consent model is one mechanism to increase recruitment and build a diverse cohort of well-informed and engaged participants. However, this approach requires refining mechanisms directed at recruiting and engaging participants and raising awareness, particularly with respect to underrepresented communities. Placing participants as co-producers in the planning, design and execution of PM also necessitates a cultural shift and the improvement of educational efforts directed at the general public (e.g. health literacy) and of healthcare professionals (e.g. cultural competence). The task at hand is thus, to identify and anticipate challenges as well as to generate strategies to leverage on the transformative power of PM to address health disparities. To that end, this presentaton will explore the use of novel technologies to increase the participation and engagement of underrepresented populations in PM.

3055W

Consent for newborn screening: Parents' and healthcare professionals' experiences of consent in practice. *S.G. Nicholls[1,2], H. Etchegary[3], L. Tessier[4], C. Simmonds[5], B.K. Potter[2], J.C. Brehaut[6,1], D. Pullman[7], R.Z. Hayeems[8,9], S. Zelenietz[4], M. Lamoureux[4], J. Milburn[4], L. Turner[10], P. Chakraborty[4,11], B.J. Wilson[2].* 1) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 2) School of Epidemiology, Public Health & Preventive Medicine; University of Ottawa, Ottawa, Ontario, Canada; 3) Clinical Epidemiology Unit, Faculty of Medicine, Memorial University, St John's, Newfoundland and Labrador, Canada; 4) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 5) Health Research Unit, Faculty of Medicine, Memorial University, St John's, Newfoundland and Labrador, Canada; 6) Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; 7) Community Health and Humanities, Faculty of Medicine, Memorial University, St John's, Newfoundland and Labrador, Canada; 8) Institute of Health Policy Management and Evaluation, University of Toronto; Toronto, Ontario, Canada; 9) Program in Child Health Evaluative Sciences, Hospital for Sick Children Research Institute, Toronto, Ontario, Canada; 10) Provincial Medical Genetics Program, Eastern Health, St John's, Newfoundland and Labrador, Canada; 11) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Background: Consent processes for newborn bloodspot screening (NBS) are variable, with a lack of descriptive research that depicts how the offer of NBS is made to parents. Methods: We explored the experience, in practice, of consent for NBS. Semistructured interviews in two Canadian provinces were held with: (1) parents of children offered NBS (n=32) including those who declined NBS; and (2) health-care professionals involved in the NBS process (n=19). Data on recollections of NBS, including consent processes, were utilized to identify emerging themes using the method of constant comparison. Results: The meaning of consent, responsibilities, and practical requirements for practice, varied between individuals. Based on reported experiences, three themes were relevant to NBS consent: (1) The 'offer' of NBS; (2) content and timing of information provision; and (3) the importance of parental experiences for consent decisions. Recollections of consent for NBS were similar between jurisdictions, with notable practice variaton between healthcare provider specialities. Excepting midwives and their patients, NBS was viewed as a routine part of giving birth, with little evidence of an informed consent process. Although most parents were satisfied, respondents suggested information about NBS be provided long before the birth. For parents who declined screening, consent was not purely an abstract decision based on benefits and harms, but was also an in-the-moment decision informed by experiences of blood draw procedures. Discussion: These results raise important questions for researchers and those responsible for NBS programs regarding the kind of information-seeking parents engage in before labor. Whose reposnsibility is this: is there any onus on parents to engage in information seeking before the birthing process? Other questions relate to why one group of practitioners view informed consent as less onerous than others, or to study actual practitioner–parent appointments with a view to explore why (or if) information on NBS is not routinely offered in prebirth appointments. These research foci could help inform practical processes for parental information provision and consent. Accounts of parents who declined screening highlight the influence of parental experiences with the heel prick process in screening decisions, and the potential importance for pain management during blood draws.

3056T

Cultural influences on genetic testing in Japanese university students.
H. Numabe[1,2], M. Wakai[2], H. Kawashima[2]. 1) Department of Genetic Councelling, Ochanomizu University, Bunkyo-ku, Tokyo, Japan; 2) Clinical Genetics Centre, Tokyo Medical University Hospital, Tokyo Japan.

In Japan, genetic testing has been carried out using similar methods with accuracy comparable to western countries. However, the implementation of testing is influenced by various factors such as population policy, economic situation, and the medical system of government, as well as the bioethical view of each individual. Especially, culture plays a substantial role in genetic medicine and reproductive medicine. It is thus imperative to take cultural influence into consideration when giving genetic counseling and genetic testing to patients and clients in Japan. We made the questionnaire survey of Japanese university students on genetic diagnoses. The questionnaire consisted of properties of the student, considerations to spiritual matters, acceptability of an artificial abortion, the time of the beginning of human life, and considerations to various genetic testing. Students were from life science courses, nursing course, and medical course of universities in Tokyo. In Japan, acceptability of genetic testing correlates closely with student's recognition of the time of the beginning of human life. Students who recognize the life start from the fertilization or within 15 days after fertilization are not acceptable to prenatal and pre-implantation tests. On the other hand, students who recognize the life start after 22 weeks of gestational age or after birth are acceptable to them. We make an interim report.

3057F

Return of secondary findings and clinical sequencing — Formulating a policy in Japan. *N. Ohashi, K. Kato.* Osaka University, Suita, Japan.

Recently genome sequencing techniques in the medical and research fields have made remarkable progress, and personalized medicine using genome information is expected to become available in the near future in Japan. However, when promoting genomic medicine such as clinical sequencing, various ethical, legal, and social issues arise. One such issue is the handling of secondary findings—usually called incidental findings (IFs)—from clinical sequencing. Although in some of the countries such as United States many genomics researchers as well as ELSI researchers have discussed the issue, there has been little analysis and policy making in Japan. Our group, supported by the new funding agency AMED (the Japan Agency for Medical Research and Development), has been working on the issue. First of all, we have investigated the actual state of advanced medical institutions that will encounter this issue. We have also conducted a survey of policies that have been proposed and implemented abroad. Through our survey and analyses, we have found the following results. □The most advanced activities are found in the field of cancer research and some organizations are returning the results of IFs to patients. □Some organizations have prepared procedures including method of consent, decision making and counseling. They may serve as guiding examples to others in Japan. □Most of the organizations consider the ACMG (the American College of Medical Genetics and Genomics) recommendations and list of genes/mutations useful. On the other hand, there are other issues such as disclosure to family members, lack of databases for interpretation, who takes responsibility for the decision making, etc. We are now preparing a tentative policy with a proposal list of genes and procedures of returning results that we hope are suitable to Japanese clinical settings. In this presentation we will introduce the draft policy and discuss its challenges.

3058W

Motivations and concerns of biobank participants in allowing family access to research samples and data after their death. *J.E. Olson[1], E.M. Winkler[2], J.T. Bublitz[1], M.A. Hathcock[1], J.E. Pacyna[3], J.B. McCormick[1], C. Radecki Breitkopf[1], R.R. Sharp[3].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 3) Biomedical Ethics Program, Mayo Clinic, Rochester, MN.

   Introduction: Biobank data and samples retain their value after participants' deaths. Posthumous access to biobank samples may be valuable to family members of the deceased, but few studies report on participant preference for family access to samples and research results after participants' death. At consent, 2% of Mayo Clinic biobank participants elected to not allow their family to access their biobank samples after their death. **Study Aim**: To understand the motivations, reasons, and concerns of persons who participate in Biobanks but <u>do not</u> want their samples and/or data shared with family members after their death. **Methods**: We conducted a mailed survey of 840 participants in the Mayo Clinic Biobank who had restricted family access to samples (restrictors) after their death and an age- and sex-matched sample of 840 participants who had allowed family access to samples (allowers). Surveys were mailed twice with an interval of 30 days, followed by phone calls to non-responders. Included in the survey were investigator-initiated items reflecting motivations and concerns about family access and published measures of self-concealment, and family environment (cohesiveness, expressiveness and conflict). **Results**: In all, 1164 (69%) responded and had usable data. Stability of family access choice was strong among the original allowers: 95% (553/580) continued to allow access to their sample. In contrast, only 23% (136/584) of original restrictors continued to restrict sample access. Among current allowers, "the potential for my family to benefit" (78.7% reported it was very important) and "the possibility of learning something new about our family history" (56% reported it was very important) were strong motivators. Among current restrictors, privacy of their health information was the most important factor for their choice (39% reported it was very important). Current allowers scored higher on family cohesion than current restrictors (7.8 vs 6.1, respectively; p<0.0001) and lower in self-concealment (15.7 vs 20.6, respectively; p<0.0001). **Conclusions:** Biobank participants hold different preferences regarding family access to their samples, and these preferences may change over time, particularly among those who initially wish to restrict access. At the time of consent, participants may benefit from anticipatory guidance regarding their views on family dynamics and issues related to privacy, health history and benefit within the family.

3059T

The sharing of genomics and health-related data from children: A case of big data and little ethics? *V.N. Rahimzadeh[1], C. Schickhardt[2], J. Friedman[3], M.C. Cornel[4], S. Pfister[5], D. Avard[6], D.F. Vears[7], K. Sénécal[8], W. Pinxten[9], C. Fernandez[10], S. Plon[11], A. Thorogood[12], B.M. Knoppers[13], E. Winkler[2].* 1) Centre of Genomics and Policy, McGill University, Montreal, QC, Canada; 2) University Hospital Heidelberg, Heidelberg (Germany); 3) Co-Lead GA4GH Paediatric Task Team, University of British Columbia (Canada); 4) Co-Lead GA4GH Paediatric Task Team, VU University Medical Center, Amsterdam (The Netherlands); 5) Professor and Head of Division Pediatric Neurooncology at the German Cancer Research Center Heidelberg; 6) Centre of Genomics and Policy, McGill University (Canada); 7) Centre of Biomedial Ethics and Law, KU University, Leuven (Belgium); 8) Coordinator, Paediatric Task Team GA4GH, Centre of Genomics and Policy, McGill University (Canada); 9) University UHasselt (Belgium); 10) IWK Health Centre, Dalhousie University (Canada); 11) Texas Children's Cancer Genetics Clinic (Clinic), Baylor College of Medicine,; 12) Coordinator Regulatory and Ethics Working Group (GA4GH); 13) Co-Lead, Regulatory and Ethics Working Group (GA4GH), Director, Centre of Genomics and Policy, McGill University (Canada).

   With the sequencing of the human genome, biomedical research has increasingly become a data-intensive venture that requires collaboration between researchers and institutions. Understanding how data sharing practices and policies (dis)enable scientific collaboration in genomics and related –omics disciplines is therefore critical to accelerating clinical innovation in Canada. Genomic research involving children can further accentuate this need, as their inability to consent underpins their status as a vulnerable population in research but nevertheless warrants their inclusion in research. Sharing research data fulfills dual imperatives inherent to any bench-to-bedside translational endeavor: a scientific imperative arises in the need to determine underlying genetic determinants of disease, and an ethical imperative arises to deliver on the anticipated benefits upon which the risks to research participants rest. For research ethics review committees (REC) and data access committees (DAC), genomics raises chiefly informational rather than physical risks. These risks require that RECs and DACs ensure data-specific, albeit proportional protections for children. Concerns regarding data privacy and security can be accentuated when children are involved [3], particularly when use of data for future, as-yet unspecified research is the norm rather than the exception in the data-intensive sciences. Collaboration among researchers who share data has nevertheless led to significant advancements in understanding childhood diseases; most notable of these include in pediatric oncology, rare genetic disease and autism. Children enjoy a right to informational privacy despite their inability to consent alone to participate in data-intensive science research. The data to be shared may also be of immediate or future relevance to the child's health, such as genetic data, and which may take on new significance as data-intensive science unearths new understanding(s) between genomics and human disease etiologies. This research investigates specific regulatory requirements for pediatric data sharing in relation to data protection, data usages and the ethical justifications for doing so in the data-intensive science research disciplines. Based on this scoping regulatory assessment, we discuss the ethical implications of children as data donors and active agents in the advancement of pediatric research in the data-intensive sciences.

3060F

Protection of genomic information by amended Japanese Personal Information Protection Laws. *N. Yamamoto, K. Kato.* Osaka University, Suita, Osaka, Japan.

In Japan, genomic information has now been put under legal protection for the first time ever because of the amendment to the Personal Information Protection law last year. There are mainly three Information/data protection laws in Japan: the "Act on the Protection of Personal Information (PPI Act)", the "Act on the Protection of Personal Information Held by Administrative Organs (PPIHAO Act)", and the "Act on the Protection of Personal Information Held by Incorporated Administrative Agencies (PPIHIAA Act)". The PPI Act includes the basic principle for all of the Acts, so an amendment of the PPI Act has started in 2013 prior to the other two, and enacted in September 2015. The PPIHAO and PPIHIAA Acts are now under review. In the amended PPI Act, genomic information was included in the definition of personal information. This is a welcome change because until now patients, research participants and consumers of genetic testing have been exposed to the risk of an invitation of privacy by breach. In addition, the amendment has raised concerns about genomic privacy, security and genomic industries among people at a deep level that has not been witnessed before in Japan. This offers an important opportunity to develop social trust and understanding between researchers and the public regarding how genomic information should be used in the future. However, some problems must be addressed as soon as possible. For example, it is an urgent matter to harmonize some particular aspects of the amended PPI Act and conventional research ethics such as how to handle informed consent when a subject's data is to be used in further research. We must ensure that patients' privacy is protected while at the same time encouraging scientific research. We have been involved in various activities to deal with these urgent matters. In this presentation, we will introduce our Genomics and Society Unit that was established in April 2016 to strengthen genomic ELSI as well as some practical solutions (e.g., a new informed-consent form) that we have proposed in every-day genomic research, including data sharing through public databases.

3061W

Personalized genomics clinics: A model for delivery of genetics care. *M. Sabbadini, J.T. Shieh, A.M. Slavotinek.* Department of Pediatrics, University of California, San Francisco, San Francisco, CA.

By targeting numerous genes simultaneously, whole exome sequencing (WES) has had a transformative impact on patient care in the few years since its inception as a clinical test. Its ability to identify a diagnosis in approximately 25% of patients with a suspected, but previously unidentified, genetic condition is revolutionary. The availability of WES and its use for screening without an identified genetic diagnosis has also increased the accessibility of genetic testing to providers who are not trained in this specialty. However, the nuances of the consent process, the complexity of test ordering, the difficulty in accurate variant interpretation, and the need for re-analysis and longitudinal follow-up for patients with variants of unknown significance (VUSs) or negative exome results make the presence of a clinician trained in genetics an invaluable asset towards realizing the full potential of genomic testing. The optimal clinical model for the incorporation of WES into patient care is not yet known. The Individualized Medicine clinic at the Mayo Clinic and Sick Kids Genome Clinic in Toronto are two published models of genomics clinics centered on the provision of genomic testing and both of them have a strong emphasis on research and data collection. In 2013, at UCSF, we established a personalized genomics clinic (PGC) for the consent process and testing, genetic counseling, results interpretation and long-term follow up for patients who undergo WES. Patients are referred to our clinic by UCSF genetics providers as well as by providers from other specialties from within our institution and externally. Patients can be referred for evaluation, consent/testing and for results provision/interpretation. Most commonly, cases were referred to PGC at the stage of results interpretation or for re-analysis following previous WES with VUSs or no variant reported. The clinic also includes a registry function whereby patients with VUSs can be contacted for annual results review. We present our clinic structure as a model for accurate and comprehensive patient-centered clinical implementation of genomic testing. Our clinic model provides support for the intrinsic advantages of establishing dedicated clinics for WES. Moreover, it delineates the value of genetics provider involvement in the genomic testing process with the goal of taking advantage of the full potential of WES while limiting the possible risks of such test.

3062T

Attitudes towards centralized biorepositories among patients in Cleveland, OH: Implications for the Precision Medicine Initiative Cohort Program. *J.N. Cooke Bailey[1,2], W.S. Bush[1,2], A. Slaven[3], M. Schachere[3], D.C. Crawford[1,2], J.R. Sedor[3,4,5], J.F. O'Toole[3].* 1) Epidemiology and Biostatistics , Case Western Reserve University, Cleveland , OH; 2) Institute for Computational Biology, Case Western Reserve University, Cleveland , OH; 3) Division of Nephrology, Department of Medicine, MetroHealth Medical Center, Cleveland, OH; 4) Department of Physiology and Biophysics, Case Western Reserve University, Cleveland , OH; 5) Department of Medicine, Case Western Reserve University, Cleveland , OH.

   The Precision Medicine Initiative, announced by the White House in early 2015, promises to revolutionize patient care by incorporating clinical, environmental, personal, and 'omics data in electronic health records (EHRs) for decision support at the point of clinical care. To accelerate this vision, the National of Institutes of Health in consultation with academic, healthcare, mHealth, and patient advocacy leaders, is launching the PMI Cohort Program (PMI-CP), an ambitious national effort to ascertain one million Americans for precision medicine research. A requirement of this cohort is the submission of electronic health records and biospecimens to a central biorepository facility funded by the government for storage and further processing for data generation and analysis. While this centralized biorepository model provides quality control and is cost-effective compared with a federated model, it is unclear what impact this requirement will have on ascertainment, which is expected to be socioeconomically, geographically, and racially diverse. To understand the potential impact this model has on ascertainment, we are surveying patients participating in a precision medicine research project at MetroHealth in collaboration with the Institute for Computational Biology at Case Western Reserve University in Cleveland, Ohio, a diverse metropolitan area with 20% of residents self-described as African American in the 2010 U.S. Census. MetroHealth is an academic, public provider, integrated tertiary care system serving northeast Ohio with a vendor-based EHR. The payer mix includes 21% uninsured, 24% Medicaid, 18% Medicare. Results currently include responses from 37 patients; 72% were African American, 55% were female, and average age was 61. Regarding willingness to participate in the PMI, 61% of patients responded favorably to submitting health records and genetic data to a national biorepository coordinated by the government. Only ~39% of patients indicated that return of results was very important whereas ~23% did not want results at all. When asked if they would be willing to install an application on their mobile device to track their health and to send that data to a national center, ~52% responded favorably to both. These responses, while preliminary, indicate that the attitudes of patients in a diverse health care environment towards the PMI-CP are varied.

3063F

A growing market for clinical exome sequencing. *G.W. Hooker[1], T.A. Murphy[1], K.E. Woodall[2], J.D. Schneider[1].* 1) NextGxDx, Franklin, TN; 2) University of Tennessee, Knoxville, TN.

   Introduction: Innovations in DNA sequencing are driving rapid growth and change in the genetic testing marketplace. As the cost to sequence a single base pair goes down, many laboratories have been able to enter new, competitive sectors of the market. One particularly competitive sector of the market is clinical exome sequencing. These tests command some of the highest per-test costs in the market, likely due to the greater analytic infrastructure needed to return high-quality reports. A more recent development in the market has been increasing willingness on the part of health insurance plans to pay for exome sequencing. In this analysis, we aimed to describe growth in the market and in the exome sector specifically, from Sep. 1st, 2014 to June 1st, 2016. We also studied pricing of clinical exome tests, to understand pricing changes in light of market growth and evolutions in reimbursement policy. **Methods:** Drawing from online lab catalogs, we curate and maintain a data resource of genetic tests offered by US CLIA-certified clinical labs. We collect information on all tests offered by these labs, as well as other test attributes such as methodology and test prices. We tracked new tests entering and exiting the market on a month-to-month basis and compared average costs per test from Sep. 2014 to June 2016. **Results:** Over this time period, we saw a net of 6,583 new tests enter the market (9.5 new tests/day). Among exome tests, we saw a net of 22 new products enter the market. The number of labs offering exome sequencing increased from 16 in Sep. 2014 to 26 in June 2016. Overall, there was a 20% decrease in the price per clinical exome from $7085/exome to $5700. Notably, however, since July 2015, the price per clinical exome has risen by 8% from a low of $5265, driven mainly by new tests entering the market at higher prices. Also during this time, coverage for exome sequencing has expanded, with 3 of the top 10 commercial health insurance plans covering exomes for eligible patients in 2015. **Conclusions:** Overall, costs in the exome sequencing sector decreased while the supply of tests on the market increased. However, this trend partially regressed in the presence of increasing private insurance coverage for these tests. The dynamic nature of the market suggests payers and clinicians can achieve value through contract negotiation, lab selection and utilization management. Ultimately, this has the potential to increase patient access to genetic tests.

3064W

Genetic counselors' experiences with clinical whole genome sequencing in a diverse healthy population. *M. McGinniss[1], E. Ramos[1], E. Thorpe[1], L. Fosler[1], K. Schahl[2], K. Trzupek[2], A. Hata[2], T. See[2], D. Eklund[3], H. Bixenman[3], D. Morton[3], D. Wellis[3].* 1) Illumina, Inc, San Diego, CA; 2) InformedDNA, St. Petersburg, FL; 3) San Diego Blood Bank, San Diego, CA.

Clinical whole genome sequencing (cWGS) for healthy individuals is rapidly gaining acceptance as cost is decreasing and success rates in identifying disease-causing variants are increasing. Studies are assessing the impact of cWGS, but the majority of early adopters have a narrow profile of racial diversity (mostly Caucasian). This lack of diversity is a barrier to understanding how genetic counseling for cWGS can be optimized for a population as diverse as that of the United States. The growing availability and uptake of cWGS highlights the importance of re-evaluating current genetic counseling methods to provide innovative approaches to address the increasing demand of genetic counseling services. We performed an IRB-approved pilot project between the San Diego Blood Bank (SDBB) and Illumina, where cWGS was offered to 70 SDBB blood donors whose demographics represented the diverse population of San Diego. The SDBB cohort was 27% Caucasian, non-Hispanic (19/70), 26% Hispanic/Latino (18/70), 13% Black/African-American (9/70), 11% Asian (8/70), 7% American Indian/Alaskan Native (5/70), and 4% Native Hawaiian or other Pacific Islander (3/70). The remaining 11.5% of participants (8/70) reported two or more races. Certified genetic counselors at InformedDNA conducted post-test telephone genetic counseling sessions, which included family and medical history collection, results disclosure, follow-up recommendations, and psychosocial assessment. Counseling metrics, including preparation time, counseling time, and time spent on documentation were tracked and compared to the 2016 National Society of Genetic Counselors (NSGC) Professional Status Survey (PSS). Average time spent counseling participants was 66 minutes with 39% of counseling sessions taking 45-60 minutes. Average time spent preparing for patients and completing a consultation summary was 18 and 35 minutes. These patient care metrics are comparable with what has been reported for in the 2016 NSGC PSS (41.8% of initial counseling sessions taking 45-60 minutes; 40.2% of preparation time taking 15-30 minutes; 31.7% of follow-up taking 15-30 minutes). This pilot study demonstrates the feasibility of providing consistent genetic counseling services to a diverse healthy cohort with cWGS results. The ability to integrate telephone genetic counseling services into cWGS also supports increasing diversity in study populations as it decreases issues of availability of and access to local genetic counselors.

3065T

Do primary care physicians manage genome sequencing results appropriately? Results of an expert panel evaluation. *E.F. Schonman[1], M.F. Murray[3], J.B. Krier[1,2], C.L. Blout[1], D.W. Bates[1,2], R.C. Green[1,2,4,5], J.L. Vassy[1,2,6], The MedSeq Project.* 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Geisinger Health System, Danville, PA; 4) Partners Healthcare Personalized Medicine, Cambridge, MA; 5) Broad Institute of MIT and Harvard, Cambridge, MA; 6) Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA.

Background: As genome sequencing (GS) becomes increasingly available, there are concerns that primary care physicians (PCPs) may mismanage patients' variants, particularly since insufficient empiric evidence exists to develop practice guidelines for most genes and phenotypes. We used quantitative expert review to rate PCP management of real GS variants. **Methods:** The MedSeq Project is a trial of GS among healthy adult patients. Patients and their PCPs received a GS report, including monogenic disease variants classified as pathogenic, likely pathogenic, or variant of uncertain significance (VUS): favor pathogenic. For each patient with a monogenic variant, we created a clinical vignette summarizing his/her health history, variant pathogenicity, and the PCP's clinical management of the result. We recruited a panel of 11 clinical genetics experts to rate the PCP's management in each vignette, using the 9-point RAND/UCLA Appropriateness Scale (1-3: inappropriate; 4-6: neither appropriate nor inappropriate; 7-9: appropriate). In round 1, panelists independently rated PCPs' actions and overall management. In round 2, panelists convened to discuss actions for which a quantitative disagreement index indicated disagreement about how to rate appropriateness. After discussion, panelists could change their initial ratings. After each round, panelists gave recommendations for general guidelines for variant management by PCPs. **Results:** Of 50 GS patients, 11 received a new monogenic result. In round 1, panelists agreed how to rate the overall management of all 11 vignettes: 8 were rated as managed appropriately, 2 inappropriately (pathogenic variants in *LXH4* and *F5*), and 1 neither appropriately nor inappropriately (VUS in *ARSE*). Panelists disagreed how to rate the appropriateness of 3 of 26 specific actions from 3 different vignettes, but in round 2, they reached consensus on 2 of these actions (stating no evidence of porphyria and chondrodyslplasia punctata in patients with *PPOX* and *ARSE* variants, respectively). They remained in disagreement about a PCP's explanation of inheritance of an *HFE* variant. Panel discussion resulted in a preliminary 6-item guideline for PCPs managing GS variants. **Conclusion:** According to expert geneticist review, PCPs managed GS variants appropriately in the majority of cases in our study. Analysis of case mismanagement, including misinformation and errors of omission, may help shape future genomic medicine practice guidelines for PCPs.

3066F

Recommendations of researchers and physicians about inherited chromosomally integrated human herpesvirus 6. *V. Noël[1], R. Drouin[2], L. Flamand[3], C. Bouffard[1].* 1) Université de Sherbrooke, Sherbrooke, QC, Canada; 2) Université du Québec à Montréal, Québec, QC, Canada; 3) Université Laval, Québec, QC, Canada.

The human herpesvirus 6 (HHV-6) has the unique capacity to integrate telomeres of the chromosomes (ciHHV-6) of nucleated cells after infection. ciHHV-6 can therefore be found in gametes and be transmitted from one generation to the next and through organ and tissue donation. This worrisome situation affects between 40 and 70 million people around the world and may be associated with illnesses (encephalitis, infertility, cancers, etc.), of which some are fatal. Despite its consequences, this phenomenon has attracted little interest in scientific and public health milieus. To highlight the importance of this issue, several researchers and physicians go so far as to publish warnings in basic research articles in purely scientific journals. This uncommon phenomenon, prompted us to investigate these concerns. Thus, our objective was to know the solutions they propose in order to identify their clinical and socio-ethical concerns. METHOD: Qualitative design used asynchronous online text-based interviews (combine questionnaire and chat box) with 13 researchers and physicians working on HHV-6. Qualitative generale inductive analysis with the use of NVivo11 software. RESULTS: Researchers and physicians suggested four types of interventions: 1) the establishment of a screening program in the context of grafts and transplants; 2) medical information on the integrative capacity of HHV-6 and pathogenicity of ciHHV-6; 3) reevaluation of types of antiviral drugs and doses to counteract the adverse effects of pharmacological treatments and standardize the therapeutic treatment plan; 4) standardization of diagnostic tests to prevent false positives or negatives. CONCLUSION: It was very interesting to note the convergence of opinions and the recommendations of the participants. These first data on the topic respond to needs expressed by researchers and physicians concerned by the ciHHV-6 in their practices. In this sense, the results of this research can guide the thinking and decision-making process to improve and harmonize practices and the wellbeing of affected individuals and populations. .

3067W

The NIGMS Human Genetic Cell Repository: An invaluable resource and a research participation opportunity for individuals with inherited diseases and chromosomal abnormalities. *S. Wert, M. Berkowitz, L. Brown, T. Schmidlen, D. Requesens, N. Turan.* Coriell Institute for Medical Research, Camden, NJ.

The NIGMS Human Genetic Cell Repository was established at the Coriell Institute for Medical Research, an independent non-profit research institution, in 1972 by the National Institute of General Medical Sciences (NIGMS). The NIGMS Repository provides a readily accessible, centralized resource for cell lines and genetic material from individuals with genetic disorders, including heritable diseases and chromosomal abnormalities, and from individuals from various geographical locations for studies on human diversity. Samples are also collected from unaffected, first-degree relatives of individuals with genetic disease, and from apparently healthy individuals for the study of human variation. The primary goal of the NIGMS Repository is to stimulate and facilitate biomedical research, teaching, and the diagnosis and prevention of human genetic diseases by establishing and maintaining a repository of high quality, uncontaminated, well characterized, and clinically well-documented cell lines and DNA and making these resources available to investigators throughout the U.S. and abroad. The NIGMS Repository contains more than 11,300 cell lines, including human iPSC lines, and 5,600 DNA samples, representing over 900 OMIM phenotypes. The NIGMS Repository is an invaluable resource and samples have been used for genetic testing assay development, validation and proficiency, identification of novel genes, exploration of gene function and expression, and generation iPSC lines. To date, there have been over 6,000 publications citing NIGMS Repository samples. The NIGMS Repository is also an important resource for clinicians and genetic counselors as it offers an opportunity for patients and families to contribute to the research process in addition to, or in the absence of, the availability of clinical testing or disease-specific research. Sample donation requires a blood or tissue specimen and completion of an informed consent, a submission form, and a clinical data summary.

3068T

Benefits, risks, and perceived utility of newborn genomic sequencing: Comparing parents' and physicians' baseline attitudes in the Baby-Seq Project. *H.A. Peoples[1], A.M. Gutierrez[1], S. Pereira[1], J.O. Robinson[1], L. Frankel[1,2], M. Towne[3], M. Helm[5], C.A. Genetti[3], I.A. Holm[3,4], A.H. Beggs[3,4], R.C. Green[4,5,6,7], A.L. McGuire[1].* 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine , Houston , TX; 2) Department of Psychological, Health, and Learning Sciences, University of Houston, Houston, TX; 3) Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 6) Partners Healthcare Personalized Medicine, Cambridge, MA; 7) Broad Institute of MIT and Harvard, Cambridge, MA.

   Background: Efforts are ongoing to explore the impact of integrating genomic sequencing (GS) into newborn clinical care. Concerns have been raised about the risks of providing incidental GS results to parents of newborns yet parents' perceptions of this technology compared to newborn screening (NBS) are unclear. We examined attitudes toward NBS and GS among newborns' parents and physicians. **Methods**: The BabySeq Project is a randomized controlled trial exploring the utility of GS for parents of newborns in the NICU and Well-Baby nursery. Parents of enrolled infants complete surveys at 4 time points in their baby's first year of life. Physicians are surveyed throughout the study. Survey domains for parents and physicians include attitudes toward and perceived utility of NBS and GS. Chi square tests were used to measure between-group differences in perceived utility and attitudes toward NBS and GS. Open-ended responses were categorized by theme. **Results**: The majority of parents (81%, n=159) and physicians (97%, n=67) agreed that there are health benefits associated with NBS. Fewer parents (18%) than physicians (33%) agreed that there are risks associated with NBS (p=.022). A smaller majority of parents (67%) and physicians (57%) also agreed that there are health benefits associated with GS. More physicians (73%) than parents (35%) agreed that there are risks associated with GS (p<.001). In open-ended responses, physicians listed early intervention as the main benefit of GS while parents also listed family planning and knowing what to expect in the future. Both parents and physicians identified psychological distress, genetic discrimination, impact on family, and the nature and potential uncertainty of GS results as risks. Parents felt that receiving their child's GS results will be more important in 10 years (82%) compared to now (61%). Physicians reported that GS will be more useful in 10 years compared to now for diagnosis (90% vs. 64%) and for managing their patients' care (85% vs. 58%). **Conclusions**: At baseline, parents reported fewer perceived risks of NBS and a more favorable risk/benefit tradeoff for GS than physicians. Parents also identified additional expected benefits of GS. Results suggest that parents may have a broader conception of the utility of GS than physicians. Continued data collection will determine whether these attitudes persist after parents receive their baby's GS results.

3069F

Genomic newborn screening: Public health policy considerations and recommendations. *K. Sénécal[1], J.M. Friedman[2,3], M.C. Cornel[4,5], A.J. Goldenberg[6], K.J. Lister[7], D.F. Vears[8], Global Alliance for Genomics and Health Regulatory and Ethics Working Group Paediatric Task Team.* 1) Centre of Genomics and Policy, Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 3) Child and Family Research Institute, Vancouver, Canada; 4) Section Clinical Genetics, Department of Clinical Genetics, Amsterdam, Holland; 5) EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, Holland; 6) The Center for Genetics Research Ethics and Law, Department of Bioethics, Case Western Reserve University, Cleveland, Ohio, USA; 7) Office of Population Health Genomics, Public Health Division, Department of Health, Government of Western Australia; 8) Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium.

   Recent research has demonstrated that genomic technologies, such as exome or whole genome sequencing, can identify genetic causes of rare paediatric diseases more effectively than conventional clinical and laboratory methods. However, the nature of this technology and its potential to identify additional information about the child being tested raise a range of ethical concerns, particularly when these technologies are applied in a public health setting, such as newborn screening (NBS). The Paediatric Task Team of the Global Alliance for Genomics and Health's Regulatory and Ethics Working Group considered these issues, developed a framework addressing key issues, reviewed the literature against this framework and then refined their findings in an iterative manner. This led to the development of recommendations to guide clinical laboratory scientists and policy makers regarding the use of genomic technologies for population-based NBS. We recommend that NBS by genomic methods should only be considered as an add-on to current screening programs, and should *not* replace current techniques unless equal or better sensitivity and specificity has been shown. Genomic NBS that has appropriate clinical utility should be equally available and accessible to every infant in the jurisdiction. Sharing of data on pathogenic and benign variants within the population(s) screened is needed for interpretation of variants. Publicly-funded universal newborn screening by genomic methods should be limited to diseases that can be effectively treated or prevented early in life. The program should guarantee treatment and follow-up of infants found to be affected by the targeted diseases.

3070W

Preemptive clinical pharmacogenetic testing in children: Patient and parent preferences. *S.K. Ring[1], C.E. Haidar[1], N. Kornegay[1], J.M. Hoffman[1], U. Broeckel[2], W.E. Evans[1], K.E. Nichols[3], M.V. Relling[1], K.R. Crews[1].* 1) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN; 2) Department of Pediatrics, Medical College of Wisconsin, Milwaukee WI; 3) Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN.

Consent processes for research protocols that include genetic testing ideally address return of primary results as well as secondary or incidental findings. The PG4KDS study, which clinically implements preemptive pharmacogenetic testing in children, genotypes pharmacogenes in a CLIA-certified laboratory. PG4KDS provides a process for implementing gene/drug pairs into the electronic health record (EHR), and for returning test results to patients. We describe parent and patient preferences regarding participation and return of results. The informed consent process provides education (through verbal and written materials) on differences between pharmacogenes and disease-risk genes and on reporting test results. Procedures are defined for sharing primary pharmacogenetic results and incidental findings that may link pharmacogenes to future disease risk. Pharmacogenetic genotype results are placed in the EHR and accompanied by interpretations and clinical decision support for gene/drug pairs. Secondary or incidental findings considered for return to participants are those that carry a higher risk than baseline, are clinically actionable, and contribute information on risk that is beyond what would be found in the course of normal health care. Thus, parents may not opt out of receiving incidental findings for minors. Patients over age 18 may select whether or not to receive incidental findings related to disease risk. Patients and families may select whether they want to be contacted when a gene test result is added to the EHR; notification is through a letter describing the resulting phenotype and an interpretation. Of 3333 patients approached, 3209 (96%) have enrolled. Reasons for refusing participation included privacy concerns (n=12), enrollment in too many studies (n=12), concern over extra blood collection (n=11), family members not agreeing (n=11), distrust of DNA research (n=5), progressive disease (n=4), religious beliefs (n=1), other reasons (n=2) or no reason given (n=66). To date, 3121 participants (97%) have elected to receive their gene test results; 13,768 letters have been mailed and placed in the EHR. Of 224 participants who reached age 18 years, 223 (99.6%) agreed to continue participation as adults; 221 (99%) have elected to be notified of disease risk. These findings suggest a high rate of willingness to be informed of preemptive genotype results and incidental disease risk findings, and a low rate of concerns that prevent participation.

3071T

Sharing genetic information with family: Responses before and after receiving pharmacogenomics results. *M. Smith[1], S.A. Aufox[1], A. Espinoza[1], L.J. Rasmussen-Torvik[2].* 1) Ctr Gen Med, Northwestern Univ, Chicago, IL; 2) Department of Preventive Medicine, Northwestern University, Chicago, IL.

Pharmacogenomic (PGx) variants have been well established as an important cause of variable drug response. However, implementation of PGx testing in clinical care has been slow. The goal of this study was to begin to develop strategies for the optimal implementation of PGx data into the clinical environment. We assessed the utility of PGx information to patients and whether they intended to share results with family. Methods: Patients were offered PGx testing through their primary care physician's (PCP) office. Testing included clinical grade results on 3 drug-gene pairs: Simvastatin, Clopidogrel, and Warfarin. Patients were selected based on likelihood of being prescribed one of the three drugs in a 3-year period. Participants completed three surveys: A 37 item survey at the time of consent and a 31 item survey that was sent to participants at 1 and 6 months post return of results. The surveys included questions about demographics, plans to use and share PGx results (pre) and how results were used and shared (post). We report here the results of participants' intent to share PGx results with family and their reports of sharing results with family at 6 months post-result return. Results: 750 participants were enrolled, received results on all three PGx drug-gene pairs, and completed at least the baseline survey. 352 completed the 1-month follow up survey and 279 completed the 6-month survey. Participants were 57% male; 48% White, 25% African American, and 6% Hispanic; 75% reported they had children and 60% had a college or post graduate degree. At baseline, 39% of participants planned to share results with family. At six months, 33% of participants reported they had told family about their results. Participants with children were more likely to have shared their results with family (37%) vs participants without children (20%) (p=.01). Thirty-five percent of non-Hispanics had shared results by 6 months versus 6% of Hispanics (p=.02). Having a genetic result for increased risk for variable drug response was not correlated with having shared results with family. Intention to share PGx test results with family at baseline was not highly predictive of having shared those results at 6 months. Summary: Being of non-Hispanic ethnicity and having children were two factors correlated with sharing PGx results.

3072F

Patient-centered deep phenotyping using the Human Phenotype Ontology. *M. Haendel[1], N. Vasilevsky[1], M. Engelstad[1], E. Foster[1], C. Mungall[2], S. Köhler[3], P. Robinson[3].* 1) Oregon Health and Sciences University, Portland, OR, USA; 2) Lawrence Berkeley National Laboratory, Berkeley, CA, USA; 3) Charite – Universitätsmedizin Berlin, Germany.

In rare or undiagnosed diseases, physicians rely upon genotype and phenotype information in order to compare abnormalities to other known cases and to inform diagnoses. To maximize the usefulness of accurate phenotyping for clinical diagnosis and to build cohorts of patients for gene discovery, a standard vocabulary is essential. Towards this end, the Human Phenotype Ontology (HPO) (http://www.human-phenotype-ontology.org/) was developed as a standardized vocabulary of phenotypic abnormalities encountered in human disease to facilitate "deep phenotyping", whereby symptoms and characteristic phenotypic findings (a phenotypic profile) are captured using a logically constructed hierarchy of over 12,000 terms. The HPO has been utilized to great success for assisting computational phenotype comparison against known diseases, other patients, and even model organisms to support diagnostics. Although the cost and ease of collecting and analyzing genomic data has improved rapidly, collecting structured phenotypic data has not become more standardized, convenient, or less expensive, limiting algorithmic approaches. Thus a major challenge in clinical care and research aimed at understanding genetic diseases is phenotyping patients accurately, yet efficiently. Conversely, patients are often the best sources of information about their symptoms and phenotypes and there is an opportunity to have them perform self-phenotyping using the HPO. However, the labels and synonyms in the HPO primarily use medical terminology, which can be difficult for patients and their families to understand. In order to make the HPO more accessible to non-medical experts, we systematically added new synonyms using non-expert terminology (i.e., layperson terms) to the existing HPO classes or tagged existing synonyms as layperson. As a result, the HPO contains over 6,000 layperson terms. This patient-centric version of the HPO is being utilized in patient-centered rare disease websites, in patient community platforms and registries, to validate patient phenotyping clinical instruments, and to enhance the deep phenotyping being performed in the clinic.

3073W

Between open and controlled? A registered access model. *S.O.M. Dyke[1], E. Kirby[2], M. Shabani[3], A. Thorogood[1], M. Linden[4], I. Lappalainen[4], J. Rambla De Argila[5], M. Fiume[6], S. Sherry[7], K. Kato[8], B.M. Knoppers[1].* 1) Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 2) Public Population Project in Genomics and Society, Montreal, Quebec, Canada; 3) University of Leuven, Leuven, Belgium; 4) ELIXIR, Hinxton, UK; 5) Centre for Genomic Regulation, Barcelona, Spain; 6) DNAstack, Toronto, Ontario, Canada; 7) National Centre for Biotechnology Information, US National Library of Medicine, Bethesda, MD, USA; 8) Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Osaka, Japan.

As data sharing policies in genomics strive to keep pace with the state of the science, we currently face a situation in which only two established mechanisms exist for data to be consistently shared: *open access,* when data are freely published on the World Wide Web; and *controlled access,* where qualified researchers apply for access on a project-by-project basis and their research plans are reviewed, often by a committee. While both models have served the research community and scientific progress well, plans for greater integration of datasets and informatics platforms call for a new, intermediate model of data access that would enable rapid access for a wider range of users to data presenting a level of ethical and legal risk. In order to expand data access options in this way, we conducted an ethical-legal analysis of research ethics and other legal and administrative frameworks applicable to data sharing and access. In particular, we determined which elements of controlled access mechanisms should be retained to provide proportionate protections involving both user identification and agreement to a general set of responsibilities while considerably simplifying the data access application process. The novel data access model we propose is called *registered access* and comprises a three-stage registration process: Authentication, Attestation and Authorization. It is based on the notion that broad categories of users can be granted access to data according to their roles (e.g., *bona fide* researcher or clinical care professional). All three stages of registered access pose ethical and legal questions which we have addressed with guidance. As we build automated systems to pilot the registered access model with ELIXIR - Europe's infrastructure for life science information - and the Scientific Demonstration projects of the Global Alliance for Genomics and Health (The Beacon Project, The Matchmaker Exchange and The BRCA Challenge), we are confronting and resolving a number of ethical and legal challenges raised by access to existing technological and administrative resources (e.g., academic infrastructure). We also report on current efforts to support the registration of a large and diverse community of clinical care professionals.

3074T

Structured care for individuals at risk for familial cancer syndromes in a family medicine residency. *W. Feero[1], E. Edelman[2], K. Reed[2], T. Ingram[2], N. Fischer[1], K. Lafferty[3], S. Miesfeldt[3].* 1) Maine-Dartmouth Family Medicine Residency, Fairfield, ME; 2) The Jackson Laboratory, Genomic Education, Bar Harbor ME; 3) Maine Medical Center Research Institute, Scarborough, ME.

Background: The patient centered medical home (PCMH) is one model for providing coordinated, longitudinal care for chronic conditions in primary care. Whether the PCMH can be adapted to care for individuals at-risk for familial cancer syndromes in primary care is unknown. Objective: To establish and evaluate clinical and educational infrastructure to support self-sustaining nursing "Cancer FHx Wellness Visits" as a PCMH-based care model for those at risk for Lynch and hereditary breast and ovarian cancer syndromes in a single large (10+10+10) family medicine residency in rural Maine. Methods: Model included: i) tailored education for common familial cancer syndromes (residents [4 hrs], faculty [4 hrs], and nursing [2 hrs]); ii) clinic-based materials to facilitate patient engagement, collection/interpretation of cancer-related family history data and appropriate referral for cancer genetic counseling services; and iii) establishment of workflows/billing processes for program sustainability. Model evaluation included i) assessment of provider education/perception of care coordination through pre-/post-surveys, ii) assessment of patient perception of care coordination through post-genetic counseling survey and iii) utilization of services. Results: Project began in the 2015-2016 academic year. Post-education, physician groups showed statistically significant improved confidence in: eliciting a family history; assessment of hereditary cancer risk; making appropriate referrals; discussing risks and benefits of cancer genetic testing; and providing risk reduction counseling. Providers experienced significant increases in genetic knowledge self-efficacy; however, content expertise as measured by knowledge questions showed little change. Providers reported improved knowledge of where to refer patients for genetic services. Preliminary analysis of provider perceptions of care coordination at 6 months showed statistically higher scores from baseline for faculty but not residents. Conclusions: A PCMH-based model of "Cancer FHx Wellness Visits" and targeted educational interventions improved provider confidence in identifying and managing individuals at-risk for familial cancer syndromes. Additional data are needed to determine if the program is sustainable and improves patients' perceptions of care.

3075F

GenomeConnect: An update of engaging patients in data sharing efforts through an online patient registry. *J. Koenig[1], D.R. Azzariti[2], D.H. Ledbetter[1], C. Lese Martin[1], V. Rangel Miller[3], H. Rehm[2,4,5,6], E. Rooney Riggs[1], K. Wain[1], W.A. Faucett[1].* 1) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Laboratory for Molecular Medicine, Partners Personalized Medicine, Boston, Massachusetts, USA; 3) PatientCrossroads, San Mateo, California, USA; 4) The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 5) Harvard Medical School, Boston, Massachusetts, USA; 6) Department of Pathology, Brigham & Women's Hospital, Boston, Massachusetts, USA.

Rapid advances in and increased use of genetic testing are resulting in the identification of genomic variation at a rapid rate, but the impact of much of this variation on human health often remains indeterminate, particularly if the variants are novel or rare. Connections between laboratories, clinicians, researchers, and patients are needed to facilitate genomic discovery, variant classification, and patient support efforts. GenomeConnect (genomeconnect. org), an online patient registry developed as part of the NIH-funded Clinical Genome Resource (ClinGen) project, serves as a secure way for individuals and families to connect with one another, participate in research, and engage in data sharing efforts which enhance variant interpretation. By enrolling in GenomeConnect, participants' share their genomic and health information through an online health survey and by uploading a copy of their genetic testing report. Their data are prepared for de-identified data sharing with approved, public databases to facilitate interpretation of genetic variants. Participants also have the ability to connect with one another, as well as with clinicians and researchers. Using the secure online portal, participants can search based on diagnosis, age, or US state. Participants also can elect to receive information about clinicians, laboratories, or researchers trying to connect with them. Individuals with variants in candidate genes also can connect through GenomeConnect's involvement in Matchmaker Exchange, a resource for patients, clinicians, and researchers to search for genetic and phenotypic matches in an effort to determine if a genetic variant in a candidate gene is the explanation for a particular condition. As of May 2016, the growing community currently includes 557 participants from 49 states and 20 countries. Participants have reported more than 200 different diagnoses, including chromosome abnormalities, childhood and adult onset single gene disorders, mitochondrial conditions, and genomic variants in a total of 153 genes. Five percent of GenomeConnect participants have the potential to match with another participant based on gene, 26% could match on a specific diagnosis, and 86% could match on US state. Overall, GenomeConnect provides a means for participants to connect with others while providing clinicians, laboratories, and researchers with genetic and health information that will allow for advances in genomic medicine.

3076W

Engaging families in research by targeting their concerns. *B. Smith-Pack-ard, H. Rocha, K. Fultz, S. Smith, S. Martin, W.A. Faucett, Simons Variation in Individuals Project (Simons VIP).* Genomic Medicine Institute, Geisinger Health System, Danville, PA.

The Simons Variation in Individuals Project (Simons VIP) is an online support community and registry for individuals with a known genetic etiology related to autism. Recruitment goals were met for the initial copy number variants (CNVs) of interest, including 16p11.2 and 1q21.1, and the community expanded to include a number of genes reported to be associated with autism risk, with the continued goal of characterizing the phenotype and natural history of these genetic changes. Various techniques to engage families and encourage enrollment have been explored, including newsletters, webinars, article summaries, social media interaction and addressing family concerns. These efforts serve a dual purpose: they provide resources to support the community and feedback to engage participants. Investigating family-raised questions and treating families as community researchers are unique methods of patient engagement. Fostering a collaborative relationship gives families the opportunity to voice concerns they have noted and allows the study team to acknowledge and explore these concerns. The process of addressing family questions begins with an informal survey to collect more information. Data are shared back with the group, and depending on responses, a pilot project may be developed to investigate further. For example: • 16p11.2 families shared that their children "run hot;" this information was corroborated by responses to an initial survey and a project was developed to track body temperature continuously over three days using a wearable monitoring device. • *SCN2A* families asked for information about the seizures their children experience. To capture this information, an online survey was implemented to allow families to view de-identified responses. The study team can also match family respons-es to lab reports to explore whether the specific variant plays a role in seizure development. Patient engagement has been a key component of Simons VIP. Developing research around questions raised by families and sharing results has helped to strengthen the partnership between families and the study team. Being able to grow a large cohort of families with rare genetic changes is an important first step in facilitating research and may help attract the interest of researchers and aid in the development of new studies and possible targeted clinical interventions. Developing strategies to encourage family interest and participation is a key factor in keeping families engaged.

3077T

Transitioning NIPT to general practice: ELSI challenges in the healthcare organization. *L. Parham[1], L. Manace Brenman[2].* 1) Jurisprudence and Social Policy, University of California, Berkeley, Berkeley, CA; 2) Medical Genetics, Kaiser Permanente Northern California, Oakland, CA.

The use of non-invasive prenatal testing (NIPT) using cell-free DNA (cfDNA) is rapidly expanding in prenatal care. The majority of patients receive con-sultation and ordering for this screening test in a general obstetrics practice, rather than a specialty genetics or maternal fetal medicine setting. As such, general practice clinicians, including obstetricians, prenatal nurses, and health education personnel, are educating and consenting pregnant women on their options regarding this novel genetic technology with complex clinical utility and limitations, particularly vis a vis traditional maternal serum and integrated prenatal screening options. NIPT use, and public demand, is also quickly mov-ing from high-risk to general prenatal populations, just as the options for what components of the test are included is increasing from chromosome aneu-ploidy with or without fetal sex to higher resolution chromosomal copy number variation screening. This presents ethical, legal, and social implications (ELSI) challenges to healthcare organizations to adequately educate non-specialized personnel and their patients about the use of a dynamic and fast-develop-ing technology. Recently, our managed care organization transitioned NIPT education, ordering, and negative resulting for pregnant women of advanced maternal age from specialty genetics consultation to the general women's health practice. Using 30 in-depth clinician/staff and patient interviews, and 80 hours of clinic observation, this study tracks the impact of launching NIPT in a general obstetrics setting, focusing on issues of informed consent and coun-seling practices, and policy and practice adjustment as the testing steadily moves from high-risk patients to the general prenatal population.

3078F

Familial consequences in 25% at-risk individuals after presymptomatic testing for Huntington disease. *A. Bonnard[1], A. Herson[1], C. Boucher[1], S. Staraci[1], M. Gargiulo[2], A. Durr[1,3].* 1) APHP Department of Genetics, Groupe Hospitalier Pitié-Salpêtrière, Paris France; 2) Laboratoire de Psychologie Clinique et Psychopathologie, EA 4056, Université Paris Descartes, Sorbonne Paris Cité, Institut de Psychologie, Paris, France; 3) Institut du Cerveau et de la Moelle, INSERM U1127, CNRS UMR7225, Sorbonne Universités – UPMC Université Paris VI UMR_S1127, Paris, France.

   Presymptomatic testing (PT) refers to the option of being tested for an inherited disease prior to onset. Testing for Huntington disease (HD) is usually restricted to individuals at 50% risk in whom the parent is diagnosed with HD. An individual at 25% risk requests testing with full knowledge that his/her parent does not want to know his/her status. International recommendations states that extreme care should be exercised when PT would provide information about another person who has not requested it. Since the 1992, PT is offered in our center (University Hospital Salpêtrière, Paris). The goal of our study was to i) compare outcome of 25% at risk individuals to those of 50%; ii) assess how test information was disclosed in the family and to the parent. In the case of an unfavorable result, this "double knowledge" has major implications for futures relationships between offspring and the transmitting parent. Psychological issues, particularly the impact to become a "messenger of bad news" for parents and siblings is amplified in this case. We assessed information about pre-test and post-test including age, gender, parent status, motivation for testing, result of the PT and contacted 25% at risk individuals by phone and invited them for an semi-structured interview. There were 1488 individuals at 50% risk and 159 at 25% who took the test between 1992 and 2016. The mean time since testing request was similar in both groups [10.9 years ± 6.4 vs 11.4 years ± 6.5 p=0.416]. 25% at risk were younger at first contact [30.7±9.7 years (13-78) vs 35.1±12.1 years (11-90), p<0.001] and had less often children [42/153 (27%) vs 661/1437 (46 %), p=0.001]. The 50% requested their result after multidisciplinary testing procedure more often than the 25% [1041/1488 (69.9%) vs 93/159 (58.5%), p=0.001], and they were more often carriers of the mutation in *HTT* gene [42% (434/1041) vs 13% (12/93), p<0.001]. Our results showed significant differences between the 25% and 50% at risk groups. Seven interviews were conducted, 5 lost for follow up. One paradoxal reaction occured with a suicide in the 25% at risk group, in a non carrier individual. Detailed psychological outcomes will be reported. .

3079W

Long QT syndrome: A genetic test in the context of a diagnostic algorithm. *E.R. Lockhart[1], C. Greene[2], V.Y. See[3], S. Vashist[2], I.M. Lubin[1].* 1) Division of Laboratory Systems, Centers for Disease Control and Prevention, Atlanta, GA; 2) University of Maryland School of Medicine, Clinical Genetics Service and Department of Pediatrics, Baltimore, MD; 3) University of Maryland School of Medicine, Cardiovascular Medicine Division and Department of Medicine, Baltimore, MD.

   <u>Introduction</u> Long QT syndrome (LQTS) is a hereditary autosomal dominant disorder of cardiac electrical activity that can lead to sudden cardiac arrest (SCA). Timely diagnosis can lead to life-saving therapies. LQTS is typically diagnosed by finding prolonged QT interval on EKG, but a significant percent of affected individuals have normal QT intervals. DNA testing in LQTS evaluation can inform diagnosis, guide personalized therapy, and identify at-risk family members. By developing a diagnostic algorithm of LQTS as a specific use, we gain insight regarding the utility, limitations, and opportunities for integrating genetic testing into LQTS evaluation that can also be used as a model for other diagnostic scenarios.   <u>Methods</u> The LQTS diagnostic algorithm was derived from common practices at the University of Maryland School of Medicine and published professional guidelines. A use case was derived based upon the scenario of a teen with loss of consciousness during athletic activity. This use case served as a framework to address several questions: 1) What is the optimal time to order a genetic test? 2) What are the implications of using genetic tests in light of family history? 3) How are findings that do not establish pathogenicity handled? And 4) How do genetic test results influence treatment decisions?   <u>Results</u> Independent of family history, genetic testing is not immediately ordered because other testing is more expedient to establish LQTS diagnosis. Family history may not always be informative and decisions made as a consequence can influence the significance placed on ordering genetic testing. While several multi-gene panels are available, the choice of testing and the interpretation of variants that are not clearly pathogenic remain a challenge. Parameters specific to each of these decision points are further evaluated.   <u>Conclusion</u> The placement and use of genetic testing within the diagnostic algorithm could be influenced by other types of testing performed, family history, the intended use of the test selected, turn-around-time, and the perceived usefulness of the test results for informing medical decisions. Findings from this case study suggest the need to review diagnostic algorithms with respect to the optimal placement and use of molecular genetic tests for making a diagnosis.

3080T

Learned lessons one-year experience of massive-scale crowd-sourcing platform for genomes and phenomes. *Y. Erlich[1,2], A. Gordon[2], J. Yuan[1,2], D. Zielinski[2], R. Aufrichtig[2], D. Speyer[1,2], J. Pickrell[2,3].* 1) Department of Computer Science, Columbia University, New York, NY; 2) New York Genome Center, New York, NY; 3) Department of Biological Sciences, Columbia University, New York, NY.

Precision medicine is a data-hungry endeavor. However, traditional cohort ascertainment strategies poorly scale and necessitate substantial investments to obtain genomics data, conduct physical exams and lab tests, and assess familial history. But are these really required in today's world? Between 2-3 million people already have access to their digital genome. In addition, the human population has produced zettabytes (10^{21}) of digital data just in the last few years, conveying massive amount of digital phenotypes. Here, I will present our successes in repurposing participants' data for ultra-large scale genetic studies. In ASHG2015, we launched a website called DNA.Land, where participants can contribute their digital genome, genealogical information, and phenotypes. During its eight months of operation we collected about 23,000 genomes from people all over the world. Our approach relies on simple concepts of autonomy, reciprocation, and a respectful consent form to facilitate participation. In this process, we learned important lessons on developing a scalable website to collect massive amount of information, engaging the general public via digital media to address privacy concerns, and return results that matter to participants. These lessons can vastly facilitate other large-scale projects that aim to digitally recruit participants, such as the Precision Medicine Initiative.

3081F

Human genetics at the National Research Centre in Egypt: History, achievements and challenges. *M.A. El Gammal, M.S. Aglan, S.A. Temtamy.* Human Genetics and Genome Research, National Research Centre, Cairo, Dokkie, Egypt.

The National Research Centre (NRC) in Egypt is a multidisciplinary research institute that has a wide spectrum of research fields and interests. It is the largest of all institutions affiliated to the Ministry of Scientific Research in Egypt. It was established in 1956 as a national center to conduct basic and applied research. The Division of Human Genetics and Genome Research (HGGRD), as one of the most distinguished divisions at the NRC, had its origins as the Department of Human Genetics which had been established in 1976 by Prof. Samia Temtamy and small number of scientists. In our presentation we will introduce the journey of the HGGRD at the NRC through 40 years of expansion and development. We will also highlight the achievements and challenges of this unique field of science in Egypt as one of the largest Arab Mediterranean countries. .

3082W

GWS-KQ: A genome-wide sequencing knowledge questionnaire for family testing. *P.H. Birch[1], S. Adam[1], R.R. Coe[1], M.B. Connolly[2], M.K. Demos[2], E. Toyota[2], M.J. Farrer[3], J.M. Friedman[1].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Division of Neurology, Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada; 3) Centre for Applied Neurogenetics, Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada.

    Purpose: We created the GWS-KQ to assess parents' knowledge and understanding of the possible ramifications of genome-wide sequencing (GWS) in family trios (two parents and an affected child) participating in a research study. We were unable to find a published, validated tool that assessed parents' comprehension of issues relevant to trio GWS.   **Method:** A literature review, consultation with genetics health care experts, and interviews and focus groups with parents who had already gone through GWS identified topics deemed important when making informed GWS choices. The issues selected included the nature of genomes, exomes, and incidental findings; the probability of making a diagnosis; variants of uncertain significance; the impact of results on family members; recurrence risks; and insurance implications. Multiple-choice questions were formulated and subjected to several rounds of testing, first with students, staff, and genetics experts, and subsequently with parents. After item analysis and final modification, we administered the questionnaire to 102 GWS-naïve parents of children with epilepsy, both before and after a genetic counselling intervention prior to exome sequencing.   A test-retest analysis of the final 21 questions was performed, two weeks apart, on an independent data set of 27 previously-counselled parents of children having GWS for unknown disorders.   **Results:** In the epilepsy study, parents' mean knowledge scores increased from 44% on the first administration to 65% on the second administration after a genetic counselling intervention (paired t-test, $p < .001$). The questionnaire was internally consistent (Cronbach's α = .83), and responses for each item were significantly correlated with the parents' overall scores (mean $r = .48$ [range .26 to .67], $p < .01$ for all items). In the test-retest study, knowledge scores were consistent ($r = .90$, $p < .001$; mean difference = +2%). Mean time to administer is about 10 minutes.   **Conclusion:** The GWS-KQ questionnaire is quick to administer; appears to have face validity, internal consistency, and sensitivity to change; and demonstrates test-retest reliability. We used GWS-KQ in a research study, but it might also be helpful as a clinical screening tool to identify areas of patients' uncertainty prior to genetic counselling for GWS.

3083T

Physician-patient communication of genome sequencing results in diagnostic odyssey cases. *C.S. Bloss[1], D.L. Boeldt[2], C. Cheung[3], A. Torkamani[2,4], J.R. Friedman[5].* 1) Department of Psychiatry, University of California, San Diego, La Jolla, CA; 2) Scripps Translational Science Institute, Scripps Genomic Medicine, Scripps Health, La Jolla, CA; 3) The Qualcomm Institute of Calit2, University of California, San Diego, La Jolla, CA; 4) Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA; 5) Rady Children's Hospital, San Diego, CA.

    Introduction: Many believe it will ultimately be the job of physicians to communicate genomic information to patients given the relative dearth of genetic counselors and medical geneticists in the U.S. Often cited concerns, however, are that physicians as a group are not trained in genomics and thus their ability to understand, use, and communicate genetic results to patients is unclear. Therefore, we conducted a controlled, quantitative study of physician-patient communication in a case series of diagnostic odyssey patients.   **Methods:** We evaluated physician return of results in the context of an ongoing study of the utility of genome sequencing for diagnosis of rare and idiopathic diseases. Using a modified version of the Medical Communication Competence Scale (MCCS), we leveraged a pre/post experimental design that compared patients' ratings of their physician's communication generally (pre-sequencing) versus his/her communication of genomic results (post-sequencing). These ratings were then directly compared to physicians' self-ratings of their communication of the genomic results.   **Results:** The first 10 cases sequenced in the parent study participated in this study. Across all 10 cases, patient ratings of their physicians' general communication were high (mean pre-sequencing MCCS=123.3 out of 140), suggesting these patients/families had good relationships with their physicians and perceived them to be good at communicating medical information generally. Post-sequencing, however, all 10 patients/families rated their physician's communication of the genetic results as worse than their general communication (mean post-sequencing MCCS=107.6; pre vs. post, p=.010). Furthermore, physicians themselves (with two exceptions) rated their own communication of the genetic results lower than did the patients (physician self-rating MCCS=96.8). We did not observe a relationship between the results of the sequencing (e.g., whether a diagnosis was found) and the ratings of physician communication.   **Conclusions:** Physician-patient communication of genetic information is increasingly important as personal genome sequencing becomes more widely available in clinical settings. Findings from our controlled, quantitative case series study, however, suggest that even physicians with strong communication skills, good patient relationships, and experience with challenging diagnostic odyssey cases may have difficulty relaying genomic information to their patients.

3084F

Recruiting students to the field of genetic counseling – A local effort.
C.A. Campbell[1], S. Feely[2], T.K. Grider[2], J. Phetteplace[2,3], K. Shipley[4], H.M. Bombei[5], J. Marcy[5], J.L. Mancuso[6], K.M. Panzer[5], S. Romoser[5], K. Wiens[5], R.J.H. Smith[1]. 1) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 2) University of Iowa, Department of Neurology, Iowa City, IA; 3) University of Iowa, Department of Internal Medicine Division of Cardiology, Iowa City, IA; 4) University of Iowa, Holden Comprehensive Cancer Center, Iowa City, IA; 5) University of Iowa, Stead Family Department of Pediatrics, Division of Medical Genetics, Iowa City, IA; 6) University of Iowa, Department of Obstetrics and Gynecology, Iowa City, IA.

There is an increasing demand for genetic counselors throughout healthcare, industry, and research. Efforts are underway to increase the number of genetic counselors trained each year. The Iowa Institute of Human Genetics (IIHG) sought to proactively attract students to the field with the hope of growing our future local workforce. We will describe local efforts undertaken to recruit students to the field of genetic counseling. The main aims of this initiative were: 1) raise awareness among students about genetic counseling; 2) prepare students for genetic counseling graduate school. To date, we have utilized a strategic approach for implementation of diverse activities to accomplish these aims. We will present our programmatic framework and considerations when implementing activities such as target populations, genetic counselor engagement, cost, support, activity evaluation, and benefits. Activities developed to date include; Careers in Human Genetics Day, Applying to Genetic Counseling Programs event, Genetic Counseling Summer Internship, and assistance with graduate school interview preparation, clinical shadowing and present at local schools upon request. Target populations include high school and college students, recent graduates, educators.   Success of the programs is measured by participant evaluation. IIHG Genetic Counseling interns have been highly successful in applying to graduate school 7 of 8 interns accepted to graduate school, and one student elected to take a gap year. In addition, a study conducted by the Association of Genetic Counseling Program Directors found in 2014 Iowa led the nation in recruiting students to the field of genetic counseling. Although there is not a genetic counseling graduate program in Iowa, this initiative has allowed us to begin to prepare for a program, and increase our local workforce 1.5 fold in 3 years. These activities can be implemented by genetic counselors who wish to be involved with student development, or to assist in overview genetic counselor workforce development. Our activities utilize multiple genetic counselors, but can be modified based on local resources or be implemented by a single genetic counselor. We believe that communities without genetic counseling training programs can help aid in the national efforts by increasing awareness and recruiting high school and college level students to the field of genetic counseling.

3085W

Satisfaction with receiving genome sequencing results in a preconception population: Implications for future large scale sequencing.
T.L. Kauffman[1], K. Bergen[1], M.J. Gilmore[2], P. Himes[2], M.C. Leo[1], F. Lynch[1], L.M. Amendola[3], S. Punj[4], P.D. Robertson[5], C. McMullen[1], C.S. Richards[4], D.A. Nickerson[5], G.P. Jarvik[3,5], B.S. Wilfond[6], K.A.B. Goddard[1]. 1) Center for Health Research, Kaiser Permanente, Portland, OR; 2) Department of Genetics, Kaiser Permanente Northwest, Portland, OR; 3) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 4) Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Seattle Children's Research Institute, Treuman Katz Center for Pediatric Bioethics, Seattle, WA.

A variety of modalities is available for disclosing results of genetic testing, including in-person or telephone visits, telemedicine, and mailed letters. Studies evaluating various approaches to pre-test genetic counseling in the context of either cancer syndromes or carrier and prenatal testing have found a high level of satisfaction with all of these modalities, but methods of post-test counseling and results disclosure have not been explored. As part of the Clinical Sequencing Exploratory Research (CSER) consortium, we looked at satisfaction with post-test counseling and results disclosure from genome sequencing for preconception carrier status via in-person visits (for abnormal results) or letters (for normal results).   Women eligible for the study had a preconception or prenatal carrier test ordered by their provider, were planning a pregnancy, and were Kaiser Permanent Northwest members. Women could receive carrier results on ~750 conditions; they could also receive additional findings on ~100 medically actionable conditions. Abnormal findings were disclosed during an in-person genetic counseling session. Normal results were disclosed via a letter. Participants were surveyed 2 weeks after receipt of results about their satisfaction with the method by which the carrier and additional findings were disclosed. The 6-item satisfaction scale has a possible score of 6-30, with higher values indicating greater satisfaction.   Eighty women completed a survey following disclosure of carrier findings: 61 had in-person visits, and 19 received letters. Sixty-six women who received letters notifying them of normal results for additional findings completed a survey. Mean and standard deviations (SD) of satisfaction scores for the letter and in-person visit for disclosure of carrier results were 22 (SD=6.1) and 26 (SD=3.7), respectively. Mean satisfaction for the letter for disclosure of normal additional findings was 23 (SD=3.7).   Overall satisfaction was high for disclosure of results via both in-person genetic counseling visits and letters, although there was slightly less satisfaction with the letters. Based on individual survey item responses, the language in the letters could be refined for clarity. While we returned only normal results via letters, future studies could explore the possibility of using letters to disclose abnormal results. This could save substantial cost and time, and make preconception genome sequencing more feasible in a clinical setting.

3086T

Genetic counseling training and services in the Asia region. *M. Laurino[1,2], D. Sternen[3], J. Thompson[4], K. Leppig[5,6]. 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) University of the Philippines, Malia, Phillippines; 3) Seattle Children's Hospital, Seattle WA; 4) BC Cancer Agency, Vancouver, British Columbia; 5) Group Health Cooperative, Seattle WA; 6) Department of Pathology, University of WA, Seattle, WA.*

The Genetic Counseling Pre-Conference Workshop (GCPCW) was held on September 16, 2015 in Hanoi Vietnam, prior to the 11th Asia Pacific Congress for Human Genetics. A total of 71 signed registrants attended the GCPCW. Representatives from Australia, Hong Kong, India, Indonesia, Malaysia, Philippines, Singapore, Taiwan, Thailand, and Vietnam reported on the status of genetic counseling in their respective country. Five countries (Australia, India, Philippines, Singapore, and Taiwan) have developed master's level genetic counselor training programs. Of the remaining five countries, four provide other models of genetic counselling training, ranging from genetic disease-specific courses lasting 3-5 days to a master's program in biomedical science with a genetic counseling focus.   The pre-conference also included a case-review breakout session and a discussion of needs for those providing genetic counseling in the Asia Region. Pre-conference and post-conference questionnaires were returned by 57 and 44 individuals, respectively. Of the 42 registrants who did not identify themselves as current students in training, 86% stated that they were currently providing genetic counseling and had training ranging from an undergraduate degree to MD with fellowship training. 100% of registrants strongly agreed or agreed that their expectations of the pre-conference were met, 98% would attend a similar pre-conference with preferred topics including best practices in genetic counseling, East vs. West counseling differences, communication skills, cross- cultural training, and research studies pertinent to genetic counseling .   The GCPCW provided a forum for those who currently provide genetic counseling in the Asia region to address the challenges and needs of genetic counseling in the region. Participants cited that the most useful information obtained during the GCPCW was the availability and status of genetic counseling in the Asia region, followed by learning of shared challenges, professional networking, opportunities for collaboration, the need to promote awareness of genetic counseling, and the development of accreditation. The formation of the Professional Society of Genetic Counselors in Asia was announced at the GCPCW and will provide an ongoing resource for education, advocacy and collaboration for genetic counselors working in Asia.

3087F

Carrier screening program for Cree Encephalitis and Cree Leukoencephalopathy: Long-term knowledge and carrier status retention in high school students. *J. Le Clerc-Blain[1], V. Gosselin[2], A. Bearskin[3], J.E. Torrie[2], G.A. Mitchell[1,4], B.J. Wilson[5], A. Richter[1,4], A.M. Laberge[1,4]. 1) CHU Sainte-Justine Research Center (Montreal, QC, Canada); 2) Cree Board of Health and Social Services of James Bay (Chisasibi, QC, Canada); 3) Eeyou Awaash Foundation (Chisasibi, QC, Canada); 4) Medical Genetics Division, CHU Sainte-Justine and Department of Pediatrics, University of Montreal (Montreal, QC, Canada); 5) School of Epidemiology, Public Health and Preventative Medicine, University of Ottawa (Ottawa, ON, Canada).*

BACKGROUND: Cree Encephalitis (CE) and Cree Leukoencephalopathy (CLE) are two neurodegenerative autosomal recessive conditions with high carrier rates in the James Bay Cree communities of Northern Quebec (Canada). Education/counselling sessions are offered to high school students (≥ 14 years) through the CE-CLE carrier screening program (CSP), a population-based CSP developed by local health authorities in collaboration with the Eeyou Awaash Foundation (a community family support group). OBJECTIVE: To describe high school students' knowledge of the CE-CLE CSP 6-12 months after their education/counselling session, and evaluate their carrier testing result retention. METHODOLOGY: Surveys were handed out before (survey A), after (survey B), and 6-12 months (survey C) following the CSP education sessions to Quebec secondary 3-5 (grades 9-11 equivalent). Data was collected on demographics, knowledge of CE-CLE, attitudes about screening, and carrier status retention. Knowledge questions were compared between participants who answered all three surveys using t-tests (n=50). Participants' answers about their carrier status were compared to their test results in the clinical database. RESULTS: Of 359 eligible students, 134 (37%) answered survey A. Of these, 71 (53%) answered survey B and 72 (54%) survey C. Ages ranged from 14-20 years. Twenty-three (17%) were either pregnant or already had at least one child. Twelve (9%) reported a positive family history for CE or CLE. Out of 14 knowledge questions, the mean number of correct answers was 4.5 (survey A), 7.3 (survey B) and 6.9 (survey C). Between pre-screening (A) and immediate post-screening (B) surveys, we observe a significant increase in the number of correct answers (p=0.003). The number of correct answers 6-12 months post-screening (C) is slightly lower than right after screening (B), but this difference is not significant (p=0.68). Knowledge remains significantly higher 6-12 months post-screening (C) than pre-screening (A) (p=0.01). Out of 72 students who answered survey C, 62 (86%) remembered correctly participating or not in the screening. Out of the 54/72 (75%) students who did the screening test and for whom results were available, 34/42 (81%) accurately remembered their carrier status. CONCLUSION: Post-screening knowledge and carrier status retention is encouraging, pointing to the lasting impact of the CE-CLE CSP education sessions on high school students. This program may serve as a model for other CSPs.

3088W

Development of a practice model of genomic counseling for actionable complex disease and pharmacogenomics. *K. Sweet[1], A.C. Sturm[1], S. Hovick[2], T. Schmidlen[3], E.S. Gordon[4], K.E. Ormond[5], B.A. Bernhardt[6], J. O'Daniel[7], J. McElroy[8], L. Scheinfeldt[3], A.E. Toland[1], J.S. Roberts[9], M. Christman[3].* 1) Division of Human Genetics, Ohio State University Wexner Medical Center, Columbus, OH; 2) School of Communication, Ohio State University, Columbus, OH; 3) Coriell Institute for Medical Research, Camden, NJ; 4) 23andMe, Mountain View, CA; 5) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 6) Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, PA; 7) Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC; 8) Center for Biostatistics, Ohio State University, Columbus, OH; 9) Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI.

Given the movement towards both multi-gene testing and online patient access, optimization of genomic counseling (GC) services to support multiplex results delivery is needed. More information on patients' motivations, preferences, and informational needs are essential to guide the development of new, more efficient practice delivery models that capitalize on the existing strengths of a limited genetic counseling workforce. To explore this, we sought to define essential content and process elements through formative research with participants receiving multiple (n= 26) online actionable complex disease and drug response results via the OSU-Coriell Personalized Medicine Collaborative. Semi-structured telephone interviews were conducted with study participants who had either telephone GC (TGC; community cohort, n=31; mean age 46.8 years) or in person GC (IPGC; chronic disease cohort, n=20; mean age 60 years). Interviews were used to a) clarify participant expectations and impressions of GC; b) determine the most and least valuable components of the IPGC and TGC models; and c) identify preferences for follow-up GC. Transcripts were analyzed using a Grounded Theory framework. Although all participants felt satisfied with the GC received, there was expressed need for a more participant-driven GC model, more control over the results discussed, more personalized risk reduction recommendations, and the option of follow-up counseling for new results. Based on these findings we developed a novel GC practice model which involves patient completion of an online survey, after receipt of online results and before GC, to select test results and provide personal questions they would like to cover in GC, and to indicate preferences for counseling modality (telephone, online or in-person). The individual preferences and questions are used to tailor the GC session and personalize result communication and risk reduction recommendations. Newly developed visual aids and result summary reports break down areas of risk (genetic, family history, lifestyle) for each disease to facilitate viewing of multiple risk results at once. Post GC, summary reports are actively routed to the participant and physician to encourage review and follow up on recommended disease prevention/risk reduction actions. The GC practice model has been evaluated in 44 new study participants, and is adaptable for use in large-scale genomic sequencing efforts.

3089T

Exploring the content and process of returning results from exome sequencing. *S.A. Walser[1], A. Werner-Lin[1], S. Biswas[2], R. Mueller[1], B.A. Bernhardt[1].* 1) University of Pennsylvania, Philadelphia, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA.

In an NIH-funded Clinical Exploratory Sequencing Project, the PediSeq study, we are offering exome sequencing to children with unexplained hearing loss, sudden cardiac arrest, mitochondrial disease, platelet function disorders or neurodevelopmental disability. In addition to diagnostic results, families can opt to learn findings that are actionable in childhood and adulthood, and carrier status. We analyzed 25 audiotaped return of results (ROR) sessions conducted by one or several of 7 genetic counselors or 6 physicians. Eleven families received pathogenic and 7 possibly pathogenic results relating to the indication for testing, 3 received medially actionable secondary findings and 14 received carrier status results. Sessions ranged from 7 to 100 minutes. In most sessions, clinicians presented results by following a 14 page results report and providing technical details about the variants found and sequencing limitations. Families asked many clarification questions attempting to understand and personalize results. Many families scanned ahead of clinicians in the report to learn additional findings. Clinicians presented diagnostic pathogenic mutations as "we found the answer". Families responded with satisfaction that their diagnostic odyssey was over, disappointment results would not modify treatment, or relief from guilt. Both negative and uncertain diagnostic findings were presented by the clinician with a promise that scientific advances might provide future clarity. Families responded to such results with disappointment, confusion, resignation, suspicion, hope or surprise. Although most received carrier status results, many families commented that they had expected more secondary findings, particularly for common adult diseases. Some believed their risk for conditions such as cancer was reduced because of the lack of secondary findings. Multiple families incorrectly expected exome sequencing results would be available on all family members who provided a study-related blood sample. We conclude that structuring ROR sessions around a complicated results report and the clinicians' use of excessive technical language may confuse and overwhelm families, and not address their main concerns: management based on test results and risk of common diseases. Pre-test informed consent sessions focusing on realistic expectations of sequencing results and the limitations of sequencing might better prepare families for the kinds of results returned. .

3090F

Raw DNA availability and downstream implications for genetic service providers. *C. Wang[1], N. Thomas[1], T. Norkunas[1], M. Flynn[2].* 1) Boston University School of Public Health, Boston, MA; 2) Boston University School of Medicine, Boston, MA.

Raw DNA data and corresponding online interpretation services are widely available to the public, yet little is known about their downstream impact on genetic service providers. We conducted a survey to examine the extent to which consumers of these personal genetic services are reaching out to genetic counselors and to identify issues stemming from any clinical encounters. A total of 83 genetic counselors completed a web-based survey that was administered via listservs of the National Society of Genetic Counselors in Spring 2016. Among respondents, 53% (N=44) had been contacted by a patient to schedule an appointment following the use of an online service to interpret raw DNA data; mean number of patients was 3.84. Among the counselors contacted, 72% indicated they had seen at least some of these patient cases in clinic. Disease categories for counseling requests included cancer (52%), neurology (21%), metabolism (17%), pediatrics (10%) and psychiatric (10%). In preparation for these cases, 64% of counselors researched the DNA interpretation service, 52% consulted genetic databases, and 16% contacted the service that interpreted the raw DNA. Average amount of time to prepare for these cases was 2 hours (range 0-6 hours). Online services used by patients included Promethease (58%), LiveWello (23%), Genetic Genie (8%), and GEDMatch (4%). Reports generated from these services were rated poorly by genetic counselors based on clarity of format (mean 1.83; 1=unclear to 5=clear) and informativeness (mean 1.54; 1=not at all informative to 5=extremely informative). When asked to weigh the benefits versus limitations of raw DNA interpretation services, 77% of counselors felt the limitations outweighed the benefits, 12% were neutral, and 11% felt the benefits outweighed the limitations. Notably, counselors who had never seen a case were more likely to feel that the benefits outweigh limitations compared to those who had seen a case (16% vs. 4%). Qualitative responses highlight challenges to counseling, including unrealistic expectations for counselors to interpret the (long) reports or raw DNA, overemphasis by patients on the legitimacy of the information within the reports, and patient frustrations if the session did not provide them with answers they were searching for. Overall, the availability of raw DNA data poses several challenges to genetic counselors who will increasingly be faced with requests by consumers to assist in data interpretation.

3091W

A novel framework for optimizing the value of personalized medicine co-dependent technologies. *P. Akude, R. Mahjoub, M. Paulden, C. McCabe.* Emergency Medicine, University of Alberta, Edmonton, Alberta, Canada.

The arrival of personalized medicine has led to an increasing recognition of the need to reconcile the evidence requirements for test and treatment technologies. In this research, we develop a method for combining evidence on the test(s) and treatment components of co-dependent technologies that identify the cost effective cut-points on the test components for pre-specified values of the Willingness to Pay for Health. The framework we propose consists of three tests (genotypic diagnosis, phenotypic expression and therapy responder status) and a treatment. The first test identifies the existence of the health condition. Based upon the presence of the condition of interest, the second and third tests characterize the ability to respond to therapy and the phenotypic expression (which places a limit on the ability to benefit from therapy) respectively. We identify two decision variables – the cut-point for the probability of responding to therapy and the cut-point for the phenotypic expression that leads to treatment. For a given probability of response, patients who are identified to benefit from treatment are administered a new treatment while other patients are prescribed standard care. The effectiveness of the therapy in responders and non-responders is exogenously determined. We show that for a given probability of response, the optimal cut-point for the test of phenotypic expression is identified at the point at which benefits to a responding patient means that the patient is indifferent between the new treatment and standard care. We present a series of analyses exploring the relationship between the distributions of the probability of responding to therapy, phenotypic expression and the net benefit from the new technology. Our analyses demonstrate that the benefit from the adoption of personalised medicine co-dependent technologies can be optimized by treating response probability and phenotypic expression as decision variables not exogenously determined parameters.

3092T

Predictors of anxiety and depression in caregivers of children with chronic conditions of unknown but suspected genetic origin. *N.I. Drago-jlovic[1], A.M. Elliott[2], J.M. Friedman[2], L.D. Lynd[1], CAUSES Study Group.* 1) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

Background: Caregivers of children with chronic health conditions tend to have increased anxiety and depression when compared to parents of healthy children. The lack of a diagnosis has been shown to act as a major psychological stressor for parents, but it remains unclear whether this is purely a result of the associated uncertainty, or whether the additional burden imposed by diagnostic investigations plays a role. **Objective:** To explore the association between healthcare utilization of children with idiopathic chronic conditions of suspected genetic origin enrolled in the CAUSES genome-wide sequencing clinic at British Columbia Children's Hospital and caregiver mental health. **Method:** One parent of each child was invited to complete an online survey that measured her or his mental health and the child's healthcare resource utilization. The Patient Health Questionnaire depression scale (PHQ-8; range 0-24) and the GAD-7 scale (range 0-21), a screening tool for generalized anxiety disorder, were used to assess the presence and severity of depression and anxiety, respectively. Parents reported their child's number of healthcare system encounters in the 6 months prior to survey completion and how many of these were primarily related to obtaining a diagnosis. The number of hours spent providing extra care due to the condition was used as an indicator of the condition's severity. Ordinary least squares (OLS) regression was used to test for associations with depression or anxiety scores. **Results:** 52 parents completed the questionnaire. Average healthcare resource utilization was 8.7 [IQR: 4 to 14] encounters, with 3.4 [0 to 4] being diagnosis-related and 5.2 [2 to 8] otherwise. The two types of encounters were not correlated (r = - 0.15, p = 0.29). Average scores were 5.0 [1.0 to 7.5] for depression and 5.7 [2 to 7] for anxiety. *Table 1:* OLS regression coefficient estimates for potential drivers of parental mental health. Standard errors in parentheses. + p < 0.10 * p < 0.05

	Depression	Anxiety
Total # of medical encounters	0.31 (0.12)*	0.19 (0.10)+
Hours of extra care per week	-0.02 (0.02)	-0.01 (0.01)
Child's age (years)	-0.34 (0.16)*	-0.06 (0.13)
N =	47	51
Adjusted R^2 =	0.18	0.02

Discussion: The number of healthcare encounters of the child was positively associated with the parent's levels of both depression and, to a lesser extent, anxiety.

3093F

Evaluating the cost-effectiveness of sequential monitoring tests. *R. Mahjoub, P. Akude, M. Paulden, C. Hollman, C. McCabe.* University of Alberta, Edmonton, Alberta, Canada.

A monitoring test is a test that is repeated in a patient over a period of time to identify changes in the patient's medical status. In this study we extend upon the existing literature for monitoring tests, allowing the test to be administered repeatedly for more than two periods, as well as incorporating the "treat without testing" and "skip test" strategies in each period. We also include "treat" or "do not treat" sub-strategies under the "skip test" strategy. Then for this setting, we examine the cost-effectiveness of a monitoring test and investigate the optimal strategy in each period. We also investigate the optimal cut-offs for test scores in those periods, in which it is optimal to apply the test. We illustrate with a numerical example using data from a Ca125 test for monitoring ovarian cancer. Our results show that the net health benefits resulting from the monitoring test and the treatment will be maximized if we optimize simultaneously for the test score cut-off points, re-test intervals, and the treat or do not treat sub-strategies. We conclude that optimal test cut-offs in a monitoring regime are population case-mix and health system specific and should not be assumed to be portable. To assume portability is to reduce the population health impact (value) of test technologies.

3094W

Primary care providers' experiences and perceptions of personalized medicine. *J.C. Carroll[1,2], T. Makuwaza[1,2], D. Manca[3], N. Sopcak[3], J.A. Permaul[2], M.A. O'Brien[1], R. Heisey[1,4], E.A. Eisenhauer[5], J. Easley[6], M. Krzyzanowska[7,8,9], B. Miedema[6], S. Pruthi[10], C. Sawka[8], N. Schneider[N/A], J. Sussman[11], R. Urquhart[12,13], C. Versaevel[3], E. Grunfeld[1,14] on behalf of CanIMPACT.* 1) Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada; 2) Ray D. Wolfe Department of Family Medicine, Sinai Health System, Toronto, ON, Canada; 3) Department of Family Medicine, University of Alberta, Edmonton, Alberta, Canada; 4) Department of Family and Community Medicine, Women's College Hospital, Toronto, ON, Canada; 5) Department of Oncology, Queen's University, Kingston, ON, Canada; 6) Department of Family Medicine, Dalhousie University, Fredericton, NB, Canada; 7) Department of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Toronto, ON, Canada; 8) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 9) Cancer Care Ontario, Toronto, ON, Canada; 10) General Internal Medicine, Mayo Clinic, Rochester, Minnesota, US; 11) Department of Oncology, McMaster University, Hamilton, ON, Canada; 12) Beatrice Hunter Cancer Research Institute, Halifax, NS, Canada; 13) Department of Surgery, Dalhousie University, Halifax, NS, Canada; 14) Ontario Institute for Cancer Research, Toronto, ON, Canada.

Context: Lack of knowledge in addition to systemic and practice issues impair the ability of primary care providers (PCPs) to embrace genomic medicine. Better understanding of PCPs' attitudes towards genomic medicine is needed to inform implementation of personalized medicine (PM) into primary care. **Objective:** To assess PCPs' experiences and perceptions of PM and their desired role, with a focus on cancer. **Design:** Qualitative study. **Setting:** Urban and rural inter-professional primary care team practices in Alberta and Ontario, Canada. **Participants:** Family physicians, nurses, nurse practitioners, physician assistants. **Method:** Semi-structured focus groups were conducted, audio-recorded, and transcribed. Analytic techniques were informed by grounded theory including coding, interpretations of patterns in the data and constant comparison. **Findings:** 5 focus groups with 51 participants in total were conducted; 2 in Alberta, 3 in Ontario. PCPs described limited experience with PM, citing breast cancer and prenatal care as main areas of involvement. They expressed concern over their lack of knowledge, particularly identifying appropriate patients for referral for genetic counseling and testing and were often unaware of the benefits of genetic testing. In some circumstances PCPs relied on personal experiences to inform their attitudes and practice. Participants anticipated an inevitable role in PM primarily because patients seek and trust their advice. Concern was expressed about the magnitude of information and pace of discovery in PM, particularly in direct-to-consumer personal genomic testing. PCPs described unfamiliarity with genetics specialists and clinics, which affected referrals. Increased knowledge, closer ties to genetics specialists and relevant, reliable PM resources accessible at point of care were considered important for successful implementation of PM. **Conclusions:** PCPs are prepared to discuss PM but they require better resources. Models of care that support a more meaningful relationship between PCPs and genetics specialists should be pursued. Continuing education strategies need to address knowledge gaps including identification and management of patients at increased genetic risk and direct-to-consumer genetic testing. PCPs should be mindful of using personal experiences to guide care. .

3095T

Status of trainee funding in ABMGG laboratory training programs. *T. Cowan[1], A. Cherry[1], M. Basehore[2], R. Schreck[3], L. Jeng[4].* 1) Stanford University School of Medicine, Stanford, CA; 2) Fullerton Genetics Center, Asheville, NC; 3) Cedars-Sinai Medical Genetics Institute, Los Angeles, CA; 4) University of Maryland School of Medicine, Baltimore, MD.

Funding sources for American Board of Medical Genetics and Genomics (ABMGG) laboratory fellows has been an ongoing concern for laboratory training in Biochemical Genetics, Cytogenetics and Molecular Genetics. With anticipated growth in Personalized Medicine, there is an increasing need for laboratorians certified in genetic laboratory specialties. In order to evaluate the current funding sources for laboratory trainees in ABMGG accredited programs, the Association of Professors of Human and Medical Genetics (APHMG) Laboratory Fellowship Program Director's Special Interest Group (SIG) surveyed training programs. A survey link was sent via Qualtrics Survey Software to program directors, training directors, and program coordinators at 45 institutions. Thirty-nine responses were received from 30 programs. About half of the respondents indicated that they have guaranteed funding for their trainees; of these, the main sources utilized by programs include a combination of institutional funds (54%), laboratory operating budget (64%), research grants (18%) and NIH training grants (10%). Some respondents indicated other sources including commercial laboratories and other internal and external funding. A similar survey was administered to ABMG laboratory training program directors in 2010, and showed similar trends. Of interest, 63% of programs said they could possibly accept additional trainees if more funding was available. Identifying consistent, guaranteed funds to support laboratory fellowship trainees would increase the ability of many of our existing training programs to train additional qualified work force to provide necessary genetic testing.

3096F

Common threads: Reflective practice connecting medical genetics concepts across an integrated curriculum. *S. Dasgupta.* Department of Medicine, Biomedical Genetics, Boston University School of Medicine, Boston, MA.

Recent curricular surveys have identified increasing numbers of medical schools moving towards integrated pedagogical models. At the same time, AAMC data indicates that the average proportion of the pre-clerkship curriculum dedicated to medical genetics is approximately 2%[1]. An integrated curriculum structure risks diluting the genetics concepts until the discipline identity is lost. To help preserve the identity of genetics within an integrated curriculum, a longitudinal reflective practice was introduced into the first year Principles Integrating Science and Medicine (PrISM) curriculum. This activity was tied to a series of patient visits held in four distinct modules of the PrISM curriculum. Following each patient session, the students were given an in-class reflective writing prompt. For the first two sessions, the anonymous writings focused on the stories' connections to their own lives and values, and subsequent sessions were paired with preceding patient visits. For example, the students' fall semester writings from meeting the parent of a child with trisomy 18 were returned to their peers for further comment after a spring semester session featuring a self-advocate with Down Syndrome. The writing guided students to contrast the similar genetic etiology and dramatically different prognoses of these conditions. Likewise, the students' fall semester reflections about women harboring hereditary breast and ovarian cancer-predisposing mutations were returned to their peers in the spring semester, after meeting a guest with multiple family members affected by Huntington's Disease. While both are single gene disorders with genetic testing options, the ability to select risk reducing procedures in one was juxtaposed with receiving a prediction for an incurable disease in the other. By structuring these reflective writings to highlight the common threads and contrasting outcomes between conditions, the students engaged with the clinical relevance of the genetics content and grappled with the ethical ramifications of genetic testing.   [1]Plunkett-Rondeau J, Hyland K, Dasgupta S. Training future physicians in the era of genomic medicine: trends in undergraduate medical genetics education. *Genet Med*. 2015. doi:10.1038/gim.2014.208.

3097W

Awareness, respect, and confidence: Increasing community conversation about genetics through a multi-faceted and interdisciplinary approach. *M. Gelbart[1], J. Kung[1], L. Tomaselli[1], D. Waring[1], E. McMillan[2], T. Wu[1].* 1) Personal Genetics Education Project, Department of Genetics, Harvard Medical School, Boston, MA; 2) Sanford Program for the Midwest Initiative in Science Exploration (PROMISE), Sanford Research, Sioux Falls, SD.

The Personal Genetics Education Project (pgEd.org) is using innovative strategies to hold community conversations and raise awareness about advances in genetics and the related ethical and social issues. pgEd is launching a new project, ARC: Building Awareness, Respect, and Confidence through Genetics, that aims to partner with teachers across all disciplines to stimulate dialogue about cutting-edge scientific and social developments in personal genetics within high schools and the broader community. First, we are creating curricular resources and adaptations for the public that revolve around the theme of genetics and identity, including modules on sex and gender, race and ancestry, intelligence and cognition, and gene editing. Second, we are developing an interdisciplinary professional development workshop for educators in urban Massachusetts and rural South Dakota communities, in partnership with the Sanford PROMISE. The first implementation of this workshop will take place in summer 2016, and we are anticipating ~50 teachers to attend from in and around Boston and across the nation. ARC will be integrated into the broader platform of pgEd's activities to engage audiences through schools, libraries, communities of faith, television and film, Congressional briefings, and social media. We will report on the most recent developments in these areas including foundational work with the Cornerstones of Science and library partners, our most recent Congressional briefing on the intersection of genetics and space travel, initial conversations with faith leaders, and pgEd's Industry Forum for Forging Community Partnerships.

3098T

The impact of personal experience with direct-to-consumer genetic testing on physician readiness. *E.S. Gordon, E. Kim, J. Cintron, R. Gold, J. Pollard, E. Nelson.* 23andMe, Mountain View, CA.

In recent years, there has been a shift towards patient empowerment in healthcare, punctuated by policy changes. In 2014, HIPAA was updated to allow patients to receive test results directly from laboratories. In 2015, the National Society of Genetic Counselors and the Association for Molecular Pathology updated their position statements in support of direct access genetic testing provided certain conditions are met. In the same year, the FDA granted marketing authorization to 23andMe for its Bloom Syndrome carrier status test, the first directtoconsumer genetic test (DTCGT), paving the regulatory path forward for more DTCGTs. Due to increased access, primary care providers (PCPs) are more likely to encounter patients inquiring about this type of testing but they may not be prepared. The role of personal genetic testing (PGT) as a potential tool for improving PCPs' awareness of and readiness for genomics has been suggested by researchers but has not been demonstrated. To assess the impact of PGT on PCP comfort with DTC-GT in the primary care setting, we analyzed previously collected data from a convenience sample of PCPs (MD, DO, NP and PA) who visited the 23andMe medical professional website between Oct 21, 2015 and Jan 22, 2016. 140 PCPs consented to the survey program with 70 completing all requirements, which included a pre-PGT survey, DNA kit registration, review of their personal 23andMe reports and post-PGT survey. The surveys were administered via online survey portal QuestionPro. Survey responses from 70 PCPs were analyzed using a o ne-sample z test, to identify differences between pre- and post-PGT. Following their PGT experience, significantly more PCPs reported feeling "very comfortable" interpreting DTCGT (47% vs. 26%, P<0.05) and making medical recommendations based on DTCGT (29% vs. 12%, p<0.05); whereas prior to PGT, the majority of respondents were "somewhat comfortable". In addition, respondents reported that their firsthand PGT experience was somewhat or very important for understanding carrier testing in general (77%), genetic testing in general (82%), key genetic concepts in general (77%) and utility of carrier testing in clinical care (77%). This study suggests potential value in PGT for PCPs to enhance comfort in dealing with DTCGT results; however the generalizability of these results is limited by the convenience sample used. Further exploration of the impact of PGT on PCPs naive to DTCGT is warranted.

3099F

Genetics Home Reference website as a point-of-care resource in the genomic medicine era. *K. Greenberg[1], S. Calvo[2], H. Collins[1], L. Forman Neall[3], S. Morrison[2].* 1) ICF International, Fairfax, VA; 2) U.S. National Library of Medicine, NIH, Bethesda, MD; 3) Medical Science & Computing, LLC, Rockville, MD.

With faster and cheaper genome sequencing technologies and the President's Precision Medicine Initiative, genetic data is becoming an integral component of health care. All clinicians will need a grasp of human genetics to be able to provide care in this era of genomic medicine. Genetics Home Reference (https://ghr.nlm.nih.gov/), a trusted consumer website from the U.S. National Library of Medicine, conveys complex genetic information through brief, plain language summaries of genetic conditions and associated genes. While the site is an excellent resource for the general public, it is also a point-of-care resource for clinicians to enhance their understanding of the role of genetics in medicine. A recent redesign of the Genetics Home Reference website improves its usability on mobile devices, making the site readily available for consultation. On a smartphone or tablet, clinicians can easily access information about more than 1,100 genetic conditions, including rare disorders unfamiliar to most physicians. They can also find summaries for more than 1,300 genes, all the chromosomes, and mitochondrial DNA, which can help put into context genetic testing results and aid in understanding the genetic basis of a diagnosis. The redesign also enhances overall usability of the website, making it easier for users to find information of interest. Navigation is streamlined, with in-text links that help users more easily move between related pages, and colors and icons that help distinguish the website's different content areas. In addition, condition summaries contain text-linked images that can be useful visual aids for understanding the features of a condition. Furthermore, clinicians can provide patients with printed versions of site content, making the information available to those without access to the Internet. In addition to particular conditions and genes, patients might find especially useful the website's primer *Help me Understand Genetics*, which discusses basic genetic concepts, such as how gene mutations affect health, how conditions are inherited, and how genetic testing is done. The Genetics Home Reference website is a useful point-of-care resource for clinicians as genetics becomes more important in health care. Improvements to the site, such as better navigation, mobile responsiveness, and the addition of educational images, will enhance the clinician's experience of using the site.

3100W

An online tool for applying the ABMGG laboratory training program milestones. *L. Jeng, J. Kleinberger.* University of Maryland School of Medicine, Baltimore, MD.

Recently, the American Board of Medical Genetics and Genomics (AB-MGG) announced the release of milestones for laboratory-based training to improve consistency of training and trainee evaluation. These milestones are a required component of training evaluations as of July 2016. In order to assist programs in using the milestones as part of the evaluation process, we have developed a simple online tool that enables programs to record which milestones have been successfully achieved by a trainee. The tool is available at http://ljeng28.wix.com/jeng-resources. When initially launched, the tool displays a list of twelve competencies (eg. Perform laboratory testing) under competency domains (eg. Patient Care). Clicking on a competency expands a list of milestones under that competency. Individual milestones can then be checked off after accomplished. The tool totals the number of milestones achieved under each competency plus each competency domain and provides a percentage of the total milestones. The trainee's progress in achieving milestones can be easily summarized by the tool in either a single page "Short PDF" showing the number of milestones achieved for each competency and competency domain or a four page "Extensive PDF" showing each milestone and whether or not it has been achieved. Either PDF version can be printed or saved and provides space to enter trainees name, date, comments and signatures, either by hand on the print out or by using a PDF editing program. This tool provides a simple way for programs to quickly implement the milestones into the evaluation process at regular intervals and to develop open communication between with the trainee regarding his or her progress on the milestones.

3101T

Development of a consent resource for genomic data sharing in the clinical setting. *E.R. Riggs[1], D. Azzariti[2], A. Niehaus[3], S. Goehringer[1], E. Ramos[3], H.L. Rehm[2,4], C.L. Martin[1], ClinGen Education Working Group.* 1) Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA; 2) Partners Healthcare Personalized Medicine, Boston, MA; 3) National Human Genome Research Institute, NIH, Bethesda, MD; 4) Harvard Medical School, Boston, MA.

Data sharing between clinicians, laboratories, and patients is essential for improvements in genomic medicine. The need for broad access to individual data must be balanced with respect for patient autonomy and privacy. Acknowledging that obtaining consent for sharing is hindered by lack of time and resources, the National Institutes of Health's (NIH) Clinical Genome Resource (ClinGen) is developing broad data sharing consent resources for use by clinicians and clinical laboratories, including a 1-page consent form and brief online videos. The consent form contains language consistent with the NIH Genomic Data Sharing Policy and NHGRI Informed Consent Resource, and the videos provide information on key topics, such as risks and benefits of data sharing (clinicalgenome.org/share). ClinGen held interviews and focus groups in 4 US cities with key stakeholders [clinicians (n=40), clinical laboratory staff (n=33), and community members (n=48)] for feedback. Preliminary thematic analysis suggests that participants feel that genomic data sharing is a worthwhile endeavor. Clinicians and laboratories cited potential to aid interpretation of variants of uncertain significance, and community members viewed it as a way to contribute to the "greater good." Participants noted that a brief consent was essential, but that the videos provided valuable information. Community members requested more information on the benefits of data sharing, while clinicians and laboratory staff were concerned with "over-promising" on benefits. To assess comprehension of concepts related to genomic data sharing, participants completed a 5 question knowledge assessment before and after reviewing materials. After review, community members' knowledge significantly improved (57% vs. 91% correct, $p \leq 0.01$), suggesting the resources effectively communicate data sharing concepts. ClinGen hopes these resources will provide a straightforward way to share genetic and health information, and help the scientific community capitalize on the largely untapped resource of data generated through routine healthcare.

3102F

Insights into residents' utilization of clinical genetics skills and services and medical genetics educational experiences. *E. Syverson, T. Pollin, S. Dixon, M. Blitzer, J. Plamondon.* University of Maryland School of Medicine, Baltimore, MD.

As genetics becomes increasingly integrated into other medical specialties, it has become the responsibility of clinicians to acquire skills and knowledge that allow them to provide genetics services or to accurately refer patients to medical genetics specialists. Although there has been a recent educational push to prepare physicians, more targeted education is necessary. In order to achieve this and identify where additional education is needed, further research is needed to identify to what extent clinicians are using clinical genetics skills and services in their own practice. This study attempts to explore new physicians' usage of core clinical genetics skills, their referral patterns to medical genetics specialists, and their experiences in medical genetics education.Residents from the University of Maryland Medical Center (n=57) were surveyed regarding these topics. In regards to clinical genetics skills, pedigrees are rarely taken and approximately 40% of residents rarely or never engange in genetic counseling when confronted with a genetic indication. Approximately half of residents have ordered genetic tests in the past, but it was largely indicated that they are not confident in their knowledge of the tests they are ordering. Overall we found that educational exposures directly involving genetics specialists or specifically revolving around clinical genetics skills were associated with a higher frequency of engagement in genetic counseling. Higher numbers of exposures, as well as those involving genetics professionals, were associated with higher feelings of adequacy with medical school genetics education. We also found that educational interventions that continue on throughout residency may be helpful in encouraging the utilization of clinical genetics skills. Residents who received lectures about genetic testing during residency were twice as likely to order genetic testing than those who did not receive such lectures, and the exposures most associated with referrals to medical genetics specialists occurred during residency. This pilot study provides a broad overview of clinical genetics practices of this cohort of new doctors. Encouragingly, the vast majority of respondents utilize genetic counselors and medical geneticists. However, further education should be provided to physicians to give them the confidence and skills they need to manage genetics patients when medical genetics specialists are not available.

3103W

Genomics in family medicine: Views of community-based providers. *S.B. Trinidad[1], S.M. Fullerton[1,2], L.M. Amendola[3], W. Burke[1,3], G.P. Jarvik[2,3].* 1) Bioethics & Humanities, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA.

Introduction: As next-generation sequencing technologies move from research settings into the clinic, primary care providers (PCPs) will often be on the front lines of explaining and managing results. Previous research has reported limited PCP knowledge of genetic concepts. The growing number of patients tested, uncertainty associated with incomplete penetrance and variants of uncertain significance, and the potential for additional, non-solicited results add to the challenge. This study explored the views of community-based PCPs with respect to the implementation of genome-scale testing in primary care. **Methods**: The University of Washington NEXT Medicine Study of the Clinical Sequencing Exploratory Research (CSER) consortium conducted semi-structured telephone interviews with PCPs practicing in community settings. Topics included views about genome-scale testing and what kinds of additional findings would be useful in primary care. **Results**: We interviewed 19 family medicine physicians (58% female; 89% non-Hispanic white; mean age 44; 13 mean years in practice) affiliated with the WWAMI (WA, WY, AK, MT, ID) region Practice & Research Network. Most (18/19) participants reported being "somewhat familiar" with genetics. Most (N=13) considered pharmacogenetic results to be most useful and common disease risk least useful (N=11); opinions were mixed regarding risk for rare treatable disease (4 reported most useful, 4 reported least). Although they expected genetic specialists to order tests and provide initial counseling, most felt that PCPs would be better positioned than specialists to help patients understand and use genetic information. Participants outlined a unique role for family medicine PCPs in educating patients, clarifying relevant values, and identifying utility. They believed that their emphasis on family history and disease prevention, their long-term relationships with patients and families, and their skills in patient-centered communication and shared decision-making enable them to frame genetic results in ways that are meaningful to patients. **Conclusions**: Family medicine PCPs are interested in helping patients and families comprehend and use genome-scale results. Further research should explore opportunities for team medicine and a rational division of labor with respect to genomics in family medicine; training and education for PCPs and their staff in genetics and risk communication; and clinical decision support.

3104T

Genetics and health sciences education pipeline for elementary, middle, high school (K-12) and undergraduate underrepresented minorities (URM) in the sciences at the Louisiana State University Health Sciences Center (LSUHSC). *F. Tsien[1], M. Moore[1,2], A. Umrigar[1], A. Augustus-Wallace[3], D. Mercer[1].* 1) Dept of Genetics, School of Medicine, LSU Health Sciences Center, New Orleans, LA; 2) Dept of Biostatistics, School of Public Health, LSU Health Sciences Center, New Orleans, LA; 3) Division of Diversity and Community Engagement, School of Medicine, LSU Health Sciences Center, New Orleans, LA.

   The Louisiana State University Health Sciences Center in New Orleans (LSUHSC) has created a science and genetics educational pipeline through the Science Youth Initiative (SYI) and the Research Experiences for Underrepresented minorities (REU) programs. The long term goals are to ultimately increase the number of Underrepresented Minorities (URM) in the sciences, especially in genetics. The SYI program: The hands-on experiments, workshops, and educational videos are coordinated with elementary, middle, and high school (K-12) teachers to maximize learning and complement their curriculum. Since 2009, the SYI has reached > 2100 K-12 students. In Orleans parish, we reached 1365 K-12 students, of which 68% are URMs. Diverse LSUHSC trainees from the Schools of Medicine, Graduate Studies, Public Health, and Allied Health help educate the K-12 students. 1) Goals of the SYI: improve science academic achievements, introduce diverse role models in the sciences, increase awareness of genetics and science careers, and provide LSUHSC trainees instruction in communication skills to the lay public. Topics range from strawberry DNA isolation (elementary schools), to forensics and HeLa cells (high schools). The REU program: Since 2014, the NSF-funded REU program has provided internships to 10-13 URM high school and undergraduate students/year. General topics include population genetics, molecular and cellular biology, cancer research, chromatin instability, mutation analysis, translational research, personalized medicine, and bioinformatics. 2.) Goals of the REU program: develop scientific problem solving skills, learn state-of-the-art laboratory techniques, develop scientific communication skills, and receive resume writing instruction for competitive application to medical or graduate school. Following the REU program, most students co-author scientific publications and conference presentations, and/or obtain part-time laboratory employment during academic year. Participating schools include rural, urban, public/charter, private schools, and historically black colleges and universities (HBCUs). Assessments are by qualitative and quantitative participant surveys, mentor evaluations, and by using the NSF-REU common assessment tool. Students are tracked via an on-line NSF reporting system. These programs have provided opportunities to K-12 and undergraduates interested in the sciences, especially genetics, and helping them achieve their career goals.

3105F

The Genomics ADvISER - A Genomics decision AiD about Incidental SEquencing Results. *Y. Bombard[1,2], M. Clausen[2], E. Glogowski[3], K. Schrader[4], M. Robson[5], A. Scheer[1,2], M. Evans[1,2], J. Hamilton[5], K. Offit[5], N. Baxter[1,2], K. Thorpe[1], L. Lerner-Ellis[1,6], A. Laupacis[1,2].* 1) University of Toronto, Toronto, Ontario, Canada; 2) Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, Canada; 3) GeneDx, Gaithersburg, Maryland, USA; 4) BC Cancer Agency, Vancouver, British Columbia, Canada; 5) Memorial Sloan Kettering Cancer Center, New York, NY, USA; 6) Mount Sinai Hospital, Toronto, Ontario, Canada.

   Background: Guidelines recommend doctors inform patients of their incidental genomic sequencing (GS) results, yet there are limited tools to support patients' decisions about which results they wish to learn. **Aims:** To develop & evaluate a decision aid (DA) to guide patients' selection of incidental results. **Methods: (1)** Development: We used the Ottawa Decision Support Framework to develop a DA. Incidental results are categorized based on Berg *et al's* binning scheme. Category 1=medically actionable variants; Category 2= common disease SNPs; Category 3= Mendelian disease variants; Category 4=early-onset neurological variants; Category 5=carrier results. **(2)** Evaluation: We used cognitive task analysis, interviews and surveys to evaluate the usability, validity & acceptability of our DA. Cancer patients were recruited through clinics. Analyses include content analysis & descriptive statistics. **Results: (1)** We created an interactive online DA. It begins with an animated video on GS, the categories of incidental results & their harms/benefits, and then engages participants in a values clarification exercise. After providing summaries of participants' preferences, it tests their understanding of the categories, with correct answers provided after. It ends by asking participants to select the categories of incidental results they want to learn. **(2)** We evaluated the DA's usability with 11 patients in 2 rounds; the majority were <60 yrs old (6/11), female (7/11), had high levels of health literacy (6/11) & education (6/11) but low computer literacy (7/11). Interviews demonstrated strong face validity & content comprehension. Patients could describe the categories well & liked their organization but had trouble recalling which number corresponds with each category. Most patients found the amount of information 'just right' (8/11), clear (9/11) & balanced (10/11). All patients felt that the information was sufficient to reach a decision; that the DA was easy to use & would recommend it. We modified our DA to aid patients' recall of categories & are re-testing it. **Conclusion:** We created and evaluated an innovative patient-centered tool to support the clinical delivery of incidental GS results. Our DA fills critical care gaps given limited genomics expertise & capacity to convey the large volume of incidental GS results & their myriad of implications. Our DA may ultimately improve the quality of patients' decisions, save clinicians counselling time & health care costs.

3106W

P4'' medicine as a social phenomenon. *C. Bouffard[1], G. Lapointe[1], R. Drouin[2,3].* 1) Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) Université Laval, CHU de Québec, Québec, Canada; 3) Université du Québec à Montréal, Montréal, Canada.

Just a few decades ago, Personalized Medicine (PM) was barely a pharmacogenomics research & development project, based on the idea that a medication's effects vary in accordance with the individual genetic profile. Subsequently, precision and genomic medicine developed under the term PM. Nowadays, in the synergy of the academic, medical, industrial and political institutions, PM is on the way to becoming "P4" medicine. In this context, several governments, in partnership with the private and parapublic sectors have made PM a research and development priority in order to improve both the wealth and health of individuals and the overall economy.   However, the success of this project implies raising technological, clinical, economic, organizational, ethical, legal and social challenges for which we are not prepared. P4 medicine is not limited to genomic technologies or to a change in medical paradigm, it is rather a social project, solidly structured and financed and involves important changes in medical practices, explanatory models of the disease, economic and social behaviours; meaning of responsibility in terms of health and well-being and health products and services regulation. Despite the expected benefits, P4 medicine leads to clinical, social and ethical issues that we have to be able to identify. In order to be aligned with coherent long-lasting solutions, in-depth knowledge of its development and institutionalization conditions are required as well as knowing what interests are at stake. We need to have adequate approaches to anticipate and respond effectively to the changes that the P4 medicine will bring about. Nonetheless, it is firstly necessary to 1) establish a clear definition of PM, 2) recognize it as a social phenomenon and, 3) develop functional inter- and transdisciplinary collaborative approaches.

3107T

Transethnic genetics: Implications for precision medicine and the study of complex traits across diverse populations. *B.C. Brown[1], N. Zaitlen[2].* 1) Department of Computer Science, University of California Berkeley, Berkeley, CA; 2) Department of Medicine, University of California San Francisco, San Francisco, CA.

Complex disease phenotypes vary in their global distributions due to a combination of genetic and environmental differences, but the mechanisms underlying these differences remain largely unknown. While the majority of genetic studies of disease have been conducted in European populations, the number of multi-ethnic disease cohorts is increasing. This provides a new opportunity to compare complex trait genetics across populations. Some researchers have argued that disease genetics are conserved between populations, because GWAS hits frequently replicate. In recent work, we examined this hypothesis by developing a novel statistical method to estimate the correlation of per-SNP causal variant effect sizes (ρ_{ge}) between populations. This correlation is 1.0 when SNP effect sizes are identical in both populations. We found that ρ_{ge} between European and East Asian cohorts is only 0.46 (0.05) for rheumatoid arthritis and only 0.62 (0.08) for type-2 diabetes. Similarly, we found that the mean ρ_{ge} of the genetic effects influencing gene expression between Yoruban and European cohorts is only 0.32 (0.01). Here, we explore the implications of these discoveries for clinical practice and complex disease research in diverse populations. We argue that while there may be many SNPs with statistically significant effects in both populations, the magnitudes of these effects can differ dramatically on average. We focus on the case of using complete knowledge of the effects sizes in one population to predict disease risk in another population, rigorously quantifying the area under the receiver operating characteristic curve as a function of ρ_{ge}. If a goal of precision medicine is to implement accurate predictors of disease risk into clinical practice, then our results demonstrate that the failure to study diverse populations is a gross inequity, because classifiers based on European studies will not be accurate enough to be used in non-Europeans. Thus, in the current paradigm people of European descent will disproportionately reap the benefits of precision medicine. Finally, we discuss the genetic mechanisms that could result in a genetic correlation less than 1.0, including differential tagging, rare variant and non-additive effects. In light of this we argue that understanding the causes of transethnic disease differences is required to fully understand complex trait genetics. All together, this necessitates increased genetic studies of disease in non-European populations.

3108F

Attitudes regarding personal genome sequencing among healthy early adopters: Findings from the PeopleSeq Consortium. *E.D. Esplin[1], E. Haverfield[1], D.E. Nielsen[2,3,4], E.F. Schonman[2,3], E. Ramos[5], M. McGinniss[5], S. Tucker[6], D.G.B. Leonard[7,8], R.C. Green[2,3,4,9] for the PeopleSeq Consortium.* 1) Invitae, San Francisco, CA; 2) Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Illumina, Inc., San Diego, CA; 6) Tucker Medical, Singapore, Singapore; 7) Department of Pathology and Laboratory Medicine, University of Vermont College of Medicine , Burlington, VT; 8) University of Vermont Health Network, Burlington, VT; 9) Partners Personalized Medicine, Boston, MA.

Background: With growing opportunities for healthy adults to access personal genome sequencing (PGS), it is increasingly important to understand their attitudes about and intended applications for PGS. The PeopleSeq Consortium is surveying ostensibly healthy individuals who have undergone PGS through commercial or research avenues to understand their attitudes toward and expectations for PGS. Methods: Respondents included in this analysis purchased PGS through Illumina's Understand Your Genome® program, which provides a clinically focused report to participants through an ordering physician, followed by access to their raw genome sequence introduced at an educational symposium. Survey items examine decision-making about pursuing PGS and general attitudes regarding PGS. Results: Respondents (n = 237) were on average 53 years old, white (90%), and highly educated (61% doctorate or professional degree), with high household income (89% ≥ $100,000). Most were married (52%), male (58%), and had children (76%), and 41% were healthcare providers or clinical researchers. The majority (54%) indicated that it took them one day or less to decide to pursue PGS. The most common reason for pursuing PGS was "finding out about personal disease risk" (23%, n = 150). Of the 52% who reported talking with someone when making their decision, most (81%) reported speaking with family members. Fifty-two percent agreed that health insurance should cover PGS, and 62% agreed that personal genomic information should be part of a standard medical record. Thirty-eight percent agreed that the government should put more effort into regulating PGS. Conclusion: The majority of individuals made the decision to pursue PGS quickly and many did so without meaningful medical consultation or counselling. A minority favored increased government regulation of PGS, possibly reflecting the libertarian perspectives of these early adopters. A majority agreed that there should be insurance coverage of PGS for healthy individuals and supported incorporating genomic information into the medical record. The future goals of the PeopleSeq Consortium are to assess medical, behavioural and economic outcomes by surveying additional varied PGS cohorts that return genomic information to ostensibly healthy individuals.   .

3109W

Genomic patents, past and present: Empirical findings, challenges, and implications. *C.J. Guerrini[1], A.L. McGuire[1], M.A. Majumder[1], R. Cook-Deegan[2].* 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Sanford School of Public Policy, Duke University, Durham, NC.

In 2013, the U.S. Supreme Court determined that naturally occurring DNA cannot be patented because it is a product of nature that is categorically excluded from patenting. Meanwhile, in a succession of contemporaneous cases, the Court articulated a test for evaluating the patent eligibility of innovations that has been criticized as so manipulable as to be almost meaningless. The conventional wisdom that has emerged in the wake of this modern legal history is that the ability to patent gene-related compositions and methods for identifying and manipulating genetic information has been severely limited. With funding from the National Human Genome Research Institute, we set out to determine whether this understanding is empirically justified. Utilizing a mixed-methods approach, first, we analyzed the file histories of over 300 gene-related patents to identify challenges to genomic patenting resulting from the application of evolving legal rules and professional practices. Second, we conducted semi-structured interviews of over 25 leading genomic innovators, patent lawyers, and patent scholars to obtain first-hand knowledge of legal, policy, and business strategies and trends relevant to genomic patenting. During this session, we will report baseline quantitative and qualitative findings and discuss the implications of our data for genomic innovation and investment. Our findings include evidence that patent protection of genomic innovations has become less robust and more costly in recent years as well as divided expert opinions about the practical and ethical implications of the shifting legal landscape on genomic research and clinical practice.

3110T

Determining critical communication issues for the genetic testing process. *M.L.G. Hallquist[1], A.H. Buchanan[1], W.A. Faucett[1], M. Cho[5], K. Brothers[2], C.R. Coughlin II[3], L. Hercher[4], L. Hudgins[5], S. Jamal[6], H. Levy[7], H. Peay[8], M. Roche[9], M. Stosic[10], M. Smith[11], W. Uhlmann[12], K. Wain[1], K.E. Ormond[5].* 1) Genomic Medicine Institute, Gesisinger Health System, Danville, PA; 2) University of Louisville, Louisville, KY; 3) University of Colorado, Denver, CO; 4) Sarah Lawrence College, New York, NY; 5) Stanford University, Palo Alto, CA; 6) Hospital for Sick Children, Genome Diagnostics Lab, Toronto, ON, Canada; 7) Johns Hopkins University, Baltimore, MD; 8) RTI International, Washington, DC; 9) University of North Carolina, Durham, NC; 10) Natera Inc. San Carlos, CA; 11) Northwestern University, Evanston, IL; 12) University of Michigan, Ann Arbor, MI.

　Given the lack of genetic specialists world-wide and an increasing use of genomics across many medical specialties, genetic testing is increasingly being ordered by clinicians without training in genetics. These clinicians often have limited experience with the complex ethical, legal, social, medical, and logistical issues (ELSIPlus) associated with genetic testing. ClinGen's Consent and Disclosure Recommendations (CADRe) workgroup is developing tools to improve communication and patient experience of genetic testing. We have developed a rubric to identify an appropriate communication strategy at the time of consent based on the nature of the gene in question and empiric evidence about such testing. Where possible, we depart from the familiar paradigm of pre- and post-test counseling in deference to the need for a stream-lined process. We are using a mixed-methods design of surveys and focus groups to conduct a formative evaluation to assess responses to the rubric from genetics clinicians, non-genetics clinicians, and individuals who have undergone genetic testing. Using snowball recruitment for an anonymous online survey, 60 genetics professionals reacted to the CADRe-recommended communication approaches for common testing indications for five genes. For all genes surveyed, the majority agreed with the CADRe rubric recommendation for the indications of diagnostic testing, confirmatory testing, and familial variant testing. When testing an *unaffected consultand* with an unavailable proband, respondents recommended an *increased* level of communication due to concerns about residual risk, communication with family members, and ensuring patient understanding of benefits and limitations. To date, we have conducted four focus groups, and will complete five more by early fall. Participants in the non-genetic provider and patient focus groups thought most genetic testing discussions should begin with a trusted clinician (e.g., primary care provider) and that the conversation should be tailored to the interests of the patient, with additional discussion, referral to genetic counseling, or educational materials as appropriate. Genetics providers stressed the importance of their involvement in return of results. Results of the survey and focus groups provide formative input to guide the continued development of the rubric and associated materials as we aim to address ELSIPlus considerations of genetic testing ordered by a diverse group of clinicians.

3111F

Views of patients and their parents regarding the incidental or secondary findings obtained from whole exome and genome sequencing: A literature review. *A. Inaba[1], E. Takamine[1], M. Matsukawa[1], A. Yonei[1], S. Kosugi[1,2].* 1) Kyoto University, Kyoto, Japan, Kyoto, Japan; 2) Clinical Genetics Unit, Kyoto University School of Public Health, Kyoto, Japan.

　[Background] Whole exome sequencing (WES) and Whole genome sequencing (WGS) with next-generation sequencing technology have started to be used as a clinical diagnosis. These sequencing technologies generate a large amount of data including both primary results and incidental or secondary findings (IF/SF). There is no general consensus on how to manage IF/SF in Japan. Therefore, it is necessary to discuss how IF/SF by WES/WGS should be treated in research or clinical settings. **[Purpose]** The purpose of this literature review is to understand and analyze views of patients or their parents regarding the results particularly on IF/SF obtained from WES/WGS. **[Method]** We searched original articles dealing with IF/SF by WES/WGS on online databases (PubMed, Web of science, ICHUSHI web) which were published from 2010.11 to 2016.5. **[Results]** Twelve articles among a total of 261 met our criteria. Five articles of the twelve investigated patients or their parents who have undergone WES/WGS. The majority of the patients and parents desired to receive not only the primary results but also IF/SF particularly related to diseases which are treatable or preventable. They felt responsible for managing their own or their child's health conditions. However, some did not want to know IF/SF because it is a psychological burden for them. It seems that patients and parents tend to consider benefits more importantly than the risks of receiving IF/SF. Some patients and parents desired to choose which findings should be returned. **[Conclusion]** This review suggested that additional support before and after WES/WGS regarding IF/SF is required. We will continue to discuss how to manage the results of WES/WGS in clinical setting in Japan.

3112W

Expert's opinion on child's capacity for informed assent in genomic research: A focus group study. *I. Ishiyama[1], Z. Yamagata[2].* 1) Tokoha University, Fuji, Shizuoka, Japan; 2) University of Yamanashi, Chuo, Yamanashi, Japan.

[Background] Current opinion on the ELSI concerning genetic testing in children by the ASHG points that children should be included in the informed-assent or -consent process to the extent that they are capable, in certain approaches. Previously, we conducted a survey and clarified the attitude of the Japanese general public on children's capacity for informed assent in genomic studies. While the answers were incoherent, people with high level of scientific literacy could express their views and tended to answer that younger children are capable for assent. A qualitative study focusing on experts with high level of literacy was required for seeking good process of assent. In this study, we aimed to explore the views of experts engaged in a medical study on child's capacity for assent in genomic researches. [Methods] A semi-structured focus group interview was conducted in March 4, 2016. The study participants were comprised of 6 female research coordinators (age, 30-49) who engaged in a birth cohort study in Japan (years of experience, 3-5). They all had a nurse qualification, 1 had a public health nurse and 1 had a school counselor. We asked to talk about their views on obtaining informed assent, capable child's age for assent, and how to assess the child's capacity for assent. Particular attention was given to the interview, such as voluntary consent and participant's anonymity. The two-hour group discussion was audio-taped, transcribed and thematically analyzed. [Results] A thematic analysis showed the following six themes: 1) Recognition of importance at the same time difficulty of obtaining informed assent; 2) Perception of the potential risk to decrease in follow-up rate due to children's valid objections; 3) Expectation of the relation between child's capacity for assent and learning social studies as well as science studies in school; 4) Prediction of the association between child's capacity for assent and psychological independence; 5) Objection to assess a child's capacity before obtaining assent by using the development inspections; 6) Intention to deal with obtaining assent, if needed, in an intelligible language tailored to the smaller age children. [Discussions] Experts emphasized the individual differences of children and disagreed to use the development inspections. It was indicated that a child's capacity for assent might be assessed by a measure individualized for each research project, and specialized for the typical research contents.

3113T

Clinician views of public health newborn screening using next generation sequencing technologies: Implications for policy. *G. Joseph[1], F. Chen[2,3], C.F. Young[4], B.A. Koenig[2,3].* 1) Anthropology, History, and Social Medicine, UCSF, San Francisco, CA; 2) Institute for Health and Aging, UCSF, San Francisco, CA; 3) Institute for Human Genetics, UCSF, San Francisco, CA; 4) California State University, Stanislaus.

Advances in genomic technologies, such as whole genome sequencing (WGS), present both practical and ethical challenges to traditional public health newborn screening (NBS). While broad debate exists among clinicians surrounding how, when, and even if to return findings from clinical WGS, little research has investigated clinicians' perspectives on the potential expansion of public health NBS with WGS. As part of a larger study of stakeholder views on NBS with WGS, we conducted three focus groups with 23 clinicians (physicians, certified nurse midwives, and nurse practitioners) representing specialties traditionally involved in NBS (neonatology, obstetrician/gynecology, pediatrics, midwifery, maternal fetal medicine, pediatric bone marrow transplantation, and pulmonary care). Our design included clinicians practicing in both the prenatal and postnatal settings. Discussion topics explored clinician perspectives on: 1) current NBS; 2) the implications of expanded NBS for a) return of results; b) consent practices and c) storage and use of results for research and NBS program quality improvement; and 3) policy recommendations. Audio recordings of the focus groups were transcribed verbatim and analyzed for thematic content. Three key themes emerged: 1) contested clinical responsibility with respect to patient education and the interpretation of NBS results; 2) perceived tension between the practical implementation of WGS NBS (informed consent, return of results, costs for follow-up care) and the obligation under public health policy to balance benefits and burdens of population-wide screening; and 3) the need for greater interdisciplinary collaboration, specifically with genetic counselors, to interpret and return genomic results. Participants expressed enthusiasm about some types of information that can be gleaned from WGS NBS data, such as actionable pharmacogenetic variants. This enthusiasm was tempered however, by clinicians' experience of feeling overwhelmed with current responsibilities and their belief that WGS NBS should not be implemented without additional informed decision-making and parental consent processes. Importantly, this study contributes to an understanding of how clinicians across practice specialties interact with parents and each other regarding NBS. Clarifying stakeholders' goals and values with regard to public health NBS is critical to developing policies needed to appropriately incorporate WGS into NBS.

3114F

Gene editing technologies: What's new in policy. *E. Kleiderman[1], R. Isasi[2], B.M. Knoppers[1].* 1) Centre of Genomics and Policy, McGill University, Montreal, Quebec, Canada; 2) Dr. John T. Macdonald Foundation, Department of Human Genetics, Institute for Bioethics and Health Policy, University of Miami, Miami, FL, USA.

With the arrival of CRISPR, human genome editing is easier and more accessible than ever before, offering both potential for scientific and therapeutic advancement, while bringing ethical, legal and social (ELSI) tensions back to the forefront. Similar apprehensions have been expressed with recombinant DNA technology, embryonic stem cell research, human cloning and more recently, preimplantation genetic diagnosis (PGD) and mitochondrial replacement therapy (MRT). The purpose of our study is to assess and compare the current regulatory and policy frameworks and issues surrounding human germline editing, PGD and MRT. Against the backdrop of a published 16 country study (Science 2016; 351:337-339), we conducted a more in-depth comparative analysis of the regulatory frameworks encompassing gene editing technologies in four countries: Canada, France, the United Kingdom (UK) and the United States (US). These countries were selected, as they represent a spectrum of approaches to the different technologies being compared, and all have recent regulatory/legislative developments (2015-2016). We highlight and address key similarities and differences between the three technologies within the regulatory frameworks of each country, such as varying levels of acceptance and regulation, common concerns, and overall limits and thresholds. Furthermore, we pinpoint the overarching ELSI questions and suggest solutions for tackling the discussion of risks associated with editing the germline. We support a cautious approach that inclusively engages stakeholders in discussion. These findings support the idea that PGD and MRT can be used as potential models for approaching the regulation of human germline genome editing and help address society's concerns about the coming of an era of slippery slopes, 'designer babies' and playing God.

3115W

Content analysis of third party interpretation tools for direct-to-consumer genetic testing data. *S.C. Nelson[1,2], S.M. Fullerton[3].* 1) Public Health Genetics, University of Washington, Seattle, WA; 2) Biostatistics, University of Washington, Seattle, WA; 3) Bioethics and Humanities, University of Washington, Seattle, WA.

Background: Direct-to-consumer (DTC) genetic testing gives people access to interpreted reports based on their genetic data and the ability to directly download their (uninterpreted) genotype data file (GDF). Many third party tools and websites have been created where DTC customers can pursue additional, self-directed interpretation of their GDF. It is important for the genetics community to understand why and how consumers are using their genotype data in DTC/third party systems, as it may provide insight into what will happen as additional routes to personal genetic data access open up. For example, recent regulatory changes (45 C.F.R. § 164.524) allow individuals to directly access their full record from HIPAA-covered laboratories, which for clinical sequencing tests could include vast amounts of uninterpreted sequence data. In addition, some participants and researchers have expressed support for giving ("raw") genotype data back to participants. **Methods:** We conducted a structured content analysis of over a dozen online third party interpretation tools, defined as tools/sites where users can upload their GDF and receive some type of information/interpretation. Tools were identified from a variety of sources, including media stories, blog posts, peer reviewed scientific literature, conference presentations, and personal sources. We tracked features such as types of information returned (ancestry, health, relatedness); who created the tool and why; ability to contribute to research; and privacy and security measures. **Results:** We found that most tools were created by either academic groups, for-profit companies, or citizen science genealogists. Many sites focus on returning genealogy and ancestry information (e.g., GEDMatch and GenomeMatePro). A few provide some health-related interpretation (e.g., Promethease, openSNP, and Interpretome). For several platforms, a prominent feature is the ability to share or donate data, either to other users (e.g., openSNP) or to the scientists who run the site (e.g., DNA.land). **Discussion:** The variety of third party tools suggests a wide range of motivations among creators and users, in addition to a spectrum of validity for the information returned. Future research will explore how and why DTC customers use these tools, with the goal of gaining insights into how expanding access to genotype data from research and medicine may unfold.

3116T

Adolescent attitudes towards genetic testing for adult-onset conditions from essays submitted to a national competition. *A. Rahm[1], L. Bailey[1], O. D'Accordo[2], Y. Munishor[2], C. Miller[2], E. Davidson[2], B. Wolf[3], K. Pulliam[3], H. Zhang[3], M. Dougherty[3], L. Hercher[2], M. Williams[1].* 1) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Human Genetics, Sarah Lawrence College, Broxville, NY; 3) American Society of Human Genetics, Bethesda, MD.

 Background: Genetic testing has become increasingly popular, however, questions and concerns surround when to test children for adult-onset conditions. A position statement from the ASHG published in 2015 reaffirmed its original position that testing for adult-onset conditions should be deferred until adulthood. Information is lacking on the attitudes and opinions of adolescents themselves about the subject. The 2016 ASHG DNA Day Essay Contest asked 9-12th graders to choose an adult-onset genetic condition and defend or refute the ASHG position statement. **Methods:** Essays submitted to ASHG's 2016 national DNA Day Essay Contest by teachers on behalf of students underwent thematic analysis. Teachers could submit up to 6 student essays per class for up to 3 classes. Participants were informed that essays may be used for research purposes and that winning essays may be published online. This study was reviewed by IRB and determined to be exempt. All essays submitted by students were analyzed for the condition chosen, agreement or disagreement with the ASHG statement, and reasons for their position. Demographic information including gender, grade, school type, and state were also analyzed. **Results:** 1241 student essays were submitted from 44 US states (87%) and other countries (13%). More females submitted essays than males (67% vs. 33%). Over 100 adult-onset conditions were discussed across all essays, although 60 conditions were each discussed only once. The most commonly discussed conditions were Huntington's disease (N=466 essays, 38%), followed by BRCA-related breast or ovarian cancer (N=205, 17%), and Alzheimer's disease (N=118, 10%). Students were evenly split whether they agreed or disagreed with the ASHG recommendation, however, slightly more agreed to defer testing for Alzheimer's (59%) or Huntington's (57%) compared with BRCA (42%). Reasons for deferring testing included potential psychological harm and anxiety, while reasons for wanting testing included prevention and life planning. Factors such as family history and personal anxiety were given as reasons for the need for choice. Additional analyses are ongoing and will be presented. **Conclusions:** The DNA Day essays provide valuable insights into the attitudes of adolescents regarding genetic testing for adult-onset conditions. Additional research is needed to understand the psychosocial impact of learning genetic information for adult onset-conditions in healthy adolescents.

3117F

Measures of self-reported outcomes from receiving genetic results among underserved racial and ethnic populations: A systematic review. *A.E. Tattersall, J-H. Yu.* Department of Pediatrics, University of Washington, Seattle, WA.

 The majority of translational research on return of genome sequencing results has focused on European Americans affected by rare genetic conditions. As studies of clinical translation of genome sequencing are expanded to include and focus on underserved populations including racial and ethnic populations, establishing the relevance of outcomes currently under investigation, the applicability of existing self-reported measures, gaps in available measures and their measurement properties will be important to ultimately assess equitable benefit. To this end, we are conducting a systematic literature review of studies that measure self-reported outcomes from receiving either genetic test results or research results among patients and participants from underserved racial and ethnic populations. Abstracts of potential primary research articles were identified from MEDLINE, PsychInfo, and CINHAL using MeSH search terms to capture parameters of interest including genetic information, outcomes, and populations. A total of 1332 unique references have been compiled and their abstracts are being reviewed to identify articles that meet the following criteria: English language, published between January 1, 1986 and February 24, 2016, and report quantitative measures. Based on abstract review, approximately 15% will meet initial inclusion criteria, while the remaining either do not constitute a research article or fail to report an outcome of genetic results disclosure. After potential articles are reviewed, the final set will be coded for their studied measures (e.g., definition, number of items, response method, type of measure), the context of their use (e.g., study, population, mode, languages, population comparisons), and measurement properties (e.g., validity and reliability). This analysis will inform availability of measures for particular domains, their evidence of validity in specific populations, and degree to which differences between populations have been observed and are to be expected. This systematic review will also propose recommendations for future measures development and related outcomes research.

3118W

Laboratory directors' opinions regarding the FDA's proposed regulatory oversight of Laboratory-Developed Tests (LDTs). *A. Taylor[1], K. Johansen Taber[2], E. Leeth[3], K. Leuer[3], C. Wicklund[1].* 1) Northwestern University, Chicago, IL; 2) American Medical Association, Chicago, IL; 3) Lurie Children's Hospital, Chicago, IL.

Background: In October 2014 the FDA released a draft guidance defining its framework for regulatory oversight of Laboratory Developed Tests (LDT's). The introduction of the FDA's draft guidance generated considerable debate with several medical and laboratory organizations releasing formal statements in response. Since laboratory directors are a major stakeholder in this discussion, and will be considerably affected by any change of LDT oversight, we interviewed laboratory directors in a structured manner to provide their insights on the debate. *Methods*: Participants were drawn from a variety of laboratory settings including academic institutions, private companies and hospital based laboratories. A semi-structured interview guide was developed to explore director's opinions on current oversight of LDTs, the FDA's draft guidance and suggestions to modify oversight. The guide was piloted with a senior laboratory member of Lurie Children's Hospital in Chicago, to ensure we were obtaining the desired information. No changes were made to the interview guide. Interviews were conducted over the telephone, recorded and transcribed verbatim. Transcripts were then coded for themes. *Results*: Twelve interviews were conducted. Twenty themes were identified and organized into five categories; concerns with the draft guidance, opinions on current oversight, recommended modifications to oversight, considerations with developing oversight and opinions of the FDA as an organization. Most notably, participants were concerned about the increased laboratory burden associated with adhering to the FDA's draft guidance. They were worried about the unintended consequences such as closure of laboratories and the negative impact on patient care. Participants want the FDA to release an updated draft guidance with more detail before it is finalized. However, they would prefer to see a modernization of current oversight and an increase in resources for the relevant governing bodies such as CLIA and CAP. *Conclusion*: The results of this study contribute to a better understanding of the concerns held by laboratory directors regarding not only the FDA's draft guidance but oversight of LDTs in general. Several themes emerged in our data that should be considered by any governing body developing oversight for LDTs. Discussion between the FDA and professional laboratory organizations is warranted as many of the concerns held by our participants will have an impact on patient care.

3119T

Minorities" in precision medicine: Perspectives from federally qualified health centers and community-based organizations. *J.-H. Yu, A.E. Tattersall.* Pediatrics, University of Washington, Seattle, WA.

Efforts to build large precision medicine (PM) cohort studies are motivated, in part, by an ethic of social justice, inclusion and partnership with underrepresented and underserved communities. Yet, in addition to the promise of benefits through genomic medicine, inclusion of these communities into genomic studies presents potential for harm such as group stigma and discrimination. As precision medicine researchers engage federally qualified health centers (FQHC) and community-based organizations (CBO) to recruit underserved populations, how to transparently present and weigh such potential benefits and harms to patients and members is an ethical and practical challenge for FQHC providers and CBO leaders and staff. Based on a developing framework for considering benefits and harms (GBH), we have and are continuing to conduct interviews (n=23) with FQHC medical directors and CBO executive directors about the feasibility of recruiting their patients and members for a national PM cohort study, attitudes toward genomic research participation, and potential risks and benefits of genome sequencing including return of incidental findings. Many participants viewed PM research through a social justice lens, expressed concerns about contributing to group harms such as racial discrimination and stigma, and expected community- and individual-level engagement by researchers about study goals and methods. While all recognized great potential for PM-tailored treatments, most were highly skeptical about their patients' or members' abilities to benefit from genomic results in practice and offered both barriers and potential facilitators. Interviews further clarified the important role FQHC-providers and CBO-staff play in helping patients/community members learn about and weigh potential benefits and harms, and making participation decisions in the context of individuals' health issues and community health priorities. Healthcare providers and community leaders of underserved populations are increasingly partnering with PM researchers and playing a key role in framing the potential benefits and harms of research participation. Relationships between providers-patients and community leaders-members offer unique opportunities to support decisions about genomic and precision medicine research participation. Their perspectives will be important to ensuring that the benefits of genomic and precision medicine are equitably realized by underserved and underrepresented communities.

3120F

When to take a family history: Assessing impact on genetic counseling outcomes. *C. Slomp[1,2], E. Morris[1,2], A. Inglis[1,2], A. Lehman[1], J. Austin[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada.

   BACKGROUND: Documentation of family history is a cornerstone of genetic counseling (GC) practice. Various strategies for obtaining family histories have evolved (e.g. pre-GC appointment online questionnaires), often to increase the efficiency of generating a product for assessing genetic risk and informing genetic testing. However, the process of patient counsellor interaction to document a family history can be useful in building rapport, which is key to optimal patient outcomes of GC. No studies have evaluated the effect of the timing or method of obtaining family history on patient outcomes of GC. HYPOTHESES: Patients whose family history was obtained by phone prior to GC (Group 1) would have greater increases in a) empowerment and b) self-efficacy than patients whose family history was collected during the GC appointment (Group 2). METHODS: We conducted a retrospective chart review in a specialist psychiatric GC clinic, using data from Feb 2012-March 2016. Family history is either documented by phone prior to GC, or at the beginning of the GC session, and patient outcomes are routinely tracked: pre-GC, and 1 month after GC empowerment is measured with the Genetic Counseling Outcome Scale (GCOS) and self-efficacy is measured by the Illness Management Self Efficacy scale (IMSES). Patients who completed GCOS and IMSES at both time points were included in the analyses (independent T-Tests and Mann-Whitney tests comparing change in GCOS and IMSES scores from pre- to post-GC between Groups 1 and 2). RESULTS: 130 patients completed scales at both time points, and were included in analyses. There were significant increases in GCOS and IMSES scores from pre- to post- GC but no difference between groups in the magnitude of change in score. CONCLUSIONS: This study suggests that obtaining family history prior to versus during the GC appointment results in similar patient outcomes for empowerment and self-efficacy. Future research could compare in person and web-based family history documentation.

3121W

Can our Manga cartoon, as a medium for promoting family health history and human genetics to the public, cause anxiety of learning their own family health history? *E. Takamine[1], Y. Hiraoka[2], M. Torishima[1], A. Inaba[1], M. Matsukawa[1], A. Yonei[1], Y. Tabara[3], T. Wada[1], S. Kosugi[1].* 1) Department of Medical Ethics and Medical Genetics, Kyoto University Graduate School, Kyoto, Kyoto, Japan; 2) Riken; 3) Center for Genomic Medicine, Kyoto University Graduate School.

   [Background] Today, genetic technology has been rapidly advancing and genetic testing is largely available in medical practice. In Japan, we have little chance to learn human genetics at school, and we, genetic counselors and medical geneticists, should play a vital role to promote human genetic literacy of the public for providing optimal genetic services. "Family Health History (FHH)" (1) is an accessible tool that can capture genetic information, lifestyle and environmental factors, (2) allows health care providers to diagnose conditions and to assess risks and (3) increases health and genetic knowledge for the individual and the family. Therefore, we developed a manga booklet, as a new educational implement of human genetics and FHH. We conducted a pilot study to undergraduate students last year. The results showed that about half of them have more interest in talking about FHH with their family. However, approximately 40% of them answered that they would feel anxious about learning their own FHH. [Purpose and Method] The purpose of this study is to investigate the details of the anxiety caused by learning about FHH through the manga, which will be expected to give genetic experts ideas of how to approach clients' anxiety during genetic counseling sessions. We will send the manga to 40 public health nurses and several hundred adults throughout Nagahama City, Shiga, Japan this year. Participants will be asked to answer to a questionnaire after reading the manga. [Results and Discussion] According to a previous study conducted on college students in the U.S. about learning their own FHH, some answered that they would feel destined to develop the diseases which run in their family. Our study targets a large population of ordinary people. Also, there are no studies reported which measure anxiety of learning their own FHH in Japan. Therefore, this study will be the first one which focuses on this topic and could give genetic counselors and medical geneticists ideas of easing the different types of anxiety which clients could experience. Additionally, we are in the process of translating the manga into English.

3122T

What's the best way to educate people about genomics? Applying user centered design to consumer genomics education. *S. Briggs, R. Goldfeder, C. Liu.* Helix, San Carlos, CA.

The frequency of genetic testing has risen in recent years in both clinical and non-clinical (direct to consumer) settings. In direct to consumer genetics alone, Ancestry.com and 23andMe have each reported over 1 million individuals genotyped through their services to date. Despite this rise, many consumers remain unfamiliar with the technology and what insights can and cannot be gained from DNA. Furthermore, while significant research has been done comparing self-reported understanding of genetic concepts and results pre and post genetic testing, less has focused on how people would prefer to learn in this context. Our aims were to: 1. Test comprehension of genetic terms, concepts, and limitations, segmented by consumers who had either prior experience with or no experience with DNA testing; 2. Understand preferences and attitudes towards learning in general and specifically in genetics; 3. Gauge levels of perceived trustworthiness when different entities create and deliver genomic content. We surveyed 1000 people to understand user preferences around genomics education. The questions covered self reported learning styles, behavioral preferences, and knowledge about genetics. The results of this study provide both a qualitative and quantitative approach to designing effective resources for individuals to learn about their own DNA. This will be of critical importance when consumers will have the choice to pick genomic products interpreting specific parts of their genome.

3123F

The business of direct-to-consumer genetic testing: A complex moral challenge. *B.R. Haddad[1], S.D. Zaatari[2], L. Wathieu[3], K.T. FitzGerald[4].* 1) Department of Oncology, Georgetown University Medical Center, Washington, DC; 2) School of Medicine, Georgetown University, Washington, DC; 3) McDonough School of Business, Georgetown University, Washington, DC; 4) Pellegrino Center for Clinical Bioethics and Department of Oncology, Georgetown University Medical Center, Washington, DC.

The emergence of companies offering direct-to-consumer (DTC) genetic testing continues to generate significant debates and controversies among geneticists, medical professionals, ethicists, and in the public. Consequently, many moral challenges arise that are complex, ambiguous, and potentially devastating. Our study addresses the question: what do companies need to do to get this business "morally right" while the industry is still in its infancy? Our interdisciplinary team, with expertise in marketing, ethics, and genetics, identified and analyzed key practices of DTC genetic testing companies that raise moral challenges inherent to the industry (e.g., data ownership, research participation, presentation of results, marketing appeals used, etc.). Our goal in addressing these challenges is to set these companies on an effective course towards sustainable market success. To help us understand the latest industry developments, we conducted extensive phone and in-person interviews with seven key players at two leading DTC genetics testing companies, at a company specialized in testing for ethnic origins, and at a research organization that uses genomic data voluntarily provided by DTC users. We reached recommendations in the following areas, critical for the success of the DTC genetic testing industry: (1) the imperative to report genetic results to consumers in a manner that sufficiently and tangibly integrates known relevant contextual determinants (e.g., family history, behaviors, environment); (2) the imperative to report results in a manner that empowers consumer decision making (e.g., by referring to the efficacy and costs of actions that consumers might want to undertake); (3) the imperative to address the heavily emotional implications of genetic information ahead of testing and at the point of disclosure, and to avoid manipulating this emotional context for the sake of recruiting customers; (4) the imperative to pursue scientific validity and certification by public and private agencies; (5) the imperative to involve consumers fairly in the profits derived from the use of their data. Instead of reaching a final set of guidelines for industry self-regulation, our purpose is to engage companies offering DTC genetic testing, the public, scientists and medical professionals, ethicists, and government regulators, in a broad and diverse conversation around best business practices in DTC genetic testing which can help DTC companies "do it right.".

3124W

108,000 and counting: A review of the consenting process for Geisinger's MyCode® Biobank. *C. Flansburg, M.L.G. Hallquist, W.A. Faucett, D.J. Carey, The. MyCode Team, D.H. Ledbetter.* Genomic Medicine Institute, Geisinger Health System, Danville, PA.

The Geisinger MyCode® Community Health Initiative biobank began enrolling participants in 2007. The consent was updated in 2013 to include the return of actionable results with placement in the electronic health record (EHR). Support from the Geisinger patient community has been strong, with more than 108,000 individuals enrolled in the project toward a goal of 250,000 participants. ~90% of individuals agree to participate. Consent includes permission to: collect a sample when the participant has a clinical blood draw, perform genetic analysis, use information from the EHR for research purposes, recontact, share samples and health information with other researchers, and disclose actionable results to the participant and their providers. Over 40,000 participants were enrolled in 2015 by an average of 18 consenters working throughout the Geisinger Health System. Consenters have access to the EHR, which is used to track eligible patients and to record consent status. Consenting occurs in the clinic waiting room, at patient registration, once patients are roomed, or at checkout, based on the individual clinic flow. When approaching a potential participant, a consenter reviews the important aspects of the MyCode consent, answers any questions, and provides the written consent for review by the participant. Successful consenting requires collaboration with clinic leadership and clinicians in a model that adapts to the unique needs of each clinic. Consenters are the front line of MyCode, the first contacts with patients and the main community representatives of this project. Ongoing educational opportunities, such as Geisinger's C.I.CARE training and reviews of returned results, are provided to consenters for their education. Patient Satisfaction surveys indicate that individuals have positive experiences when approached by consenters. Consenters have many resources for responding to patient questions, including genetic counselors, laboratory staff, and study personnel. In addition to consenters embedded in clinics, in Nov 2015, we established online consenting for MyCode which requires an individual to have an account on our patient portal, MyGeisinger. To date, 2,752 participants have consented online and many have contacted the MyCode team for information. Additional opportunities for individuals to enroll in MyCode include events such as health fairs and consent via mail. Lessons learned in the development of the consenting process for MyCode will be shared.

3125T

Decision-making around secondary findings in genomics: Insights from whole-genome and whole-exome sequencing recipients. *M. Mackley[1,2], M. Parker[3], H. Watkins[1,2,4], E. Ormondroyd[1,4].* 1) Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK; 4) NIHR Oxford Biomedical Research Centre, Oxford, UK.

The return of secondary findings (SF) in whole-genome and -exome sequencing (WGS) is a subject of much debate. As WGS moves into routine clinical practice, policies around SF must take into account values and preferences of stakeholders. To inform this issue, we recently completed a systematic review of qualitative and quantitative studies that investigated stakeholder views on SF in WGS. We found that, while there was agreement that 'actionable' findings should be available, there was no consensus on what was considered 'actionable'—the balancing of rights and responsibility variably influenced stakeholder views. Furthermore, we found that many studies were probing the wrong knowledge in assessing literacy; our findings suggest that truly informed decisions require an understanding of the implications of WGS and possible findings, rather than technical details. The majority of studies came from the US, emphasizing a need for wider perspectives on SF in WGS to be explored. Molecular Genetic Studies of Individuals and Families at Risk of Inherited Disease (MGAC) is a WGS study that preceded the NHS England 100,000 Genomes Project in Oxford, and recruits adults and children with diverse rare diseases. Frequently a trio design is used, wherein unaffected parents and affected child are recruited. During informed consent, adult participants are offered choices about whether to receive secondary findings: incidental, and/or additional screening of selected genes based on the ACMG list. In order to address gaps highlighted by our systematic review we are investigating the views of adult participants in MGAC. Participants are sent a questionnaire eliciting understanding and motivation for WGS, and invited to participate in semi-structured interviews. In particular, decision-making around SF disclosure and participant appreciation for the implications of SF will be investigated. Thematic analysis of interview data is being undertaken. Eighty-eight adult participants are eligible to participate, and questionnaires have been distributed by post (current response rate = 28%). Data will be presented from analysis of interviews completed to date. Preliminary analysis suggests that participants feel empowered by their ability to make decisions around SF, but want to share responsibility. Some participants, based on their experience of genetic disease, appreciate the context-dependency of WGS results—suggesting that broader or dynamic forms of consent should be considered.

3126F

Patients' perceived utility after receiving whole genome sequencing. *J.O. Robinson[1], K.D. Christensen[2], P.M. Diamond[3], L. Jamal[1], K. Lee[1], H. Peoples[1], C. Blout[2], J.L. Vassy[2,4], R.C. Green[2,5,6], A.L. McGuire[1] for the MedSeq Project.* 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) University of Texas School of Public Health, Houston, TX; 4) VA Boston Healthcare System, Boston, MA; 5) Partners HealthCare Personalized Medicine, Cambridge, MA; 6) Broad Institute of MIT and Harvard, Cambridge, MA.

　Background: Genome sequencing is increasingly being integrated into clinical care as a diagnostic and screening tool. Patients' perceived utility of this information is unclear. **Methods:** The MedSeq Project is a randomized clinical trial comparing whole genome sequencing (WGS) to standard of care which included a review of family history (FH) information, in two study cohorts, healthy primary care patients and cardiology patients with cardiomyopathy diagnoses. Patients were surveyed at baseline (BL) and 6-months (6-mos.) after results disclosure to assess multiple domains, including 6 questions about perceived utility of the study results (WGS or FH). **Results:** Of 200 enrolled patients, to date 168 have completed the 6-month follow-up survey. The percentage of participants who perceived utility in study results decreased from baseline to the 6-month follow-up on all items: impacting medical decision-making (87% BL vs. 32% 6-mos.), affecting medical treatment (85% BL vs. 50% 6-mos.), identifying disease risk (80% BL vs. 60% 6-mos.), influencing medications (73% BL vs. 16% 6-mos.), impacting end-of-life planning (46% BL vs. 13% 6-mos.), and influencing reproductive decisions (33% BL vs. 12% 6-mos.; all $p<0.05$). Decreases in perceived utility differed across randomization arms for 2 of the 6 utility items: identifying disease risk (3% WGS vs. 28% FH, $p<0.001$) and influencing treatments (16% WGS vs. 37% FH, $p=0.008$). Changes on all other items did not vary significantly by randomization status ($p>0.05$). Adjusted logistic regression model analysis showed that participants who were randomized to WGS, compared to FH, were more likely to report utility for identifying disease risk (OR=7.7, 95% CI=2.8-21.3) and influencing medical treatment (OR=2.6, 95% CI=1.4-5.0), regardless of study cohort. **Conclusions:** Six months after receiving WGS or FH information participants in both arms perceived less utility in their study results than they anticipated there to be at baseline. However, both WGS and FH were perceived to have some utility in all areas assessed at follow-up. WGS was perceived to have greater utility than FH for identifying disease risk and influencing medical treatments. These findings have implications for pre-test counseling. Clinicians should help set realistic expectations about the types of information patients will receive from WGS and the utility of WGS, and even FH information.

3127W

Reporting practices for next generation sequencing technologies: A qualitative study with laboratory personnel. *D. Vears[1], K. Sènècal[2], P. Borry[1].* 1) Center for Biomedical Ethics and Law, KU Leuven, Leuven, Belgium; 2) Centre of Genomics and Policy, McGill University, Montreal, Canada.

　Next-generation sequencing (NGS) technologies are revolutionizing genetic diagnosis, both in research and clinical contexts, by identifying causative mutations in previously undiagnosable individuals, influencing their treatment and providing important information to their families. However, these technologies also have the potential to identify additional genomic information that is unrelated to the illness under investigation, such as variants in genes with unknown significance (VUS) and mutations in genes causing phenotypes extraneous to the clinical question (incidental findings, or IFs). Recommendations by professional bodies outlining which findings should be reported by laboratories to the clinician requesting the test are inconsistent, particularly regarding the duty to search for variants unrelated to the clinical question. Little is known about which variants laboratories are reporting and how they are using guidelines to make decisions about this.　To address these questions, in-depth interviews were conducted with 25 laboratory personnel in Europe, Canada and Australasia to explore actual reporting practices for diagnostic NGS. Interviews were transcribed, deidentified and analysed using inductive content analysis. Reporting practices were then compared with published guidelines to determine whether their testing policies were in line with current guidelines.　Reporting of IFs differs between laboratories. Some laboratories limit their reporting to variants considered to be causative of the phenotype under investigation. Others report IFs to varying degrees. Decisions about which variants to report often involves detailed discussions, involvement of multidisciplinary committees, and consideration of factors such as actionability or medical relevance. There was also considerable variation in the filtering strategies that participants described. While some laboratories use quite stringent filtering in order to limit the number of IFs that are identified, others are less strict, allowing identification of these variants. Only a small number of laboratories actively search for variants unrelated to the clinical question.　Our study highlights that laboratories are still grappling with decisions about which variants to report from NGS. These findings will assist laboratories to learn from each other's experiences and refine reporting procedures.

3128T

Negotiating the research-clinical interface in genomic medicine: Analysis from the CSER Consortium. *W. Burke[1], L. Amendola[2], E. Clayton[3], R. Green[4], J. Harris-Wai[5], G. Henderson[6], G. Jarvik[2], B. Koenig[5], L. Lehmann[7], A. McGuire[8], P. O'Rourke[9], C. Somkin[10], B. Wilfond[11], S. Wolf[12]. 1) Bioethics and Humanities, University of Washington, Seattle, 98195, WA; 2) Medicine (Medical Genetics), University of Washington, Seattle 98195, WA; 3) Center for Biomedical Ethics and Society, Vanderbilt University, Nashville 37232, TN; 4) Medicine (Genetics), Brigham and Women's Hospital and Harvard University, Boston 02115, MA; 5) Institute for Health Aging, UCSF School of Nursing, San Francisco 94118, CA; 6) Social Medicine, University of North Carolina, Chapel Hill 27599, NC; 7) Center for Bioethics, Brigham and Women's Hospital and Harvard University, Boston 02115, MA; 8) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston 77040, TX; 9) Human Research Affairs, Partners HealthCare, Boston 02115, MA; 10) Division of Research, Kaiser Permanente, Oakland 94612, CA; 11) Treuman Katz Center for Pediatric Bioethics, Seattle Children's, Seattle 98101, WA; 12) Consortium on Law and Values in Health, Environment and the Life Sciences, University of Minnesota, Minneapolis 55455, MN.*

The NIH-funded Clinical Sequencing Exploratory Research (CSER) Consortium seeks to support translation of genomic and exome sequencing into clinical applications. CSER studies generally occur in clinical settings and results may have implications for patient care. These studies provide insights into the challenges of combining research and clinical activities in translational genomics. Prior literature has contrasted norms and rules of research vs. clinical care, but little is known about how investigators navigate these issues. **Methods**: Each of nine CSER study teams completed a survey on how the research-clinical interface was manifested in their study, problems raised and solutions devised. The survey queried study consent procedures; involvement of research and clinical personnel; use of a research or clinical lab; return of research results; where results are recorded; legal issues raised; procedures coverd by the research budget vs. billed to insurer; and views of the study team on the interface between genomic research and clinical care. <u>Results:</u> All CSER studies use consent procedures to identify research processes. Several studies use dedicated research spaces and research personnel to differentiate research activities. Studies vary in their use of a CLIA-certified clinical lab vs. CLIA confirmation of research results. Most studies return research results to participants, but the scope of return differs. Studies vary in placement of results in the research record, medical record, both, or depending on participant choice, and in whether results are returned by research or clinical personnel. Studies note participant confusion related to the research-clinical distinction, e.g., confusion between research and clinical consents or a participant directing questions about clinical follow-up to research personnel. Legal issues considered include privacy, malpractice, HIPAA, genetic discrimination, and CLIA. Studies vary in what is billed to the research budget vs. insurers. Problems noted include limited availability of clinical follow-up to pursue genomic findings, and divergent perspectives held by researchers and clinicians. **Conclusions:** Developing an evidence base for clinical sequencing compels researchers, clinicians and participants to navigate the interface between research and clinical care. The CSER studies reveal key challenges and evolving strategies.

3129F

Implementing genomics into clinical practice: A collaborative approach to removing barriers. *C.L. Gaff[1], I. Macciocca[1,2], M. Martyn[1,3], E. Forbes[1], S.M. White[2,4], Z. Stark[2], D. Schofield[3,5], N. Mupfeki[1], K. Alam[3,4], P. James[4,6], A. Roberts[6,7], M. Ryan[8], T. Day[6], M. Walsh[2], P. Kwan[6], P. Perucca[6], A. Boussioutas[4,6], T. Tan[2], W. Wilson[9], G. Taylor[2], A. Oshlack[3], G. Brett[1,2], E. Creed[1,7], E. Wilkins[1,6], N. Thorne[1], T. Bakker[1], Melbourne Genomics Health Alliance. 1) Melbourne Genomics Health Alliance, VIC, Australia; 2) Victorian Clinical Genetics Services, VIC, Australia; 3) Murdoch Childrens Research Institute, VIC, Australia; 4) University of Melbourne, VIC, Australia; 5) University of Sydney, NSW, Australia; 6) Melbourne Health, VIC, Australia; 7) Walter and Eliza Hall Institute, VIC, Australia; 8) Royal Children's Hospital, VIC, Australia; 9) CSIRO, NSW, Australia.*

Widespread change within and across organisations is needed to resolve the numerous barriers to genomics in healthcare. The Melbourne Genomics Health Alliance is taking a collaborative approach to this task. Ten independent healthcare, research and academic organisations are working together to embed change in practice "starting with the patient and working backwards". We conducted a 2-year demonstration project to build genomics capability, prospectively evaluate the impact of genomic sequencing and identify paths to full implementation. We assessed the processes of genetic counselling, variant interpretation and reporting, information management, data sharing and data linkage. Common systems, tools and policies were developed and applied across the participating hospitals and laboratories. Clinicians offered whole exome sequencing (WES)-with targeted gene analysis in parallel with standard care to 316 patients with one of five clinical indications, enabling condition-specific and common approaches to be identified. Three accredited laboratories collaborated in sequencing and reporting of results. Evaluation data were collected from clinical records, participant surveys and interviews with 32 clinicians (7 disciplines) and 14 other stakeholders. The impact of WES varied by indication. Targeted WES increased detection for 3 of the 4 germline conditions and changed patient management in 2 of these to date. WES was highly cost effective for infants with features of known monogenic conditions: AUD $5,047 per diagnosis for singleton WES compared to AUD$27,050 for standard care. WES as a first tier test would achieve a saving per additional diagnosis of AUD$2,182 (95%CI: 129.92 - 5,855.02). Clinicians identified processes that are perceived as critical for the successful application of genomics (e.g. patient selection) and others that support adoption.(e.g. multidisciplinary variant review). There was no consensus on return of predictive information. Pre-test genetic counselling was effective; more than 75% of participants answered knowledge questions correctly. 93% agreed to share their data for research. 6% expressed concerns related specifically to genomics. This comprehensive evaluation provides crucial insights into pathways to progress the introduction of genomic medicine collaboratively. Models proposed by clinicians will be presented. The results are informing implementation of systems that meet the needs of both the workforce and patients.

3130W

Improving access to genetic services: Needs assessment data from the National Regional Genetic Collaboratives System. *A.S. Keehn[1], D.M. Maiese[1], M. Lyon[1], J. O'Leary[2], S. Alexander[2], C. Kaye[3], NCC Evaluation Workgroup Members.* 1) American College of Medical Genetics and Genomics, Bethesda, MD; 2) Genetic Alliance, Bethesda, MD; 3) University of Colorado School of Medicine, Denver, CO.

   Purpose To strengthen and support the genetic and newborn screening capacity of the states and to improve the availability, accessibility, and quality of genetic services and resources for individuals with, or at risk for, genetic conditions and their families is the mission of the National Coordinating Center and the seven Regional Genetic Service Collaboratives (NCC/RC system). The NCC is administered by the American College of Medical Genetics and Genomics (ACMG), in partnership with the Genetic Alliance as a subcontractor for the National Genetics Education and Consumer Network. The RCs are funded under separate Health Resources and Services Administration (HRSA) grants. Now in its 12th year, the NCC/RC system has common evaluation data elements related to HRSA priorities and Healthy People 2020 objectives that focus on care coordination (e.g., medical home and transition from pediatric to adult care); telegenetics; newborn screening capacity and long-term follow-up. **Methods** In 2015, the NCC conducted listening sessions and fielded national surveys to individuals and families with genetic conditions and genetic service providers (medical geneticists, genetic counselors, metabolic dietitians, and other genetic professionals). The ACMG Member Directory, listservs from national and RC partners, and social media were used to electronically distribute the surveys. **Summary of results** Across the country genetic services needs are not being met. Our findings show: most genetic professionals work in university or children's hospitals; the majority of their time is spent in clinical care; two thirds of geneticists report that their current wait time for a non-emergency appointments is greater than one month with respondents at children's hospitals reporting the longest wait time; and use of telemedicine is low. Barriers to access include too few clinical geneticists and genetic counselors; geographic disparities; costs of testing and lack of reimbursement; and lack of education. Potential solutions include better care coordination, more genetics professionals, telemedicine, increased acceptance of billing codes by insurers, increasing genetic competency for non-genetics professionals, and making genetic counseling separately reimbursable. The NCC is funded by *U22MC24100*, awarded as a cooperative agreement between the Maternal and Child Health Bureau/Health Resources and Services Administration, Genetic Services Branch, and the ACMG.

3131T

Disseminating best practice in ELSI research to support the translation and uptake of high value precision medicine technologies in Canada: The Precision Medicine Policy Network. *S.K. Genuis[1], F. Rousseau[2], B. Knoppers[3], B. Wilson[4], M. Wolfson[5], T. Bubela[6], C. McCabe[1].* 1) Department of Emergency Medicine, University of Alberta, Edmonton, AB, Canada; 2) Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Quebec, QC, Canada; 3) Department of Human Genetics, McGill University, Montreal, QC, Canada; 4) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada; 5) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, ON, Canada; 6) School of Public Health, University of Alberta, Edmonton, AB, Canada.

   Precision Medicine promises to improve clinical care and efficiencies in health care. Despite significant government investment, this promise has not yet been achieved. There are significant barriers to the successful implementation of precision medicine technologies in Canada and internationally. Many of these barriers are related to the ethical, legal and social implications of Genomics research (ELSIs). In 2013 Genome Canada (GC) funded 17 Genomics and Personalized Health (GAPH) Large Scale Applied Research Projects focused on bringing genomics technologies to market. By 2015 it had become clear that ELSI knowledge was siloed within individual projects. Genome Canada recognized that substantial added value could be obtained by synthesizing this knowledge and experience that has developed within the research teams. On January 1, 2016, Genome Canada provided funding for the *Precision Medicine Policy Network*, a collaborative 'learning network' of the 17 Projects. Focusing on four, high priority themes (Research Ethics; Health Economics & Health Technology Assessment; Knowledge Translation & Implementation Science; and Intellectual Asset Management & Commercialization) the Network has brought together expertise from the 17 GAPH Projects and from academic, policy and private sector communities. Through critical appraisals and consensus workshops, the Network is developing good practice guides for each of the priority areas. Over the next two years, the network will be delivering pan-Canadian training workshops to support their dissemination.

3132F

Effect assessment of genome research supporting policy and program alteration in Korea. *S. Moon, K. Jang, H. Kim.* Korea Institute of Science&Technology Evaluation and Planning, Seoul, South Korea.

Most of the times, research in innovative fields such as genomics develops with close relation to public funding. This report analyzes genomics research in Korea with regard to governmental research supporting policy alteration. In very recent years, research funding programs for genomics in Korea has experienced dramatic change. Unlike other life sciences area, genomics research funding has adopted a multi-departmental program format, which emphasizes collaborational policy mix and shared roles and responsibilities. We will discuss the effect of this multi-departmental support with emphasis on budget accumulation and allocation, composition of research topics so that the envelope of research moves towards common goal. We also focus on the changes that have been made to supporting policy that can influence researchers in the field, including legal and ethical issues, and database generation, standardization and information sharing. We expect to shed light on connection between research policy and laboratory research, which could strengthen future policy design and execution to accelerate meaningful findings in human genomics.

3133W

Regional genetic service models: What's out there and what could the US learn for future genetics services delivery models? *C. Kaye[1], A.S. Keehn[2], M. Lyon[2], M.S. Watson[2], D.M. Maiese[2], S. Alexander[3], J. O'Leary[3], NCC RSSM Workgroup and Advisory Commitee Members.* 1) SOM, Dean's Office, Univ Colorado, Parker, CO; 2) American College of Medical Genetics and Genomics, Bethesda, MD; 3) Genetic Alliance, Washington, D.C.

Purpose To understand the current landscape of regional models of healthcare delivery across the United States (and the genetics service in the United Kingdom) in order to consider potential future regional models of genetics support services. The National Coordinating Center for the seven HRSA Regional Genetic Service Collaboratives (NCC/RC system), housed at the American College of Medical Genetics and Genomics (ACMG), pursued a comprehensive review and development of future regional support service models. **Methods** In 2015, a small working group sponsored by the NCC, the Regional Support Service Models Workgroup (RSSM WG) and an Advisory Committee (AC) considered a set of 35 recommended regional service delivery models, comprehensively reviewing 13 of those models via telephone interviews completed by individual workgroup members. The RSSM workgroup also reviewed national needs assessments of genetics providers and consumers of genetic services completed by the NCC and its partner, the National Genetics Education and Consumer Network (NGECN). Using reviewed models and the genetics service priorities and needs identified through the needs assessments, the RSSM workgroup then developed eight models of regional genetic services that could be considered in the future. **Summary of results** Center reviews revealed key themes: collaborative model can be very effective; support to outreach clinics can vary; no model provides mental health support; data collection—registries and robust databases—can be a very powerful tool for long-term success; and genetics and primary care workforce issues are real. The national needs assessment identified the following priorities: data collection, education and training, efficient practice, family engagement, formal partnerships, and practice support. This work led to the identification of eight regional support service models that range from a regional genetics education and TA center (similar to an ECHO model) to a genetic service data center to a quality improvement model. The models, their pros and cons, and timelines have been developed and reviewed by the RSSM WG and AC, offered for open comment for a month, revised and edited, and finalized. The NCC and NGECN are funded by *U22MC24100*, awarded as a cooperative agreement between the Maternal and Child Health Bureau/ Health Resources and Services Administration, Genetic Services Branch, and the ACMG.

3134T

Integrative personal omics profiling during periods of diet or infection perturbations. *M. Snyder[1], W. Zhou[1], B. Piening[1], K. Contrepois[1], G. Gu[1], S. Leopold[3], T. Mishra[1], D. Perleman[2], B. Leopold[3], E. Sodergren[3], T. McLaughlin[2], G. Weinstock[3], Integrative Human Microbiome Project consortium.* 1) Genetics, Stanford University, Stanford, CA 94305; 2) Division of Endocrinology, Stanford University School of Medicine, Stanford, CA 94305; 3) The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030.

While significant genetic and environmental risk factors are known to contribute to the development of Type 2 Diabetes (T2D), overall our ability to predict which individuals will develop T2D and when this will occur is woefully inadequate. To better understand these factors, we present a longitudinal multi-omic personalized profiling of comprehensive blood- and microbiome-based analytes that we apply to track the progression to T2D in a cohort of 100 individuals over periods of health, illness and weight gain and loss. Multi-omic profiling (transcriptome, proteome, metabolome etc.) revealed significant differences in multiple 'omes between prediabetics and healthy controls at steady state, implicating pathways related to chronic inflammation and insulin regulation as well as novel connections to T2D. A subset of participants was then placed on a short-term high caloric diet and profiled by longitudinal multi-omics approach. The dietary perturbation was associated with a wealth of biomolecular expression changes concomitant with weight gain and spanning multiple 'omes including the microbiome, and the omic response to weight gain differed between prediabetics and healthy controls. In another subset of participants who went through respiratory viral infections, their multi-omic profiling, including the microbiome, responded distinctly to different illness stages during the infection. Overall, multi-omic profiles of individuals are unique to themselves regardless diet or illness perturbations. Together, these large-scale longitudinal data offer a novel and comprehensive view of the dysfunction in cellular networks associated with the progression to T2D and may offer new strategies for predicting and preventing the disease.

3135F

Making humanized major histocompatibility complex (MHC) mice to study Graves' disease. *P.L. Chen[1,2,3], K.C. Chen[2], T.C. Chang[3].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 2) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Graves' disease (GD, MIM 27500), the global leading cause of hyperthyroidism and thyroid eye disease, is an autoimmune disease with both clinical and academic significance. To further investigate the mechanism of GD, and also to develop possible novel treatment and/or prevention methods, murine models have the potential to provide invaluable insights. Most immune-related human diseases actually lack good animal models. Technology advances (such as recombineering, pronuclear injection-based mouse targeted transgenesis (PITT) and CRISPR/Cas9) in recent years would open a new era of excellent humanized animal models. Previously animal models involved in MHC were mostly simple transgenic lines. There are syntenic regions between human and mouse MHC loci/genes, but the correlations are far from perfect. More sophisticate humanized MHC knock-in design holds the promise to shed light on the mechanism of human immune-related diseases. We aimed at generating mouse lines carrying various *HLA-DRB1* alleles. We started with the embryonic stem cells from the C57BL/6N line, and therefore future backcrossing to the B6 background will not be necessary. We successfully achieved first targeting; the mouse *H2-Aa* was replaced with human *HLA-DRA* gene, which was retrieved from human BAC RF11-379F19. We then accomplished the second targeting by replacing mouse *H2-Ab1* with human *HLA-DRB1* gene. It is noteworthy that the human *HLA-DRB1* gene was retrieved from human BAC CTD-2510D15, and we later noticed that the sequence at the splice donor site was "AT" (instead of "GT") at the +1 and +2 position of intron 1, which was predicted to cause abnormal splicing results. We then successfully corrected the sequence from "AT" to "GT" at the targeting vector, which was 28,770 bp in size before the final blastocyst injection. Future validation and phenotyping will be needed after we get the mice carrying human HLA genes. Humanized MHC mice could provide important help to study GD and possible other immune-related human diseases/phenotypes.

3136W

Large scale analysis of HLA-peptide interactions. *H. Miyadera[1,2,3], Y. Okabe[2,3], K. Tokunaga[3], E. Noguchi[1].* 1) Department of Medical Genetics, University of Tsukuba, Tsukuba, Ibaraki, Japan; 2) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan; 3) Department of Human Genetics, The University of Tokyo, Bunkyo-ku, Tokyo, Japan.

   Genes encoding the human leukocyte antigens (*HLA*) are associated strongly with a variety of immune related disorders including infectious diseases and autoimmune disorders. However, exact mechanisms of *HLA*-disease association have remained elusive for most of these diseases. To identify disease-relevant self- and foreign peptides and uncover mechanisms of *HLA*-disease association, we devised an assay to screen for HLA-binding peptides for a large numbers of *HLA* allele products.   Methods: We used cell-surface HLA expression assay (Miyadera, et al. 2015, JCI) to screen for peptides that are potentially presented by HLA class II, including *HLA-DQ*, *-DR*, and *-DP*. First, we established stable cells for HLA class II alpha subunit using murine fibroblast cells (NIH3T3) as a host cells. The stable HLA class II alpha-cells were transiently transduced with HLA class II beta subunit, which is fused to peptides. The level of cell surface HLA expression for each HLA-peptide combination was measured by flow cytometry and used as an estimate of HLA-peptide interaction.   Results: We validated the assay using the peptides that are known to bind strongly or weakly to certain HLA allele products. We then extended the analysis to screen for self- and foreign peptides that might be involved in certain HLA-associated diseases including narcolepsy, type 1 diabetes, and hepatitis B. In addition to peptides, allelic differences in the intrinsic stability of HLA-DQ, -DR, and -DP were also estimated to analyze potential association between HLA stability and disease susceptibility. .

3137T

A rapid genotyping research assay for determining basic ABO RhD blood types and CCR5 status. *E. Schreiber[1], T. Nong[2], J.H. Lee[2].* 1) Thermo Fisher Scientific, Genetic Sciences Division, South San Francisco, CA; 2) Thermo Fisher Scientific, One Lambda Division, Canoga Park, CA.

   Purpose: Bone marrow registries require high resolution HLA genotyping information of donor specimen for matching them with their future recipients. For optimal outcome it is further desirable to use bone marrow of identical or compatible ABO RhD/CE blood types between donor and recipient. In order to aid research in this area, we have developed a rapid research assay that utilizes DNA isolated from buccal swaps of past marrow donors to determine the genotypes of the ABO blood antigens and the presence (+) or absence (-) of the RhD antigen. In addition the assay also determines the presence or absence of a 32 bp deletion in the CCR5 gene. Homozygous carriers of the deletion are resistant to HIV-1 infection and thus could be valuable stem cell donors for HIV-infected recipients.   Methods: The research assay is based on a single multiplex-PCR reaction with 5 fluorescently-labeled and 12 allele-specific primers followed by capillary electrophoresis on an automated Applied Biosytems 3500 Genetic Analyzer. The primers for the human ABO gene target 4 SNPs in exons 6 and 7 allowing the determination of the A, A2, B, O1,2 and O3 alleles. The peak pattern generated on the sequencer is analyzed by GeneMapper® software and the resulting peak/genotype table is translated into a genotype/phenotype status report using standard spreadsheet software. Results: We have verified the assay by analyzing a panel of DNA samples with known blood group antigens and CCR5 gene types and found it to be 100% comparable. Conclusion: Taken together, this easy to use, inexpensive and rapid research assay may prove useful in the future towards the development of an assay that could be used for the stratification or identification of bone marrow donors and other areas of public health studies. For research use only. Not for use in diagnostic procedures.

3138F

Characterizing haplotype diversity at the immunoglobulin heavy chain locus across human populations using novel long-read sequencing and assembly approaches. *C.T. Watson[1,2], M. Laird Smith[2,3], W. Gibson[2], O. Rodriguez[2], M. Pendleton[2], P. Comella[2], L. Harshman[4], W.A. Marasco[5,6], E.E. Eichler[4], R. Sebra[2], J. Korlach[3], A.J. Sharp[2], A. Bashir[2].* 1) Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Pacific Biosciences, Menlo Park, CA; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Department of Cancer Immunology & AIDS, Dana-Farber Cancer Institute, Boston, MA; 6) Department of Medicine, Harvard Medical School, Boston MA.

The human immunoglobulin heavy chain locus (IGH) remains among the most understudied regions of the human genome. Recent efforts have shown that haplotype diversity within IGH is elevated and exhibits population specific patterns; for example, our re-sequencing of the locus from only a single chromosome uncovered >100 Kb of novel sequence, including descriptions of six novel alleles, and four previously unmapped genes. Historically, this complex locus architecture has hindered the characterization of IGH germline single nucleotide, copy number, and structural variants (SNVs; CNVs; SVs), and as a result, there remains little known about the role of IGH polymorphisms in inter-individual antibody repertoire variability and disease. To remedy this, we are taking a multi-faceted approach to improving existing genomic resources in the human IGH region. First, from whole-genome and fosmid-based datasets, we are building the largest and most ethnically diverse set of IGH reference assemblies to date, by employing PacBio long-read sequencing combined with novel algorithms for phased haplotype assembly. In total, our effort will result in the characterization of >15 phased haplotypes from individuals of Asian, African, and European descent, to be used as a representative reference set by the genomics and immunogenetics community. Second, we are utilizing this more comprehensive sequence catalogue to inform the design and analysis of novel targeted IGH genotyping assays. Standard targeted DNA enrichment methods (e.g., exome capture) are currently optimized for the capture of only very short (100's of bp) DNA segments. Our platform uses a modified bench protocol to pair existing capture-array technologies with the enrichment of longer fragments of DNA, enabling the use of PacBio sequencing of DNA segments up to 7 Kb. This substantial increase in contiguity disambiguates many of the complex repeated structures inherent to the locus, while yielding the base pair fidelity required to call SNVs. Together these resources will establish a stronger framework for further characterizing IGH genetic diversity and facilitate IGH genomic profiling in the clinical and research settings, which will be key to fully understanding the role of IGH germline variation in antibody repertoire development and disease.

3139W

Comparative study reveals superior performance of WGS compared to WES in a real-world study of human Mendelian diseases. *Y. Huang[1], J. Hughes[2], J. Zhuang[2], M. Lenardo[3], H. Sun[3], J. Milner[3], R. Goldback-Mansky[4], H. Runz[1], J. McElwee[1].* 1) Genetics and Pharmacogenomics, Merck Research Labs, Boston, MA; 2) Computational Genomics & Genetics, Merck Informatics IT, Boston, MA; 3) National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 4) National Institute of Arthritis and Musculoskeletal and Skin.

Next-generation sequencing has established itself as the *de facto* choice for measuring genome-wide genetic variation in clinical and research settings. The effectiveness of its practical application, however, largely depends on carefully selecting the appropriate sequencing technology, namely whole genome sequencing (WGS) or whole exome sequencing (WES). Providing base pair level resolution of the genome, WGS is more attractive for complete details of one's genetic makeup. On the other hand, WES offers a more affordable solution cost-wise. The drawback is that WES covers only ~50-65M base pairs of the genome using various capture reagents (e.g., Agilent SureSelect™ v5 and Nimblegen SeqCap v3). In order to establish a best-practice guidance for sequencing studies, it is important to understand the strengths and weaknesses of each technology. In the present study, we provide an objective evaluation on the two sequencing technologies in a large-scale, real-world human sequencing analysis. A cohort of 208 subjects exhibiting Mendelian immune diseases was sequenced with both WGS and WES. We compared WGS and WES for the identification and interpretation of genes and variants in the ensemble consensus coding sequences (CCDS) as a consistent target region of interest for Mendelian and functional genetics. WGS showed significantly better coverage, with an average 99% of CCDS covered at 10x or greater by WGS compared to 91% in WES. In addition, WGS has significantly fewer bases (20,950 for WGS vs. 37,273 for WES) that are not mapped. The majority (71.9% for WGS vs. 47.6% for WES) of these were found to be deletions in dbSNP. WGS also exhibits more uniform coverage across genome while WES shows higher level of variability across target regions. The average coefficient of variation of WGS is 10-fold smaller than that of WES, 1% vs. 10% respectively. In particular, WES suffers from low coverage in genes with high GC content (e.g., CEBPB, TNFRSF4, SOCS1) as well as GC rich regions (e.g., 5' UTR and first exons). When inspected at individual gene level, as many as 374 genes have >95% of their coding regions well-covered by WGS compared to <70% coverage in WES. These differentially-covered genes include 41 (12%) genes relevant for Mendelian disease, such as APOE, INS, and PDX1. In summary, WGS delivers superior power for variant calling in many regions of the genome, with greater mappability, more uniform coverage, and better performance in high GC content regions.

3140T

The Ion S5™ ExT Workflow: Enabling rapid, automated, long-read sequencing for the Ion S5™ sequencing system. *G.M. Lowman[1], M. Landes[1], V. Mozhayskiy[1], X. Peng[1], G. Luo[1], R. Lyons[1], M. Shenasa[2], L. Ong[2], J. Ghadiri[2], C. Buie[2], A. Gulati[2], J. Dinis[2], A.G. Jung[2], S. Lee[2], A. Pomorski[2], T. Burgess[2], P. Thwar[2], E. Linch[1], L. Miller[1], C. Hoff[2], C. Duncan[1], D. Mazur[1], A. Mongan[2], K. Lam[3], K. Beutner[3], Y. Saito[3], J. Lee[3], W. Zhang[1], M. Andersen[1], P. Vander Horn[1].* 1) Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA 92008; 2) Thermo Fisher Scientific, 180 Oyster Point Blvd. South San Francisco, CA 94080; 3) One Lambda, 21001 Kittridge St., Canoga Park, CA 91303.

The Ion S5™ ExT templating and sequencing reagents for the Ion S5™ sequencer provide a fully automated and rapid sample-to-answer solution enabling a wide range of applications including research in human leukocyte antigen (HLA) genotyping, immune repertoire sequencing, bacterial identification, and metagenomic analysis with read lengths greater than 600bp. This unique workflow permits a single day turnaround time, including sample templating, sequencing, and analysis, which is imperative for applications where time is of the essence. The Ion S5™ExT system provides flexibility in semiconductor chip selection for large-scale studies where sample number, cost, and automation is critical or smaller studies where answers are needed quickly. By combining rapid templating technology and automation on the Ion Chef™, ExT technology supports read lengths greater than 600bp with less than 30 minutes of hands-on time. When paired with the Ion Xpress MuSeek™ library preparation kit, an 8-hour DNA-to-answer workflow becomes a reality. The benefits of the ExT workflow are shown in HLA genotyping studies where longer Ion Torrent reads demonstrate improved coverage bridging exons and in high-GC regions. The concordance rate between the longer read protocol-derived genotypes and those generated by TypeStream™ auto-analysis was 99.7% for the HLA-A, B, C, DRB1, DQB1, DPB1 and DRB 345 loci combined. Importantly, reads in excess of 600bp provide a route for HLA genotyping to unambiguously establish the gametic phase between alleles. Metagenomic studies show the benefit of longer reads to improve species level characterization in 16S samples, and extending read length enables sequencing the full variable region of the TCR(α,β) immune repertoire.   *For Research Use Only. Not for use in diagnostic procedures. .

3141F

Single-cell transcriptional profiling identifies rare cell types with gene markers for classification and prognosis in acute myeloid leukemia. *A. Sartori[1], E. Falconnet[1], P. Ribaux[1], S.E. Antonarakis[1,2,3], T. Matthes[2], C. Borel[1].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) University Hospitals of Geneva, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland.

Acute myeloid leukemia (AML) is a particularly aggressive blood cancer difficult to treat because of its recurrence affecting differentiation of hematopoietic progenitors. It is hypothesized that relapse of AML is due to incomplete eradication of leukemic stem cells (LSC) with self-renewal and leukemia-initiating properties. Our ultimate goal is to characterize LSC based on their transcriptional profiling. Evidence suggests that CD34+CD38− cell population is enriched for LSCs but they remain rare and difficult to identify. Thus, we present a single-cell RNA-seq approach to characterize CD34+CD38- sorted cells in 2 AML individuals (AML1(M0), AML2(M5)) at the time of diagnosis and in 4 unaffected individuals (N). For 311 single-cells, 1764 genes were detected on average (RPKM>10). We identified 763 and 858 significantly differentially expressed genes in AML2 and AML1, respectively using D3E. We demonstrate that the captured AML cells possess the hallmarks of LSC. The top highly expressed genes are stem markers including *SOX4, CD74, CD69* and *CXCL8*. In particular, *HOXA-B* cluster genes appeared markedly dysregulated demonstrating their leukemia-initiating capability. Gene ontology revealed enrichment for annotation of apoptotic resistance and self-renewal capacity as well. Furthermore, 3 distinct clusters of co-expressed genes were identified that distinguished between N cells and AML cells with M0 and M5 (SEURAT-tool). Finally, we hypothesized that prognostic gene-expression signature is present at diagnosis. Using TCGA RNA-seq and clinical datasets, we identified genes whose expression correlated with AML patient survival. We confirmed their differential expression in AML single-cells and thus, their contribution in leukemogenesis and potential relevance for clinical-outcome prediction.

3142W

Utilizing integrative omics to identify phenotypic subtypes of childhood asthma. *R.S Kelly[1], A. Dahlin[1], A.C Wu[1], M.J McGeachie[1], D.C Croteau-Chonka[1], C.B Clish[2], J. Celedón[3], S.T Weiss[1], J.A Lasky-Su[1].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Boston, MA; 3) Division of Pediatric Pulmonary Medicine, Allergy and Immunology University of Pittsburgh, Pittsburgh, PA.

Background: Asthma is a major global public health problem, but its heterogeneous nature is poorly understood. Integrative omic studies represent a novel approach to characterize asthma subtypes and their underlying biology. **Methods:** Metabolomic and transcriptomic profiling was performed on serum of 381 asthmatic children from a genetically isolated Costa Rican population (CR). After quality control and processing, four liquid chromatography mass spectrometry platforms and whole genome sequencing yielded 8185 endogenous metabolites and 25060 unique transcripts, respectively. Hierarchical clustering was used to identify groups of asthmatic children with similar transcriptomic, metabolomic and integrated omic profiles. The epidemiological, immunological and clinical characteristics of the different clusters were compared, and the genes and metabolites driving the formation of the clusters were identified using one-way ANOVA. Validation was performed using an independent cohort of 612 asthmatic children (CAMP). **Results:** Asthmatic children were clustered into groups based on their transcriptomic (n=5 clusters), metabolomic (n=5 clusters) and integrated (n=5 clusters) profiles. There was high concordance between the clustering of children under all three profiles. The clusters could be broadly defined by asthma severity (lung function and exacerbations), age of onset, atopy and vitamin D status. The clusters were most clearly defined using the integrated profile. Interrogation of the integrated profiles identified 268 genes and 69 metabolites as drivers of cluster formation, including both novel and established asthma associations. Pathways underlying phenotypic heterogeneity in asthma; including glycerophospholipid, sphingolipid, tryptophan and linoleic acid metabolism, were enriched among the genes and metabolite variants, and the variants could be annotated to disorders of lung function and immunity. Validation in CAMP identified four clusters with similar phenotypic patterns to those in CR. Moreover, the drivers of these clusters were replicated in CAMP on both an individual variant and pathway level. **Conclusions:** As one of the first and largest integrative omics studies of childhood asthma this study demonstrates and validates the use of omic profiles to classify asthma sub-phenotypes. Interrogation of the profiles reveals biological mechanisms discriminating between the subtypes, paving the way for future biomarker development and targeted therapeutics.

3143T

Proteomic peptide screening of dried blood spots for primary immunodeficiency, Wilson disease and cystinosis. *S. Hahn[1,2], S. Jung[2], J. Whiteaker[3], L. Zhao[3], H. Yoo[4], T. Torgerson[1,2], H. Ochs[1,2], W. Gahl[5], A. Paulovich[3].* 1) Department of Pediatrics, University of Washington, Seattle Children's Hospital, Seattle, WA; 2) Seattle Children's Hospital Research Institute, Seattle, WA; 3) Fred Hutchison Cancer Research Center, Seattle, WA; 4) Department of Pediatrics, Ulsan University College of Medicine, Asan Medical Center, Seoul, Korea; 5) National Human Genome Research Institute, National Institute of Health, Bethesda, MD.

Background: Newborn screening (NBS) has proven to be highly effective at reducing healthcare costs, improving outcomes, and avoiding long-term disability in affected children. Unfortunately, many attractive candidates for NBS are characterized by the absence or modification of proteins for which there are currently no cost-effective screening methods. Among these are Primary Immunodeficiency Diseases (PIDD), Wilson Disease (WD) and Cystinosis. Objectives: Our goal is to develop a high-throughput, multiplex, proteomic screening assay for a variety of genetic conditions. We hypothesized that peptide immunoaffinity enrichment coupled with Liquid Chromatography-Selected Reaction Monitoring-Mass Spectrometry (immuno-SRM) can be utilized as a rapid, inexpensive approach to detect CD3ε (for SCID), WASP (for WAS), BTK (for XLA), ATP7B (for WD) and Cystinosin (for Cystinosis) in Dried Blood Spots (DBS).**Methods:**. Several candidate peptides were screened by *In Silico*, BLAST search, then selected based on detectability and response in LC-MS/MS analyses followed by polyclonal antibody generation for enrichment. We analyzed target peptides for BTK, CD3ε, WASP, ATP7B and Cystinosin in DBS using SRM mode with 6500 QTRAP (ABSCIEX). Blood samples from patients were spotted on filter paper, dried and stored at -80 °C until use.**Results:** Two BTK targeted peptides were absent or markedly reduced in two XLA patients compared to normal controls, while the other target peptides designed to detect SCID, WAS in the XLA patients were normal. Cystinosin peptide was absent in two Cystinosis patients (57-kb del homozygotes). Lastly, the results of DBS samples from 5 controls and 7 confirmed WD patients showed that the assay readily distinguished affected from controls. The assay showed a linear response ($r^2 = 0.99$) spanning the peptide concentration of 27 to 16,765 pmol/L (0.7 to 417 femtomoles). The median CV for all points on the response curve was 7.8%. Intra- and interassay imprecision from the normal control were 8.89% CV and 7.57% CV, respectively. ATP7B protein marker in DBS at various temperatures was stable for a week.**Conclusions:** This is very compelling evidence that immuno-SRM assays can be used as a highly sensitive platform for DBS protein analysis in the low-picomolar range. When multiplexed, there is a significant potential to screen various congenital disorders lacking specific protein markers.

3144F

Diagnosis of adenylosuccinate lyase deficiency by metabolomic profiling in plasma reveals a phenotypic spectrum. *T. Donti[1], G. Cappuccio[1,2], L. Hubert[1], J. Neira[1], P. Atwal[1], M. Miller[1], A. Cardon[3], V. Sutton[1], B. Porter[4], F. Baumer[4], M. Wangler[1], Q. Sun[1], L. Emrick[1,3], S. Elsea[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Translational Medical Sciences, Section of Pediatrics, Federico II University, Naples, Italy; 3) Section of Pediatric Neurology and Neuroscience, Baylor College of Medicine, Houston, TX; 4) Stanford Medical School, Stanford, CA.

   Adenylosuccinate lyase (ADSL) deficiency is a rare autosomal recessive neurometabolic disorder that presents with a broad-spectrum of neurological and physiological symptoms that typically include intellectual disability, seizures, and global developmental delays. Mildly affected individuals present mainly with autistic features. *ADSL* produces an enzyme, adenylosuccinate lyase, with binary molecular roles in *de novo* purine synthesis and purine nucleotide recycling. The biochemical phenotype of ADSL deficiency, accumulation of SAICAR and succinyladenosine (S-Ado) in biofluids of affected individuals, serves as the traditional approach for diagnosis, with targeted quantitative urine purine and CSF analysis employed as the predominant method of detection. Diagnosis of rare neurometabolic disorders typically requires the use of targeted analyte analyses, limiting the use of this approach in cases without typical presentation of disease. Using an alternative method, untargeted metabolomic profiling, an analytical scheme capable of generating semi-quantitative z-score values for over 900 unique compounds in a single analysis of a plasma specimen, we diagnosed of ADSL deficiency in 3 individuals from 3 families presenting with a broad spectrum of features ranging from mild to severe intellectual disability with or without seizures. A fourth child was identified in one of these families using targeted purine analysis, further illustrating the phenotypic variability, even within the same family. These findings were confirmed with targeted quantitative biochemical analysis and molecular genetic testing revealing novel and known mutations in *ADSL*. The use of untargeted metabolomic screening for neurometabolic disorders has the potential to identify patients with enzyme deficiencies that may present on the mild end of the reported spectrum, reducing time and cost of diagnosis. Undifferentiated phenotypes, such as intellectual disability, seizures, and hypotonia, are common among many inborn errors of metabolism and diagnosis dependent upon specific, targeted testing has classically stood as a hurdle in the initial diagnosis and subsequent treatment. The findings presented here demonstrate the clinical utility of metabolomic profiling in the diagnosis of ADSL deficiency and highlights the potential of this technology in the diagnostic evaluation of individuals with neurologic phenotypes. .

3145W

Characterization of the poly-T variants in the *TOMM40* gene using PacBio long reads. *A. Sethuraman, K. Eng, J. Gu, P. Baybayan, S. Kujawa.* Pacific Biosciences, Menlo Park, CA.

   Genes associated with several neurological disorders have been shown to be highly polymorphic. Targeted sequencing of these genes using NGS technologies is a powerful way to increase the cost-effectiveness of variant discovery and detection. However, for a comprehensive view of these target genes, it is necessary to have complete and uniform coverage across regions of interest. Unfortunately, short-read sequencing technologies are not ideal for these types of studies as they are prone to mis-mapping and often fail to span repetitive regions. Targeted sequencing with PacBio long reads provides the unique advantage of single-molecule observations of complex genomic regions. PacBio long reads not only provide continuous sequence data though polymorphic or repetitive regions, but also have no GC bias.   Here we describe the characterization of the poly-T locus in *TOMM40*, a gene known to be associated with progression to Alzheimer's, using PacBio long reads. Probes were designed to capture a 20 kb region comprising the *TOMM40* and *ApoE* genes. Target regions were captured in multiple cell lines and sequencing libraries made using standard sample preparation methods. We will present our results on the poly-T structural variants that we observed in *TOMM40* in these cell lines. We will also present our results on probe design optimization and barcoding strategies for a cost-effective solution.

3146T

Identification of large pathogenic structural variants using next generation genome mapping technology. *E. Vilain, H. Barseghyan, W. Tang, M. Almalvez, E. Douine, H. Lee, S. Nelson.* Human Genetics, UCLA School of Medicine, Los Angeles, CA.

Whole exome/genome sequencing (WES/WGS) has become more routine and affordable with the capability of reliably identifying single nucleotide variants (SNVs), small insertions and deletions (INDELs). However, this approach does not allow the identification of a full spectrum of genetic variants in the human genome. Primarily, WES/WGS relies on short-read sequences that are mapped to a reference human genome without genome assembly. These sequencing methods miss on identification of large insertions/deletions/copy number variations in repetitive regions of the genome and are not capable of easily detecting other structural variations (SVs) such as inversions and translocations. In order to overcome these limitations we used Next Generation Mapping through nanochannel arrays (BioNano Irys System) that produce high-resolution images of fluorescently labeled native-state DNA molecules (up to 1Mb in size) for genome assembly and SV detection on a series of patients diagnosed with Duchenne Muscular Dystrophy (DMD). Multiplex ligation-dependent probe amplification (MLPA) was previously performed on all patients and several carrier mothers to identify which exons were deleted or duplicated. Using this strategy we successfully identified the intronic breakpoints within the *Dystrophin* gene where patients had a deletion or duplication. The identified deletions ranged from 30kb – 200kb, and insertions ranged from 10kb – 150kb in size both encompassing several exons and introns. Next Generation Mapping also successfully identified heterozygous SVs in carrier mothers of DMD patients indicating the ability of the method to distinguish between homozygous and heterozygous variants. Along with SVs identified in *Dystrophin* gene we have also on average identified 800 insertions, 500 deletions, 100 inversions and 50 translocations. To filter out likely benign SVs we used a database containing SVs from healthy individuals. Although we observe that SVs are less common than SNVs and INDELs in population they account for a significant fraction of genetic variation. Next Generation Mapping is poised to identify potential pathogenic SVs in patients with negative WES/WGS and to provide, in a single test, a combination of analysis of copy number variants and of large structural variants. This approach can be incorporated in the clinical diagnostic strategy of undiagnosed patients.

3147F

Unraveling the genetic basis of atypical cerebral palsy. *I. Blydt-Hansen[1], A. Matthews[2-4], M. Tarailo-Graovac[2-4], B. Al-Jabri[5], H. Vallance[6], G. Sinclair[6], R. Salvarinova[1], C.J. Ross[3,4], W.W. Wasserman[2-4], H. Goez[5], G. Horvath[1], C.D. van Karnebeek[1,2,4].* 1) BC Children's Hospital, Division of Biochemical Diseases, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Child & Family Research Institute, Vancouver, BC, Canada; 5) Department of Pediatrics, Stollery Children's Hospital, University of Alberta, Edmonton, Alberta, Canada; 6) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

Background: Cerebral palsy (CP) is a debilitating life-long condition, often presenting with co-morbidities such as intellectual disability, movement disorder or epilepsy. Though classically deemed environmental (eg oxygen deprivation pre/perinatal brain insult), monogenic conditions can mimic CP. Timely diagnosis is essential for proper management and counseling, especially in the case of neurometabolic diseases that mimic CP and are amenable to treatment. **Aim**: To report all families with atypical CP (additional symptoms including intellectual disability and dyskinesias) in whom TIDEX whole-exome sequencing (WES) identified a monogenic disorder during the period 2011-2016. **Methods:** WES analysis with a customized bioinformatics pipeline was initiated in 23 exquisitely phenotyped families with unexplained atypical CP. Variants were filtered against public and in-house databases to identify rare variants which were also predicted to be damaging by *in silco* tools. Accounting for the growing number of CP genes, we performed targeted analysis, ie virtual gene panel of known and suspected CP and movement disorder genes, with subsequent untargeted analysis of the whole exome using our TIDEX pipeline. **Results:** Contributing to the neurogenetic presentation, an autosomal de novo mutation was found in 14 families, a recessive mutation in 6, a hemizygous variant in 2 and an X-linked de novo mutation in 1. In the cases of CSTB and PAK3 deficiency, targeted treatment with neurotransmitter supplements was enabled by a genomic diagnosis. Furthermore, our findings expand the phenotype of three conditions. In at least three other cases, the affected individual's described phenotypes appeared to be caused by variants in two distinct genes (blended phenotypes). **Discussion**: Traditionally, functional or model organism studies are preferred to establish a gene as disease causing. However, such studies are costly, time consuming and not feasible for all CP genes. Rather, to prove disease causality of unreported CP genes, identifying unrelated families with similar phenotypes and different variants in the same gene will be required. Further identification of CP genes may improve diagnosis, reveal treatment targets and unravel pathophysiology of this complex and heterogeneous disease. As our understanding of atypical CP improves, we will be able to better understand early brain development and inform families regarding disease progression and genetic counseling decisions.

3148W

Improved sensitivity for clinically relevant variants using orthogonal sequencing. *J. Thompson, E. White, N. Chennagiri, N. Vena, A. Seman, S. Hosseini, B. Wu.* Res & Development, Claritas Genomics, Cambridge, MA.

We have previously shown that using independent Next Generation Sequencing platforms provided improved sensitivity for detecting exomic variants in the well-defined regions of standard reference samples. However, it is not straightforward to compare results in other parts of the exome and extrapolate to clinically relevant variants. To assess the ability of dual platforms to improve results, we examined the likely/possibly pathogenic variants found in more than 50 exome-based Regions of Interest. As previously observed~90% of all variants were orthogonally confirmed with identical sequencing results found on both the Illumina NextSeq and the Ion Torrent Proton platforms. Variants reported to patients included those that were sufficiently well studied as to be classified as Pathogenic/Likely Pathogenic by ACMG guidelines. In addition, Variants of Uncertain Significance (VUS) were found in genes known to be disease causing but the specific variants could not be unambiguously classified as pathogenic were also reported. Among the P/LP/VUSs, most were orthogonally confirmed but there were also examples of variants found exclusively on the NextSeq or exclusively on the Proton. If only a single platform had been used, these variants would have been missed and not reported to the patient. The number of genes that are 100% covered at >20x increased 42% when the Proton coverage was added to the Illumina analysis. Thus, this method is shown to deliver confirmed variants in a timely fashion while expanding the sensitivity for identifying rare, disease-relevant variants.

3149T

Improved performance of whole genome sequencing detects a *SYNGAP1* mutation in siblings with epilepsy with myoclonic-atonic seizures and photosensitivity. *M.J. Cowley[1,2], Y. Liu[3], K.L. Oliver[3,5], G. Carvill[4], C. Myers[4], V. Gayevskiy[1], M. Delatycki[6], Y. Zhu[7], K. Ying[1], D. Miller[1], P. Morris[1], A. Statham[1], H. Mefford[4], M.F. Buckley[8], S.F. Berkovic[5,6], M. Bahlo[3], I. Scheffer[5,6,9,10], M.E. Dinger[1,2], T. Roscioli[1,2,11].* 1) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia; 2) St Vincent's Clinical School, University of New South Wales, Darlinghurst, Australia; 3) Population Health and Immunity Division, Walter and Eliza Hall Institute, Melbourne, Australia; 4) University of Washington Department of Pediatrics, Genome Sciences, Seattle, USA; 5) Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Heidelberg, Australia; 6) Austin Health, Melbourne, Australia; 7) Department of Medical Genetics, Royal North Shore Hospital; 8) SEALS laboratory, Prince of Wales Hospital, Randwick, NSW, Australia; 9) Florey Institute, Melbourne, Australia; 10) Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Australia; 11) Department of Medical Genetics, Sydney Children's Hospital, NSW, Australia.

We have recently achieved ISO15189 accreditation for clinical whole genome sequencing (WGS) for diagnosing rare Monogenic Disease, using the Illumina HiSeq X Ten platform, the first facility in the Southern Hemisphere to achieve this. Here we report WGS results in a non-consanguineous family of Ashkenazi-Jewish descent with two sisters with epilepsy with myoclonic-atonic seizures and photosensitivity. During an extensive 5-year diagnostic odyssey, they were investigated with targeted sequencing of known epilepsy genes using Molecular Inversion Probes (MIPs), followed by Whole Exome Sequencing (WES), with no aetiology identified by either technology. Using WGS, we identified a 13-bp heterozygous duplication in *SYNGAP1* (chr6:g.33400507_33400519dup) which was only present in both children, consistent with inheritance from a gonadal mosaic parent. Heterozygous mutations in *SYNGAP1* are consistent with the epileptic encephalopathy phenotype. We present an analysis of the events that confounded this molecular diagnosis. The reads containing the duplication were unmapped in the MIPs data using BWA v0.5.9, which is likely due to the large duplication at the end of the 101bp read. By increasing the gap extension parameter (-e=20) we were able to detect the mutation within the previously unmapped reads. The WES analysis used the hg19 reference genome, which contains 7 alternative MHC haplotypes. One of these contains *SYNGAP1*, resulting in two identical copies in the reference genome, and most reads having a mapping quality of 0. Mapping the WES reads to GRCh37 or with BWA-MEM v0.7.12 to GRCh38, which supports alternate haplotypes, revealed the mutation. In the WGS analysis, the longer 150bp reads, and the use of BWA-MEM v0.7.10 with the GRCh37d5 reference genome allowed the reads to map appropriately, and led ultimately to the identification of the duplication. This case study highlights that technical confounders can lead to missed diagnoses, and that the appropriate aligner for the reference genome should be used. We propose that alternative analysis pipelines, and/or WGS should be performed where a diagnosis has not been achieved using targeted sequencing and that all forms of possible inheritance should be considered.

3150F

Identification of novel Ras superfamily nucleotide-binding protein alterations in individuals with brain malformation using Missense-Depleted Regions (MDRs) in population exomes. *J. Shieh[1], X. Ge[1], H. Gong[1], K. Dumas[1], J. Litwin[1], J.J. Phillips[1], Q. Waisfisz[2], M.M. Weiss[2], Y. Hendriks[2], S.F. Nelson[3], W.W. Grody[3], H. Lee[3], P-Y. Kwok[1].* 1) University of California, San Francisco, CA USA; 2) VU University Medical Center, Amsterdam, The Netherlands; 3) University of California, Los Angeles, CA USA.

Purpose: Genomic sequence interpretation can miss clinically-relevant variants even with detailed medical genetics phenotyping of individuals. Missense variants are numerous in the exome, and affected genes may lack disease association. Methods: Here we apply population exome data to identify intragenic missense-depleted regions, MDRs, genome-wide that may be important in disease. We used MDR ranking to evaluate nondiagnostic patient exomes. Exome sequencing routinely yielded over 400 rare missense variants per individual. MDR ranking was then used to prioritize exome variants. With the addition of these these methods we identify a new gene association with pediatric brain malformation. Summary of Results: By applying these methods to nondiagnostic exome data, we identified a novel gene association for human brain malformation. We found novel *de novo* missense variants that affect the GDP/GTP-binding site of *ARF1* in three unrelated patients. Corresponding functional analysis indicates ARF1 GDP/GTP-activation, as measured by antibody pulldown, is affected by specific missense mutation associated with heterotopia. These findings expand the genetic pathway underpinning neurologic disease that classically includes *FLNA* and *ARFGEF2* and supports decreased ARF1/GEF function in disease. Using functional ontology, top-ranked MDR-containing genes were highly enriched for nucleotide-binding function, suggesting these as potential candidates for human disease. This is further supported by the recent identification of such genes in disease. We suggest routine consideration of MDR in exome interpretation may help identify strong genetic factors for rare undiagnosed conditions, and potentially infertility and embryonic conditions contributing to preterm loss.

3151W

Whole genome methylation profiles of single neurons. *R. Mulqueen[1], J. Sinnamon[2], G. Mandel[2], B.J. O'Roak[1], A. Adey[1,3].* 1) Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR 97239, USA; 2) Vollum Institute, Oregon Health & Science University, Portland, OR 97239, USA; 3) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR 97239, USA.

DNA methylation is a stable epigenetic modification with roles in regulation and differentiation. During development methylation is critical in defining neuronal cell function. However, methods used in previous studies are inherently limited; samples prepared using thousands of cells, shroud population substructure and neuronal heterogeneity. To address this shortcoming, we modified and optimized a single-cell whole genome bisulfite sequencing method (Smallwood et al. 2014; Farlik et al. 2015). Nuclei are flow sorted into individual wells, bisulfite converted and adapter tagged by random nonamer priming. Libraries are then amplified, barcoded and sequenced on the Illumina NextSeq 500 platform. By decreasing reaction volumes, using a dual indexing strategy, and reducing our protocol to a single well reaction, we improved upon previously described methods in terms of cost and scalability of library preparation. This renders cell population specific methylation analyses more tractable. We first applied our protocol on primary cultures of wild type mouse hippocampal neurons. Methylation libraries constructed for 92 single neurons had a 78.2% (72/92 nuclei) quality control pass rate with stringent filters. Counts as low as 300K uniquely aligned reads were sufficient to reproduce expected global methylation patterns and quantitation trends over CG islands and gene bodies, corroborating our strategy. We compared these to libraries constructed from a model of Rett Syndrome (RTT [MIM 312750]), a neurodevelopmental disorder caused primarily by mutations in the gene encoding epigenomic regulatory protein *methyl-CpG binding protein 2* (MECP2; Amir et al. 1999). The mutant neuronal library pool of 92 single neurons passed quality control at a rate of 71.7% (66/92 nuclei) and recapitulated known RTT methylation patterns, with no significant change in global methylation rate compared to wild type (Student's t-test, p>0.35; Gabel et al. 2015). We next intend to leverage the single cell information we have produced to quantify differences in locus specific variability and between neuronal epigenetic subtypes that are not possible to investigate with bulk preparation methods. With our high cell count, low coverage approach, we will assess unresolved changes in mean methylation levels and variability in a genomic locus and cell population specific manner. We plan to use this method as a nuanced approach to elucidate epigenomic dysregulation in complex tissue.

3152T

Deciphering the complex interplay between genetics, metabolomics and proteomics in Alzheimer's disease. *P. Proitsi[1,2], M. Kim[3], S. Policicchio[1], J. Xu[3], S. Newhouse[1,4], L. Whiley[3], M.K. Lupton[5], K. Ling[1], C. Johnston[1,4], M. Sattlecker[1,4], S. Lovestone[6], C. Legido-Quigley*[3], J.F. Powell[1], R.J.B. Dobson[1,4], The AddNeuromed consortium.* 1) King's College London, IoPPN, London, UK; 2) MRC Unit for Lifelong Health and Ageing (LHA), UCL, UK; 3) King's College London, Institute of Pharmaceutical Science, London, UK; 4) NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust, UK; 5) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 6) Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK.

Alzheimer's disease (AD) is a devastating illness, with no effective treatments and lack of an early diagnosis. Blood metabolites have emerged as promising AD biomarkers. They are easily accessible, functional intermediate phenotypes that represent a molecular "fingerprint" in health and disease, and their integration with genetic and other omics data has increased our understanding on the interplay of multiple layers of biological organization. We recently performed a comprehensive blood lipidomics study on ~300 AD patients and controls replicating previous as well as reporting novel associations between blood lipids and AD clinical diagnosis, disease progression and brain atrophy. The aim of this study was to scrutinize the complex interplay between genetic variation (>6.4M SNPs), blood lipids (2225 features generated using Ultra-Performance Liquid Chromatography/Mass Spectrometry) and protein levels (1000 proteins measured using a Slow Off-rate Modified Aptamer (SOMAmer)–based capture array ("SOMAscan", SomaLogic, Inc)) in 180 AD patients, 189 controls and 38 people with Mild Cognitive Impairment. We first 1) performed an mGWA analysis to survey regions of the genome associated with blood lipids. We then sought to 2) identify blood lipids associated with established AD genetic variants, 3) identify loci that influence AD-related blood lipids and 4) investigate statistical colocalization of genetic risk variants for AD and for AD lipids. Next, we 5) integrated lipid and genetic data with blood protein data for 198 individuals to explore lipidomic-proteomic associations and investigate whether genetic variation driving AD lipids is associated with protein levels. Finally, 6) we performed Mendelian Randomization to determine whether blood lipids lie in the causal pathway to AD. We report novel associations between a number of loci and blood lipids. We also report associations between 4 AD risk SNPs and Phosphatidylcholine and Triglyceride molecules. We found >20 AD lipids to be associated with at least one locus at $P<5*10^{-8}$ with 11 passing Bonferroni correction. Integration of lipidomic and proteomic data revealed strong associations with proteins including Lp-PLA2 and Adiponectin and MR analyses point to causal links between lipids and AD. Ours is the first study to examine the interplay of genetics, blood lipidomics and proteomics in AD and demonstrates the utility of using multi-omics integrative approaches to understand the mechanisms underlying AD.

3153F

Development of a protocol of RNA extraction from skin biopsy for RNA Sequencing studies and its application in the identification of biomarkers implicated in painful and painless Idiopathic Peripheral Neuropathy. *S. Santoro[1], I.D. Lopez[2,10], M. Sorosina[1], A. Zauli[1], A.M. Osiceanu[1], L. Grevendonk[1], S. Peroni[1], R. Lombardi[3], D. Cazzato[3], M. Marchi[3], C.G. Faber[4], M. Sopacua[4], M.M. Gerrits[5], R. Almomani[5], J.G. Hoeijmakers[4], I.S. Merkies[4,5], H. Fadavi[6], R.A. Malik[7], D. Ziegler[8], G. Boenhof[8], S.D. Dib-Hajj[9], S.G. Waxman[9], G. Comi[1,2,10], A. Quattrini[2,10], G. Lauria[3], F. Martinelli Boneschi[1,10].* 1) Laboratory of Human Genetics of Neurological Disorders, San Raffaele Scientific Institute, Milan, IT; 2) Laboratory of Neuropathology, San Raffaele Scientific Institute, Milan, IT; 3) Neuroalgology Unit, Neurological Institute Carlo Besta, Milan, IT; 4) Department of Neurology, Maastricht University Medical Center, Maastricht, NL; 5) Clinical Genetics, Maastricht University Medical Centre, Maastricht, NL; 6) Centre for Endocrinology and Diabetes, University of Manchester and Central Manchester NHS Foundation Trust, Manchester, UK; 7) Division of Cardiovascular Medicine, University of Manchester, Manchester, UK; 8) German Diabetes Center, Medical Faculty Heinrich Heine University, Düsseldorf, DE; 9) Department of Neurology, Yale University School of Medicine, New Haven, CT, USA; 10) Department of Neurology, San Raffaele Scientific Institute, Milan, IT.

Neuropathic pain is a frequent clinical symptom occurring in the context of Peripheral Neuropathy (PN) and causing a significant impact on patients' quality of life. However, it is unpredictable, since not all of the individuals with PN develop pain. We hypothesized that the identification of a transcriptomic signature in skin components from fresh human skin biopsy and peripheral blood could be used to better understand the pathophysiology of neuropathic pain and to identify individuals at risk. With this purpose, we developed and optimized a protocol for RNA extraction from human skin biopsy, usually collected to make diagnosis, to explore the transcriptome using mRNA Sequencing in 4 painful and 4 painless PN patients and 6 age- and sex-matched healthy controls. We developed a protocol characterized by the following steps: 1) collection of biopsy and storage in a RNA stabilizing solution; 2) embedding and storage in liquid nitrogen; 3) cryosection; 4) staining with hematoxylin and eosin; 5) microdissection of the following tissues: A) border between epidermis and dermis - B) dermis - C) glands - D) whole biopsy; 6) RNA extraction. From 11 processed skin biopsies, we obtained the following mean quantities and DV_{200} values of RNA respectively for the 4 tissues: A) 59 ng and 88%, B) 36 ng and 82%, C) 22 ng and 78%, D) 52 ng and 88%. Although the quantity and quality of extracted RNA was poor, they fulfilled recommended parameters for RNA-Seq. Extracted RNA was enriched using the Illumina TruSeq RNA Access Library Prep Kit and paired-end sequenced (2x101bp) on the Illumina HiSeq 2500 platform. Quality control parameters from the sequencing of a first biopsy were successful. We obtained a mean of $9x10^6$ reads per tissue with 96% of total aligned reads on hg19 and 75% on coding regions. As expected, a high correlation coefficient was found between different skin tissues with the lowest between A and C (0.929) and the highest between B and C (0.978). Sequencing on additional samples, as well as analyses of differential gene expression comparing the different phenotypes in different tissues and peripheral blood, are ongoing. We conclude that the implemented protocol is efficient in extracting RNA from skin biopsy of enough quality to be used for RNA Sequencing and that our project will help to identify novel biomarkers of disease and potential targets for focused pharmacological approaches.

3154W

High-throughput functional analysis of *PTEN* variants reveals geno-type-phenotype relationships. *T.L. Mighell[1,2], S. Evans[2], B.J. O'Roak[1,2].* 1) Neuroscience Graduate Program, Oregon Health & Science University, Portland, OR; 2) Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR.

The PI3K/mTOR pathway is a critical regulator of cell survival, growth, and proliferation. A crucial negative regulator of this pathway is phosphatase and tensin homolog (PTEN [MIM 601728]), which is frequently mutated in tumors. Interestingly, *PTEN* mutations have also been observed in individuals with autism and macrocephaly ([MIM 605309]), as well as various phenotypic presentations collectively known as PTEN hamartoma tumor syndrome (PHTS, e.g. [MIM 158350]). Despite its clear relevance to human health, we are still unable to predict the phenotypic outcome for the majority of *PTEN* genotypes. Some recent studies suggest that a *subset* of autism-associated variants retain greater phosphatase activity than cancer- and PHTS-associated variants. Making use of a *humanized* yeast model (Rodríguez-Escudero *et al.*, 2005), we seek to fully explore this hypothesis by performing saturation mutagenesis of *PTEN* and characterizing the phosphatase activity of thousands of mutations in parallel. In yeast expressing human *PIK3CA*, toxicity results from the conversion of the phosphatidylinositol (4,5)-bisphosphate (PIP_2) pool to PIP_3. Co-expression of PTEN rescues yeast growth by restoring the essential PIP_2 pool. However, catalytically dead PTEN variants are unable to rescue, thereby coupling the catalytic competence of a PTEN variant to cell growth. We adapted this system for highly multiplex mutation evaluation in an inducible competitive liquid culture system, thus permitting the assessment of thousands of variants in a single experiment using massively parallel DNA sequencing as readout. We have directly synthesized oligonucleotide probes encoding all possible codons at each position of PTEN (n=25,389) for cloning in a yeast expression vector. Using an *in vitro* recombination approach, mutation probes are introduced in parallel onto an otherwise wild-type *PTEN* background. We show in pilot experiments expected results for known catalytically dead and hypomorphic mutations. Our data allows for quantitative description of enzymatic activity. By subjecting a PTEN saturation mutagenesis library to this selection experiment, we will illuminate most or all residues that are important for phosphatase activity. This will inform understanding of basic protein properties of PTEN and will assist in the interpretation of variants observed in human genomes. Coupling genomic discovery with high-throughput functional analyses is paramount to advancing precision genomic medicine. .

3155T

Creating induced neurons directly from human adult fibroblasts provides a model system to elucidate the molecular effects of the 22q11.2 deletion. *C. Purmann[1,2], K. Tanabe[3], Y. Zhang[2], C.E. Ang[3], J. Hallmayer[1], M. Wernig[3], A.E. Urban[1,2].* 1) Department of Psychiatry & Behavioral Sciences, Stanford University, Palo Alto, CA; 2) Department of Genetics, Stanford University, Palo Alto, CA; 3) Institute for Stem Cell Biology and Regenerative Medicine and Department of Pathology, Stanford University, Palo Alto, CA.

22q11.2 Deletion Syndrome is caused by a common chromosome abnormality and a prominent point of entry for the analysis of the genetic basis of schizophrenia and autism. However, many dozen genes are affected by the deletion alone, and the effects are likely to be highly cell-type specific. Here, we present a novel method that allows for the direct induction of neuronal cells (induced Neurons, iN) from adult human fibroblast cells (previously only possible from perinatal fibroblasts). We applied this novel protocol to fibroblasts from 22q11.2 deletion patients and carried out extensive genomic and epigenomic analyses of the resulting iNs. We included fibroblasts from three 22q11.2 deletion patients and three control probands. Reprogramming into iNs via recombinant virus resulted in cells with neuronal phenotype and staining for relevant markers such as MAP2 and PSA-NCAM after approximately 21 days of culture, compared to several months for the standard route via iPSCs. RNA and chromatin were collected at the fibroblast and iN stages. Transcriptome-patterns were analyzed using RNA-Seq (also for micro-RNAs), and chromatin states were analyzed using ATAC-Seq. Established data analysis pipelines for each data type were employed. On both levels of molecular analysis, iN cells carrying the 22q11.2 deletion showed clear deletion-specific dysregulation. Genes expressed in iNs and within the deletion-boundaries were down-regulated while genome-wide both up- and down-regulation could be observed. Chromatin conformation was also clearly impacted by the presence of the 22q11.2 deletion. Principal-component-analysis showed clear clustering of patient vs. control iNs using the overall ATAC-Seq signals. For individual chromatin regions both increased and decreased chromatin compaction is observed. Our results show that with the iN method it is now possible to generate neuronal cells directly from adult human fibroblasts, in which a defined genetic lesion will lead to a well-observable molecular phenotype. For 22q11.2 Deletion Syndrome, we present gene-by-gene effects of the deletion on gene expression in human neuronal cells as well as the observation that the deletion also affects a critical level of the epigenomic regulation of gene activity.

3156F

Next-generation mapping, a highly sensitive and accurate method for interrogation of clinically relevant structural variation. *A. Hastie, A.W. Pang, J. Lee, T. Anantharaman, W. Andrews, M. Saghbini, H. Cao.* BioNano Genomics, San Diego, CA.

   Structurally complex loci underlie many diseases. These loci can be very challenging to resolve by currently available methods such as karyotyping, Arrays, PCR-based tests and next-generation sequencing (NGS). Next-generation mapping (NGM) using the BioNano Genomics Irys® System offers a high-throughput, genome-wide method, based on direct visualization of extremely long genomic fragments, able to interrogate genome structural variants (SVs) in the range of one kilobase pairs to hundreds of kilobase pairs. The Irys System uses extremely long range information to span interspersed and even long tandem repeats making it suitable for elucidating the structure and copy number of complex regions of the human genome, such as complex pseudogene and paralogous gene families. Clinically relevant regions often contain genes with paralogs and other complex repetitive structures complicating the interpretation of data and diagnosis of disease. We have used two hydatidiform mole cell lines (homozygous human cell lines) to simulate a diploid genome and measure sensitivity to homozygous and heterozygous SV detection. We found 87% sensitivity to heterozygous SVs and 99% sensitivity to homozygous SVs (2 kbp and up). We also present several examples of SVs at complex genetic loci by NGM, such as those made up of tandem repeats, paralogous gene families, and loci flanked by segmental duplications. Examples of variants at tandem repeats are tRNAs, kringle IV, and *D4Z4*. An important variable length tandem repeat is *D4Z4*, which is associated with facioscapulohumeral muscular dystrophy (FSHD). FSHD muscular dystrophy is strongly associated with a low copy number (< 10 units), occurring in 95% of FSHD cases. Copy number of tandem repeats is extremely hard to measure accurately with available methods, but we show that NGM using the Irys System can accurately measure the copy number of *D4Z4* as well as haplotype the repeat array and differentiate the gene from a paralog that occurs on another chromosome. A second class of complex SVs involves genes with paralogs such as amylase and *UGT2B17*, two genes whose copy number have been shown to be associated with human health. We demonstrate that NGM using the Irys System is proving to be a highly accurate and sensitive method for detection of clinically relevant SVs.

3157W

Mendeliome sequencing increases the diagnostic yield in patients with unexplained intellectual disability by 30% (a single center experience). *A. Rump[1], L. Mackenroth[1], A. Kahlert[1], J. Lemke[2], E. Krajcir[1], A. Kruger[1], F. Kuhlee[1], F. Stubner[1], A. Tzschach[1], E. Schrock[1], N. Di Donato[1].* 1) Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany; 2) Institute for Human Genetics, University Clinic Leipzig, Leipzig, Germany.

   Introduction Advances in deciphering the genetic causes for impaired development have always been accompanied by advances in technology. In the past, array-CGH technology has boosted the detection rate significantly; but up to 50% of children with developmental delay still remain undiagnosed. Since the latest technological advance, next generation sequencing (NGS), has the power to improve the diagnostic yield tremendously, we applied this method to 57 index patients with developmental delay or intellectual disability (ID) and pre-excluded genomic imbalances. **Method** Genomic DNA samples of 28 parent-child trios plus 29 individuals were analyzed for mutations in 4813 genes, using the TruSightOne gene panel on the MiSeq platform (Illumina, San Diego, CA). After sequencing with median target coverage of 80-fold and mapping to hg19, sequence variants were called by two independent platforms: the GATK pipeline installed on the MiSeq and the CLC Biomedical Genomics platform (Qiagen, Hilden, Germany). All *de novo* variants, homozygous variants with heterozygosity in both parents and compound heterozygous variants were analyzed as well as hemizygous variants in male patients. These variants were screened for clinical und molecular concordance (i.e. disease-association of the gene, published mutation). In addition, all splice-relevant variants, all frameshift and nonsense-mutations were screened for clinical concordance. The results were discussed in a team of clinicians and molecular geneticists, relevant variants were validated by Sanger-sequencing. **Results** Using the mendeliome in a diagnostic setting, we established a diagnosis in 16 of the 57 index patients (28%). For seven further patients, we found one or two possibly causative candidates (13%). 5 patients (9%) showed incidental findings which either made treatment or surveillance necessary (homozygous MUTYH-mutations, SDHA-mutation) or led to an increased risk for a recessive disease in children (PAH- or CFTR-mutations). **Conclusion** Mendeliome next generation sequencing significantly increases the diagnostic yield in patients with ID unsolved by previous routine testing (array-CGH, conventional karyotyping). However, variant interpretation remains challenging and requires standardized procedures for using the new technology in a standard diagnostic setting.

3158T

RNA-seq, an important tool do discover non-exonic disease causing variants. *L. Kremer[1,2], D. Bader[3], C. Mertes[3], G. Pichler[4], T. Schwarzmayr[1], R. Kopajtich[1,2], A. Iuso[1,2], E. Konarikova[1,2], G. Kastenmüller[1], P. Lichtner[1], T. Strom[1,2], T. Meitinger[1,2], A. Lombes[5], D. Ghezzi[6], P. Freisinger[7], F. Distelmaier[8], J. Gagneur[3], H. Prokisch[1,2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Institute of Human Genetics, Technische Universität München, Munich, Germany; 3) Computational Biology, Department for Bioinformatics, Technische Universität München, Munich, Germany; 4) Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany; 5) Inserm UMR 1016, Institut Cochin, Paris, France; 6) Unit of Molecular Neurogenetics, Foundation IRCCS, Institute of Neurology "Carlo Besta", Milan, Italy; 7) Department of Pediatrics, Klinikum Reutlingen, 72764 Reutlingen, Germany; 8) Department of General Pediatrics, University Children's Hospital, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany.

Whole exome sequencing (WES) has proven to be a valuable tool for genetic diagnosis and is therefore underway to be implemented at an early stage in the diagnostic algorithm of genetic disorders. Yet, in at least 50% of the patients WES does not reveal the disease causing variant and a significant part of these variants might be regulatory. In principle, whole genome sequencing (WGS) approaches allow the discovery of all variants, coding and non-coding. The drastically increased number of rare variants detected via WGS, however, makes variant prioritization and interpretation challenging. In order to study regulatory variants by means of expression outliers, abnormal splicing, or monoallelic expression we applied RNA sequencing in 96 fibroblast cell lines derived from mitochondrial disease patients from which 50% remained without a molecular diagnosis after WES. Some already underwent whole genome sequencing (WGS). A systematic analysis identified in each sample on average 7 monoallelic expressed variants, 8 expression outliers and about 4 splice defects. As an example, follow-up experiments of an expression outlier identified an intronic variant to create an additional exon. The new exon resulted in a frameshift and a premature stop codon which leads to the absence of the encoded complex I assembly factor. For systematic validation of regulatory variants we performed quantitative proteomics. In case of the complex I assembly factor defect proteomic analysis revealed global complex I reduction, which was partially rescued upon lentiviral transduction of the wildtype allele. In another sample monoallelic RNA expression resulted in significantly decreased protein level being less than 3%, confirming the causal role of this variant for the clinical presentation of the patient. In conclusion, we show that detection of expression outliers, abnormal splicing and monoallelic expression by RNA-seq can unravel non-exonic pathogenic variants. By this approach we already provided a genetic diagnosis for 10% of unsolved WES cases. Validation of further candidates is ongoing. .

3159F

Rapid, single cell RNA-Seq from mature auditory hair cells. *P.T. Ranum[1], A.M. Sheffield[2,3], A.T. Goodwin[2], H. Liu[4], D.Z.Z He[4], R.J.H Smith[1,2,3].* 1) Interdisciplinary Graduate Program in Molecular & Cellular Biology, The University of Iowa Graduate College, University of Iowa, Iowa City, IA 52242, USA; 2) Molecular Otolaryngology and Renal Research Laboratories, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA; 3) Department of Otolaryngology, Head and Neck Surgery, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA; 4) Department of Biomedical Sciences, Creighton University School of Medicine, Omaha, NE 68178, USA.

Auditory hair cells are critical mechanosensory cells at the heart of the auditory system. The presence of healthy hair cells is a prerequisite for hearing function and their absence underlies many forms of deafness. Recent advances in ultra-low input RNA-Sequencing (RNA-Seq) technologies have enabled transcriptome profiling and pathway analysis of auditory hair cells. These experiments have the potential to provide novel insight into how hair cells respond to damaging insults and may allow us to identify pathways for drug development to prevent hair cell death. Here we describe the unpublished results of our work toward the development of a rapid and affordable single cell RNA-Seq protocol that can be used to isolate and characterize the transcriptomes of rare cell types like auditory hair cells. We have performed this protocol on mouse inner and outer auditory hair cells at P15 (the onset of hearing), P35, and P70 timepoints. Our data robustly differentiate between inner and outer hair cell types, and demonstrate the expected transcriptional variability between single cells. We have also sequenced groups of pooled cells to determine how cell number impacts transcript detection sensitivity and biological variation. We utilized pool/split and ERCC spike-in controls to assess and differentiate between biological and technical variation between samples. The resulting datasets provide a wealth of temporal and cell-type specific information about mature auditory hair cells. This information provides a foundation for studies designed to map the molecular pathways of auditory hair cells in disease states with the goal of identifying pathways that can be targeted to preserve hearing function.

3160W

A custom targeted next generation DNA sequencing capture panel for age-related macular degeneration. *R. Torene[1], F. Yang[1], Y. He[1], M. Sultan[2], T. Schlitt[2], A. Fernandez[2], C. Wache-Mainier[2], T.M. Morgan[1], J. Laramie[1], L. Ferrara[1], M. Roguska[1], J.F. Reilly[1], T. Dryja[1], P. Zamiri[1].* 1) 250 Massachusetts Avenue, Novartis Institutes for BioMedical Research, Cambridge, MA, USA; 2) Fabrikstrasse 2, 4056, Novartis Institutes for BioMedical Research, Basel, Switzerland.

   Purpose: Age related macular degeneration (AMD) is the leading cause of legal blindness in the developed world. Early AMD, characterized by accumulation of drusen, often progresses to either neovascular AMD or geographic atrophy (GA), affecting about 1 million people in the US, and millions more worldwide. Twin studies have demonstrated a significant genetic influence, but whole genome sequencing is not yet cost-efficient. Thus, we designed an NGS hybridization capture solution to comprehensively identify variants related to AMD and applied it to 486 patients with GA.   **Methods:** We targeted the exons and flanking sequence for 68 loci known to be associated with AMD risk, AMD progression, or response to AMD therapies, as well as complement genes. Sequencing was performed using an Agilent SureSelect custom capture approach with 10,698 probes covering 693 kb of sequence. The captured DNA was then sequenced using the Illumina HiSeq 2500 platform. DNA from 486 patients with GA were sequenced with an average read depth in targeted regions of about 50X and concordance with TaqMan candidate SNP genotyping of >99.5%.   **Results:** The targeted NGS data were used to calculate polygenic risk scores (PRS) based on 19 SNPs previously implicated in AMD by GWAS. The PRS was significantly higher in the patients with GA than in individuals without AMD ($p < 0.01$). We also used the AMD targeted panel to identify carriers for Stargardt's variants in the *ABCA4* gene. Although these patients can have an AMD-like phenotype, they may not respond to therapies that target AMD. Many of the GA patients were carriers for Stargardt's-associated variants. Besides these known variants, we identified rare (< 1% allelic frequency in 1000 Genomes) variants in *ABCA4* and other AMD relevant genes that are predicted to be damaging to the protein product. It remains to be determined whether these rare variants impact GA progression or treatment response.   **Conclusion:** We designed a targeted sequencing panel to enable cost-efficient (approximately $200/sample) genotyping of known variants as well as novel or rare variants in clinical trials. The data generated by this panel can be used for a variety of analyses to better understand AMD incidence, progression, and response to therapies.

3161T

Recommendations for best practices for genomics knowledge management in a cloud-based contextual knowledge hub. *X. Xu, J. Hull, J. Cohen, W. Fields, P. George, M. Kaganovich, D. Gross, D. Caplan.* SolveBio, New York, NY.

   Purpose: Next generation sequencing and other high throughput data generation technologies has lead to an explosion in the breadth and quantity of experimental and curated reference data. These datasets should be actionable assets and provide valuable insight into research and clinical questions. However, the lack of systems and technologies to index, standardize, and share data, as well as robust visualization and analysis platforms, results in analysis inefficiencies. For example, a highly trained clinical geneticist currently requires 54 minutes to interpret the the clinical significance of a single variant (Dewey 2014). This analysis and interpretation inefficiency is the major bottleneck in the NGS pipeline and is a serious barrier to the promise of precision medicine. The purpose of this project was to reduce the data use burden for geneticists, translational scientists, bioinformaticians, and clinicians. **Methodology**: We have created a cloud-based contextual knowledge hub that indexes structured and unstructured genomics data from publicly available databases, commercial data providers, and private user-generated sources. The hub implements many best practices for data management, including, but not limited to, data versioning and dating, fully documented data provenance, harmonization for content and field types. Additionally we have created a browser interface and an application programming interface (API) so that the data is accessible to a wide scope of users; from scientists to clinicians to enterprise software engineers. **Results & Future Directions**: As of June 2016, the knowledge hub has integrated 71 publicly available data sources into 1.39 billion independent records and 33.7 billion datapoints into one single platform. Here, we present our recommendations for best practices, current models, and technical and logistical challenges involved in genomics data integration and sharing. Our future efforts will focus on lowering the data integration barrier to allow for more serendipitous encounters between previously silo'ed data repositories to advance genetics research and development and clinical genetics.

3162F

Identification of novel pathways linked to Bardet-Biedl Syndrome (BBS) using a multi-omic strategy to implement biomarker discovery. *R. Davies[1], D. Kelberman[1], G. Otto[1], F. Minnecci[1], J. Gratton[1], J. Kenny[1], G. Madhan[1], X. Wang[1], S. Dowle[1], T. Peltola[2], T. Tukiainen[2], S. Kaski[2], C. Bacchelli[1], P.L. Beales[1].* 1) National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children, NHS Foundation Trust and University College London Institute of Child Health, London, UK; 2) Helsinki Institute for Information Technology HIIT, Aalto University, Helsinki, Finland.

Genetic diagnosis of rare diseases has accelerated exponentially over the last 5 years due to the advent of advances in next generation sequencing technologies. It is estimated that by 2020 the genes known to cause most remaining undiagnosed genetic disorders will be identified. Despite this, development of successful therapeutic methods for both rare and common disorders has lagged behind the pace of genetic diagnostics. One factor that must be taken into account is that every individual is unique, and that looking one-dimensionally at the genome alone will provide limited information on the underlying pathobiology. To overcome the complexity of this challenge, we have established a multi-dimensional deep phenotyping framework implementing whole genome sequencing, RNA-sequencing, and targeted proteomic and metabolomic assays, combined with clinical phenotype information as a novel strategy for the study of rare diseases.Bardet-Biedl Syndrome (BBS) is a rare autosomal recessive pleiotropic disorder, characterised by loss of vision, obesity, renal dysfunction, learning difficulties and hypogonadism. This multi-system phenotype is caused by defects in genes involved in ciliogenesis and intraflagellar transport, the proteins of which localize to the primary cilia. We have obtained disease relevant tissue, including models derived from induced pluripotent stem cells, from a defined set of clinically well-characterised BBS patients. These individuals all harbor the same underlying diagnosed variant (*BBS1* c.1169T>G/p.M390R) but exhibit considerable phenotypic heterogeneity in their clinical presentation.Through RNA-sequencing of patient fibroblasts we have uncovered over 4000 differentially expressed genes leading to the identification of multiple pathways involved in cell metabolism, maintenance and signaling that are disrupted in BBS patients compared to controls. Integrated burden testing, using variants identified by whole genome sequencing, has revealed genetic variation in distinct ciliopathy related pathways that contribute to the overall mutational load of the disease. As a result, we have devised a disease-specific multi-dimensional molecular map, featuring integrated data from the phenome, genome, transcriptome, proteome and metabolome. This wealth of rich information has provided us with exciting new avenues, in which we can discover novel therapeutic targets, and define biomarkers that can be implemented to predict and pre-empt the onset of disease.

3163W

Ultra-sensitive droplet digital PCR for detecting a low-prevalence somatic *GNAQ* mutation in Sturge-Weber syndrome. *Y. Uchiyama[1,2], M. Miyajima[3], M. Masataka[4], M. Nakashima[1], N. Matsumoto[1].* 1) Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Department of Medicine and Clinical Science, Gunma University Graduate School of Medicine, Gunma; 3) Department of Neurosurgery, Juntendo University Graduate School of Medicine, Tokyo, Japan; 4) Department of Biostatistics, Graduate school of Medicine, Yokohama City University, Yokohama, Japan.

Droplet digital PCR (ddPCR), a method for detecting DNA quantity, is useful for determining somatic mutation rates utilizing TaqMan probes. In this study, the detection limit of copy numbers of genomic DNA by ddPCR was determined based on Poisson distribution. Peptide nucleic acid (PNA), which strongly hybridizes to target lesions, can inhibit target amplification by PCR. Therefore by combination of PCR with PNA and ddPCR (PNA-ddPCR), detection limit could be lowered. We revisited a somatic *GNAQ* mutation (c.548G > A) in patients with Sturge-Weber syndrome (SWS) using ddPCR and PNA-ddPCR. Importantly among three mutation-negative patients by previous next generation sequencing, two patients had the *GNAQ* mutation with less than 1% of mutant allele frequency. We also confirmed the same *GNAQ* mutation in three blood samples and one saliva sample. These data clearly indicates that ultra-sensitive method should be utilized in detecting the low-prevalence somatic *GNAQ* mutation in SWS.and that the mutation is derived from the mesoderm.

3164T

The first 200 diagnostic whole genome sequencing referrals from Australia: Results and recommendations. *T. Roscioli[1,2,4,5], M. Buckley[1,2,3], P. Budd[1,2], L. Burnett[1,2,3], L. Constantinescu[1,2], J. Copty[1,2], M. Cowley[1,2], M. Dinger[1,2], L. Ewans[1,2], M. Gonzalez[1,2], V. Gayevsky[1,2], T. Groza[1,2], W. Kaplan[1,2], M. Lubka-Pathak[1,2], B. Lundie[1,2], J. Mattick[1,2], S. Mead[1,2,3], A. Minoche[1,2], M. Pinese[1,2], D. Saroufim[1,2], D. Schofield[2,6,7], A. Statham[1,2], J. Stockmyer[1,2], B. Terrill[1,2], K. Ying[1,2].* 1) Garvan Institute of Medical Research; 2) The Kinghorn Centre for Clinical Genomics (KCCG) at the Garvan; 3) SEALS Genetics, NSW Health Pathology Service; 4) St Vincent's Clinical School, University of New South Wales, Australia; 5) Department of Medical Genetics, Sydney Children's Hospital, Australia; 6) Murdoch Children's Research Institute, Australia; 7) Faculty of Pharmacy, the University of Sydney, Australia.

Background**:** The Kinghorn Centre for Clinical Genomics (KCCG) was recently accredited for diagnosing germline pathogenic variants using the Illumina HiSeq X-Ten whole genome sequencing (WGS) system, as the first facility in the Southern hemisphere. Aims**:** To describe the results of diagnostic WGS in childhood and adult Mendelian disorders, summarise analysis methodologies and provide recommendations to maximize the utility of clinical referrals. Subjects and Methods**:** Referrals were submitted from clinical geneticists and non-genetic physicians working with genetic counsellors, including families with adult-onset movement disorders, intellectual disabilities, cardiomyopathies and immunodeficiencies. 200 consecutive samples were assessed at the KCCG diagnostic genomic facility. Sequencing libraries were prepared using Illumina TruSeq DNA preparation kits, robotic instrumentation, and sequenced on Illumina HiSeqX sequencers with 2x150bp reads, equivalent to an average genome-wide coverage of >30x, with 95% of the genome covered to >20x depth. Raw sequencing reads were aligned to the genome using BWA-MEM with SNVs and indels identified using GATK HaplotypeCaller, filtered with the Seave system and reported by genomic pathologists. Results**:** The overall diagnostic rate was 50% with the largest phenotype group, Intellectual Disability having likely causative pathogenic variants identified in 54%, skewed to *de novo* events. The majority of successful gene diagnoses were in trio-based analyses confirming additional utility from testing family units. Requests for specific single gene analyses extracted from WGS did not result in diagnostic findings also confirming that a broader panel or whole genome approach is most valuable. A number of unsuspected genomic findings altered clinical diagnoses, confirming the combined utility of a clinical genomic classifications. Extending the diagnostic pipeline to copy number variant analysis will be discussed.

3165F

5-day Rapid Paediatric Sequencing (RaPS) for rare disease gene identification in children on the Paediatric Intensive Care Unit at Great Ormond Street Hospital (GOSH). *H.J. Williams[1], S. Drury[2], L. Ocaka[1], C. Bacchelli[1], L. Jenkins[2], L.S. Chitty[2], M.J. Peters[3], P.L. Beales[1], GOSgene.* 1) UCL Institute of Child Health, London, United Kingdom; 2) NE Thames Regional Genetics Laboratory, Great Ormond Street Hospital, Barclay House, 37 Queen Square, London; 3) UCL Institute of Child Health and Great Ormond Street NHS Foundation Trust, London, United Kingdom.

The use of Whole Genome Sequencing (WGS) for the identification of genes that cause rare diseases is well documented. We sought to evaluate the utility of WGS in a clinical setting that is likely to benefit greatly from a rapid genetic diagnosis and to this end we have highlighted the Paediatric Intensive Care Unit (PICU) at Great Ormond Street Hospital (GOSH). In the first instance, we aimed to deliver a turnaround time, from sample ascertainment to candidate variants of potential clinical significance, in under a week. For the pilot project we have identified 10 patients (run as family trios) from the PICU ward that have an immediate clinical need, present with a phenotype of suspected genetic etiology and for whom the rapid identification of a causative gene could have a positive impact on clinical management. Our Rapid Paediatric Sequencing (RaPS) pipeline begins with notification from a senior clinician on PICU of a suitable case that has consented to the study. We next collect blood samples which undergo rapid DNA extraction and WGS library preparation within the accredited diagnostic laboratory at GOSH. Here the sequencing takes place on an Illumina Hiseq (proband) and Nextseq (pooled parents) both on rapid-run mode. The resulting sequence data is analysed by GENALICE using their propriety software to map and call variants. Interpretation of the data utilises QIAGEN Ingenuity Variant Analysis software. Our pilot project data has shown we are able to return actionable findings to the PICU clinicians within 5 days of the initial consent. Of the 10 patients processed as part of this pilot we have identified 3 patients with a gene known to cause the underlying disease and 2 patients with genes that are novel candidates for the disease. In the remaining 5 patients we have so far been unable to identify any actionable variants.

3166W

Whole exome sequencing re-analysis boosts diagnosis in Mendelian disorders: Recommendations for laboratory practice. *L.J. Ewans[1,2], M.J. Cowley[1,2], Y. Zhu[3], V. Gayevskiy[1], K. Ying[1], C. Walsh[4], E. Lee[4], E. Kirk[4,5,6], A. Colley[7], A. Turner[5,6], D. Mowat[5,6], L. Worgan[7], M-L. Freckmann[5,6], R. Sachdev[5,6], M. Field[3], D. Miller[1], P. Morris[1], M. Buckley[4], M.E. Dinger[1,2], T. Roscioli[1,2,5].* 1) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, New South Wales, Australia; 2) St Vincent's Clinical School, University of New South Wales, Darlinghurst, Australia; 3) Newcastle GOLD Service, Hunter Genetics, Waratah, NSW, Australia; 4) SEALS laboratory, Prince of Wales Hospital, Randwick, NSW, Australia; 5) Department of Medical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia; 6) School of Women's and Children's Health, University of New South Wales, Australia; 7) Clinical Genetics Department, Liverpool Hospital, NSW, Australia.

   Background: Mendelian disorders are clinically and genetically heterogeneous making them ideally suited to diagnosis through whole exome sequencing (WES). Diagnostic potential is estimated to be between 25-30%, with re-analysis of data proposed to improve the diagnostic success.   Aims: To assess the diagnostic potential of WES in a Mendelian cohort with comparison to subsequent re-analysis at a 12 month time point.   Methods: WES was performed in a cohort of 54 patients from 37 families with unsolved Mendelian disorders in which there had been extensive prior investigation. The majority of patients presented with intellectual disability, with the remaining diagnoses including: skeletal dysplasias, retinitis pigmentosa, haematological disorders, seizures, metabolic disorders, and dysmorphic syndromes. WES data from Illumina HiSeq 2500 sequencers was filtered using the GEMINI (Genome MININg) software underlying the in-house SEAVE platform. The most likely inheritance model was applied to each family and further filtering was performed using pathogenicity scoring systems such as CADD (Combined Annotation Dependent Depletion) to assist prioritization of disease-causing mutations. Re-analysis was performed at a 12 month time point with an improved clinical interpretation pipeline.   Results: Initial analysis of WES data identified a diagnosis in a known disease gene in 30% (11 of 37 families) of the cohort. Following subsequent re-analysis, this improved to 38% (14 of 37 families). This was due to interim novel gene discovery, extended phenotype information, and improvements to the clinical interpretation pipeline.   Discussion: The diagnostic success of WES in Mendelian disorders is significantly improved with re-analysis of data at a 12 month time point. Recommendations for clinical interpretation pipelines include: obtaining accurate phenotype information, use of in-house databases of variants, updated variant prioritisation tools including gene annotation and incorporation of population data, and collaboration between referring clinicians and laboratories. It is also recommended that resources are provided to ensure re-analysis of historical genomic data to maximize diagnostic utility.

3167T

Multi-omic approaches towards personalised medicine for rare diseases. *D. Kelberman[1,2], G. Otto[1,2], F. Minneci[1,2], R. Davies[1,2], J. Gratton[1,2], J. Kenny[1,2], G. Madhan[1,2], X. Wang[1,2], S. Dowle[1,2], T. Peltola[3], T. Tukiainen[3], S. Kaski[3], N.J. Sebire[1,4], L.R. Wedderburn[1,4], K. Nistala[1,4], N. Shah[4], J. Kammermeier[1,4], K. Pritchard-Jones[1,4], W. Mifsud[1,4], J.C. Sowden[1], M. Bitner-Glindzicz[1], S. Rahman[1,4], G.E. Moore[1], C. Bacchelli[1], P.L. Beales[1,2].* 1) UCL Institute of Child Health, University College London , London, United Kingdom; 2) National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Trust, London, UK; 3) Helsinki Institute for Information Technology HIIT, Aalto University, Helsinki, Finalnd; 4) Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK.

   Large genetic sequencing programmes for Mendelian diseases are underway internationally, but these only address the genomic landscape in a single dimension; that of defining disease-causing variants. Patients affected by rare diseases and carrying the same primary genetic defect can markedly vary in their expression of clinical phenotypes or response to therapies. We have established a new programme of research to address a major challenge in understanding how other genomic variants and non-genetic factors influence disease variability. This pilot project has developed strategies to investigate the basis of variability in human disease, initially in a select number of rare diseases with clinically well characterised patients. Selected diseases include: Bardet Biedl syndrome, Juvenile Dermatomyositis, Wilms' tumour, very early onset inflammatory bowel disease, Usher syndrome, mitochondrial disease and Silver Russell syndrome. Using multiple technologies including whole genome and RNA sequencing, mass spectrometry and targeted proteomic and methylation assays, we have undertaken in depth phenotyping at the molecular level to identify variation in coding and non-coding RNA, protein, metabolites and epigenetic factors measured in disease relevant tissues in conjunction with genetic variation. Initial preliminary investigations of individual datasets for specific projects have highlighted a number of cellular pathways that can be further explored for their role in disease pathology or in response to particular therapies. Following the initial analysis of individual omics data, the combined integration of these multi-omic datasets with clinical information within patient subgroups will allow the identification of biomarkers to assist in the prediction of disease severity, risk of developing specific complications, and/or stratify patients for appropriate or novel therapies depending on the disease in question. It is expected that the methodological strategies being developed will also offer the opportunity to show proof of principle for a broader range of applications as well as the better understanding of the molecular pathology of disease. This will suggest new avenues or targets and markers for the most efficacious treatments in a move towards stratified or personalised medicine for rare diseases.

3168F

Identification of age-related patterns in human extracellular RNA in serum. *D.F. Dluzen[1], S. De[2], W.H. Wood[2], K.G. Becker[2], A.B. Zonderman[1], T. Tanaka[3], L. Ferrucci[3], N. Noren Hooten[1], M.K. Evans[1].* 1) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD; 2) Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD; 3) Translational Gerontology Branch, National Institute on Aging, NIH, Baltimore, MD.

   Extracellular RNA (exRNA), circulating in the blood, have been identified and used as markers for human development, aging and disease. However, these studies have been limited in scope, either by profiling one subtype of RNA species (e.g. microRNA [miRNA]) or profiling RNAs of a particular transcript size (e.g. small or large). Here, we have sequenced both large and small RNAs from human serum under one comprehensive analysis within the context of normal aging. We hypothesized that age-related changes in exRNAs are important for normal aging and age-related degenerative processes. Total RNA was isolated from serum from young (30-35 yrs) and old (80-85 yrs) African American females chosen from the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study or the Baltimore Longitudinal Study of Aging (BLSA). Total RNA-seq was performed on the Ion Torrent Proton and sequences were aligned against the human genome version 19 using separate protocols to identify linear RNA, miRNA, or circular RNA (circRNA). There was an average of 9 million reads per individual. We identified mRNA, miRNA, long non-coding RNAs (lncRNA), small nucleolar RNA (snoRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), circRNA, and Y RNA transcripts. The total number of unique RNA transcripts varied by individual; however, the percentages of each transcript subtype were similar. In general, the distribution of highly abundant mRNAs was more variable in older individuals. There were a number of shared mRNAs, e.g. *NOTCH2NL*, which were abundant in both young and old. We sequenced 458 unique miRNAs throughout the entire cohort, 94 of which are present in every individual. 95 miRNAs were unique to only older individuals and 74 were present only in young. There were significant age-dependent changes in serum abundance of *C19orf48* (mRNA), *MTND1P23* and *SNHG10* (lncRNA), *SCARNA5* (snoRNA), *RNA5SP338* (rRNA), and several miRNAs, including miR-16, -191, -26a, -24, -21, and let-7g. Pathway analysis revealed that the top pathways associated with exRNAs profiled in old individuals was oxidative phosphorylation and mitochondrial dysfunction. We will examine the relationships between the identified RNA species and nutritional, physiological, psychosocial, and cognitive parameters in order to identify novel, age-related circulating RNA biomarkers and related regulatory pathways.

3169W

Construction of a highly efficient platform CNVplex for the detection of chromosomal aneuploidies. *L. Haibo[1], X. Jingjing[1], W. Ting[1], L. Hong[1], M. Jun[1], Z. Qin[1], L. Minjuan[1], L. Yinghua[1], J. Zhengwen[2], L. Deyuan[2].* 1) Center for Reproduction and Genetics, Suzhou Municipal Hospital, Suzhou, JiangSu, China; 2) Genesky Diagnostics (Suzhou) Inc., SIP, Suzhou, JiangSu, China.

   Chromosomal aneuploidies was an important reason for miscarriage at the early stage of pregnancy, and chromosome karyotype analysis of chorionic villi could provide valuable information for genetic counselling and prenatal care in future pregnancies. A novel technique CNVplex was developed by modifying the multiplex ligation-dependent probe amplification (MLPA) method. 170 pairs of probes were designed for 24 human chromosomes, and each chromosome contains at least 5 pairs of probes. Each probe is composed of a oligonucleotide complementary to the target sequence and the universal sequence to allow PCR amplification with the fluorescent primer pairs. After hybridization, ligation and multiple PCR amplification, the amplified products were separated by size and color using capillary electrophoresis, and the relative amount of target sequences were calculated according to the amount of PCR products. 236 samples of chorionic villi with chromosomal abnormalities determined by karyotype analysis and MLPA in our center before were selected and re-analyzed by CNVplex. Except for 19 samples which could not be analyzed by CNVplex, including 5 samples with quality problems and 14 samples with polyploidies or chromosome structural abnormalities, the detection rate of chromosomal aneuploidies by CNVplex is 92.4%, which is higher than that of karyotype analysis (56.4%) and MLPA (67%) (p<0.01). In conclusion, CNVplex is more sensitive and efficient in the detection of chromosomal aneuploidies than karyotype analysis and MLPA, and it could be applied clinically as an alternative to MLPA.

3170T

High accuracy genotyping with duplex sequencing. *J.C. Roach[1], D. Mauldin[1], M.J. Hipp[2], E. Smith[2], G. Glusman[1], S.R. Kennedy[2].* 1) Family Genomics, Institute for Systems Biology, Seattle, WA; 2) Department of Pathology, University of Washington, Seattle, WA.

   Identifying needles in haystacks has become increasingly important for genomic applications. As technical accuracy of key technologies increases, the size and complexity of the haystacks we may search also increase. False discovery rates are a limiting factor in high-throughput genomic pipelines, as these ultimately bottleneck at procedures performed by humans. Increases in technical accuracy lead to decreases in false discovery rates. For gene- and marker-level applications (e.g., STS marker searches for a classic disease), technical accuracy must reach ~1:500 to reduce a domain of ~20,000 genes to tens of candidates. For genome-level applications such as identification of a rare detrimental SNV responsible for a birth defect, technical accuracy must reach ~1:5,000. This is the typical accuracy of current high-throughput sequencing, and has resulted in the explosion of disease-variant identifications in the last half decade. False discovery rates are particularly poignant in forensic applications — they may lead to false convictions. We now have the opportunity, and the need, to look for needles in a haystack, where the haystack may be any haystack on the planet. Applications include identification of variants in stem cells within anlages of developing tissues, tumor subclones that are possible keys to cures, and traces of a sexual criminal diluted in other genomes. Each of these applications involves computational dissection of mixtures of DNA. If technical error is higher than the fraction of a mixture DNA component, it becomes difficult or impossible to recognize and distinguish component genomes in mixtures. Current mass-marketed sequencing technologies cannot distinguish trace genomes mixed with other genomes in excess. Duplex Sequencing is designed to decrease uncorrected technical error, and enables analysis of complex trace mixtures. Duplex Sequencing assays each strand of DNA independently. The two strands of dsDNA are labeled and separated before amplification and sequencing. The same technical error must occur on both strands in order to avoid detection and correction. Duplex Sequencing offers the ability to accurately detect a minor allele fraction of ~1:50,000, extending the reach of sequencing to previously intractable challenges in stem-cell biology, oncology, and forensics. We are developing protocols to apply Duplex Sequencing to problems in forensics, including the analysis of degraded samples and complex mixtures of multiple genomes.

3171F

Establishment of blood sample collection protocols and procedures for DNA methylation and RNA analyses in genome cohort studies. *A. Shimizu[1], R. Furukawa[1], T. Hachiya[1,2], Y. Shiwa[1,2], H. Ohmomo[1], M. Satoh[1,2,8], K. Ogasawara[3,4], M. Nakamura[3,5], J. Hitomi[3,6], K. Sobue[3], M. Sasaki[3,7].* 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan; 2) Division of Biobank and Data Management, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan; 3) Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan; 4) Department of Neurosurgery, School of Medicine, Iwate Medical University, 19-1 Uchimaru, Morioka, Iwate 020-8505, Japan; 5) Department of Internal Medicine, School of Medicine, Iwate Medical University, 19-1 Uchimaru, Morioka, Iwate 020-8505, Japan; 6) Department of Anatomy, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan; 7) Division of Ultrahigh Field MRI, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan; 8) Division of Biomedical Information Analysis, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan.

   PURPOSE: In genome cohort studies, accurate omics information that comprises SNPs, DNA methylation (DNAm) rates, and gene expression levels is often not provided mainly owing to omics status is physically unstable, differs among cell types, and changes over time. Thus, we attempted to establish optimized blood sampling protocols for omics analyses that can be used in genome cohort studies. METHODS: To examine the influence of sample collection/transportation methods on the stabilities of DNAm and RNA, we extracted DNA and RNA from blood specimens processed by multiple protocols such as with/without nucleic acid stabilizer, using different blood collection tubes, extraction methods, and transportation conditions, and then measured DNAm rates using DNA microarray and expression levels using next-generation sequencer. Second, to optimize cell-sorting protocols, cells were sorted using various cell separation solutions, collection tubes, and cell sorters, and we then examined the amount and purity of nucleic acids of the sorted specimens. Third, to evaluate the effect of time on omics variations, we withdrew blood from 2 volunteers 24 times for 3 months, and then measured the DNAm rates using DNA microarray and expression levels using next-generation sequencer. We also measured body temperature, blood count, and high-sensitivity C-reactive protein. RESULTS: We found that the refrigerated transportation of the blood cells purified immediately after blood withdrawal and containing cell solubilizer (Promega, Madison, WI, USA) was the most stable method for RNA measurements. We also confirmed that DNAm conditions were altered during specimen processing and transportation, mainly owing to variations in cell composition, which can be normalized by estimating the cell ratios in the samples. For cell sorting methods, we readily purified various blood cells through the combined use of the CPT Tube (BD Biosciences, Franklin Lakes, NJ, USA), Mono-Poly Resolving Medium (DS Pharma Biomedical, Osaka, Japan), and SH800 cell sorter (SONY, Tokyo, Japan). Regarding chronological changes, we found substantial variations in the expression levels of some genes, but verified that most DNAm sites were stable for several months and showed approximately 2% temporal variations. CONCLUSION: We established optimized protocols for collecting and transporting blood samples to improve the precision of DNA methylation and RNA analyses in genome cohort studies.

3172W

GRIN (Genomic Research and Innovation Network): A sustainable infrastructure for pediatric data sharing. *K.L. Sund[1], S. Biswas[2], M. Heinz[3], T. Glauser[1], K.D. Mandl[3], I. Krantz[2].* 1) Cincinnati Children's Hospital, Cincinnati, OH; 2) Children's Hospital of Philadelphia, PA; 3) Boston Children's Hospital, MA.

The most significant barrier to research in pediatric personalized medicine is the ability to identify sufficiently large patient cohorts with in depth, consistent phenotyping and genotyping, coupled to the computing capabilities to correlate and interpret information to advance discovery and improve care. The extreme phenotypes found in pediatric patients represent a tremendous opportunity to understand many aspects of disease, but the first step is to confirm a correlative relationship between phenotype and the underlying genomics. The Genomic Research and Innovation Network (GRIN) is a collaborative effort between 3 pediatric institutions: Boston Children's Hospital, Children's Hospital of Philadelphia and Cincinnati Children's Hospital. The vision is to accelerate genomic discoveries in pediatric populations by creating a collaborative data sharing network at the institutional level. A key principle of the network is the collective ability of our centers to pursue well-designed, carefully selected research studies through broad sharing of genetic/genomic data associated with well-characterized phenotypic information and to link this data to a federated biobank of specimens. The team approach will also make it possible to leverage the strongest expertise for any given phenotype to improve interpretation and meaningful functional research studies. GRIN is currently piloting collaborative studies across the three sites to demonstrate feasibility of data sharing between pediatric centers. Simultaneously, we are developing the necessary regulatory, legal and bioinformatics infrastructure to scale to multiple diverse projects at an institutional level. GRIN infrastructure includes aligned institutional enrollment protocols, cloud-based data sharing and a portal to query a virtual biobank/data bank to for biospecimens and genomic data related to specific variants or phenotypes. To sustain this collaborative effort over time, the GRIN has developed a strategic model to use this existing infrastructure to engage appropriate industry partners in a service model. Such engagement will enhance pediatric academic research over time so that we can foster advances in phenotype-genotype interpretation and translation to patient care. GRIN will bring a much-needed pediatric focus to cohort expansion, genomic data sharing, interpretation, and matching expertise across existing and future member institutions with the ultimate goal to improve clinical outcomes. .

3173T

Targeted sequencing using advanced molecular inversion probes (MIPs) for discovery and validation in Mendelian disease trios. *S. Suresh, K. Jefferson, H. Halvensleben, D. Green, R. Bannen, M. Brockman, T. Richmond, D. Burgess.* Roche NimbleGen, Madison, WI.

Progress in understanding the genetic components of human diseases has been bolstered by next-generation targeted sequencing technologies. Emerging techniques that promise to improve throughput, cost and turnaround time include molecular inversion probe (MIP)-based methods, which enable rapid and simple workflows to detect sequence variation in inherited disease and cancer. We have developed a streamlined HEAT-Seq targeted enrichment assay, for variant discovery and validation. It utilizes advanced versions of MIPs with unique molecular identifiers incorporated into probes to identify PCR duplicates, improve sensitivity, and reduce false-positive calls. Previous studies using MIPs failed to address a key barrier to the widespread adoption of this technology, e.g., the difficulty of designing probes and probe-sets that yield uniform sequence coverage over the target. We addressed this issue by developing an empirical database-driven approach. All of the probes in the database have been assigned empirically derived performance scores. Consideration of individual probe performance in the design process allows generation of custom probe sets which typically produce more uniform coverage relative to previous methods. The efficiency of our approach was validated using three HEAT-Seq gene panels targeting cancer and inherited disease. Our analyses indicate improved performance compared to previous methods for two important metrics: (1) percentage of functional probes in the initial probe-set, and (2) sensitivity of SNP detection. The data suggest HEAT-Seq target enrichment system is an efficient, sensitive and cost-effective tool for targeted sequencing. To demonstrate the efficacy of HEAT-Seq target enrichment assay for identifying genetic variants in Mendelian disease, we studied two trios segregating rare autosomal recessive neurological disorders: Ataxia Telangiectasia (AT) and Familial Dysautonomia (FD). The AT mutation is a G>A coding mutation in exon 55 of *ATM*. The FD mutation is a T>C transition in a splice acceptor of *IKBKAP*. The mutations in both trios were previously described. We characterized the samples using both a large gene HEAT-Seq discovery panel that included *ATM* and *IKBKAP*, and a small validation panel that targeted only *ATM* and *IKBKAP*. The results show HEAT-Seq target enrichment system is a fast and cost-effective assay for identifying genetic variation for both discovery and genotyping/validation applications in Mendelian disease.

3174F

Human gene mapping: Historical review. *A.A. Khedr[1], S.A. Temtamy[2], M.A. Awadalla[3].* 1) Human CytoGenetics , National Research Center, Cairo, Egypt; 2) Clinical Genetics, National Research Center, Cairo, Egypt; 3) Pediatric& Genetics, Faculty of Medicine, Ain Shams University.

   The science of Human genetics can be rooted 1500 years ago where the blood disorder Hemophilia was mentioned in the Talmud. However, the idea concerning genetics really started with the work of M.G. Mendel in 1865**.** His work was only appriciated, following the indepedent work of W.S.Sutton & T. Boveri in 1902, who proposed the chromosome theory of heridity and the relation of those minute colored structures with the phenomenon of inheritance proposed by him. The science of human genetics witnessed the first revolutionary advance in its feild in late fifties by the assignment of the Duffy blood group gene on an autosome , however , it was untill early eghities when it was located to ch.1q12-21. Fudemamental to progress in molecular genetics was the discovery of restriction enzymes which became the scalpel for dissecting the human genome. The launching of the human gene mapping project aiming to sequence the entire 3 billion base pairs resulted in explosion of information that the entries of McKusick Catalogue of Mendalian Inheritance (MIM) increased from 1487 in the first edition in 1966 to 4500 in the 9th edition in 1990. After 50 years of his first edition,the OMIM entries are 23,528 to date. We aim to review updates in various methods of human gene mapping, starting from Linkage sudies through cytogenetics and FISH to the late advances in the molecular technology including methylation specific PCR(m-PCR), Multiplex ligation-dependent probe amplification (MLPA) CGH, Array CGH, Microarray CGH ; their principles, imporance, genetic significance, and the benifits they added to our knowledge of the human genome in general and to genetic disease in particular.

3175W

A novel sgRNA tagging technology identifies and addresses heritable clonal heterogeneity in CRISPR screens. *A. Biton[1,2,5], M. Boettcher[3,5], S. Covarrubias[3,5], J. Blau[3], H. Wang[4], N. Zaitlen[1], M. McManus[3].* 1) Department of Medicine, Lung Biology Center, University of California, San Francisco, San Francisco, California, USA; 2) Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI, USR 3756 Institut Pasteur et CNRS), Paris, France; 3) Department of Microbiology and Immunology, UCSF Diabetes Center, University of California, San Francisco, San Francisco, CA 94143, USA; 4) Departments of Medicine and of Microbiology & Immunology, the Rosalind Russell-Ephraim P. Engleman Medical Research Center for Arthritis, and the Howard Hughes Medical Institute, University of California, San Francisco, CA 94143; 5) These authors contributed equally.

   Functional genetic screens using CRISPR/Cas technology have become an invaluable tool to discover gene function in cell lines at a genome scale. However, it is still a nascent technology with complex and sensitive elements that remain unexplored. First, in order to obtain a cell line suitable for functional screens, a selection of a subclone with stable Cas9 expression is performed. This subclone is selected from an heterogeneous population of cells with varying levels of Cas9 expression that may introduce bias by responding differently to treatment. Second, screens can be performed using either Cas9 nuclease (CRISPRwt) for knockout mutations, or dCas9-KRAB (CRISPRi) for gene silencing. The comparison of the results of these two competing CRISPR systems has started but needs to be extended. Here, we performed a pooled genome-wide CRISPR screen for resistance to TRAIL-mediated apoptosis on Jurkat cell lines. A sgRNA library targeting most of the genes in the human genome by up to 12 sgRNAs was designed for the CRISPRwt and the CRISPRi system. The screen was run in each system on two subclones selected from the same starting population of cells. In this screen, we introduce a novel multi-level barcoding technology that enables the study of the screen at the subclonal level. The first level of barcoding allows to study the behavior of the four clonal lines across time, after they were mixed in a TRAIL-treated and an untreated vessel. We demonstrate that these clones have a heritable, but heterogeneous response to TRAIL: one sensitive and one resistant clonal cell. This phenomenon likely exists in other screens and is therefore both a source of power loss and bias. A second level of barcoding introduced during the sgRNA library PCR amplification allows us to distinguish between subclones carrying a same sgRNA. Our analysis of the pooled screen across time (cells were harvested at days 4, 9, and 14 after treatment) show that both CRISPR systems identify overlapping as well as mutually exclusive sets of gene hits. On the other hand, we observe that, within a same CRISPR system, results substantially differ between the two clonal populations. We demonstrate that the barcoding technology not only provides in-sample replication but also a means to explore the effect of sequencing depth or clonal mixture on power in a CRISPR screen. We show how heterogeneous clonal populations can impact the output of a pooled screen and lower the power of candidate detection.

3176T

Advanced applications for clinical research exomes. *M. Corioni, E. Lin, K. Jeong, F. Useche, A. Vadapalli, C. Pabon, M. Visitacion, G. Amparo, M. Ramaiah, M. Isaksson, D. Roberts.* Agilent Technologies, Santa Clara, CA.

Targeted re-sequencing enables highly sensitive and comprehensive detection of variants and provides insights into the biology behind a given phenotype. With recent major advances in this technology, combined with a better understanding of biological pathways, NGS is now extensively considered for use in clinical research. Whole exome sequencing provides deep coverage of genomic content from curated databases making it an efficient solution to characterize and catalogue variants. We describe several strategies that add additional value to whole exome analysis enabling more complete coverage and advanced applications such as detection of copy number changes and phasing. We share an augmented exome for clinical research that focuses on targets relevant to constitutional disease research with input from specific databases including Human Gene Mutation Database (HGMD®), ClinVar and Online Mendelian Inheritance in Man (OMIM™). We also describe the combination of exome with copy-number probes to extend target enrichment applications to the detection of genome-wide copy-number-changes (CNCs), and copy-neutral loss-of-heterozygosity (cnLOH). Finally, we demonstrate a phased exome coupled to 10x Genomics' linked read technology. The addition of "phasing" probes allows determination of long haplotype blocks. The separation of exonic reads into haplotypes further aids in structural variant and translocation determination, especially for disease causality in complex genotypes. We describe data analysis and performance attributes for each of these platforms, including SNP/INDEL, loss of heterozygosity, copy number variation, phasing, translocation, and resolution of complex compound heterozygote samples.

3177F

Target enrichment using a neurology panel for 12 barcoded genomic DNA samples on the PacBio SMRT Sequencing Platform. *K. Eng, W. Lee, T. Hon, A. Sethuraman, J. Gu, P. Baybayan, S. Kujawa.* Pacific Biosciences, Menlo Park, CA.

Target enrichment is a powerful tool for studies involved in understanding polymorphic SNPs with phasing, tandem repeats, and structural variations. With increasing availability of reference genomes, researchers can easily design a cost-effective targeted investigation with custom probes specific to regions of interest. Using PacBio long-read technology in conjunction with probe capture, we were able to sequence multi-kilobase enriched regions to fully investigate intronic and exonic regions, distinguish haplotypes, and characterize structural variations. Furthermore, we demonstrate this approach is advantageous for studying complex genomic regions previously inaccessible through other sequencing platforms. In the present work, 12 barcoded genomic DNA (gDNA) samples were sheared to 6 Kb for target enrichment analysis using the Neurology panel provided by Roche NimbleGen. Probe-captured DNA was used to make SMRTbell libraries for SMRT Sequencing on the PacBio RS II. Our results demonstrate the ability to multiplex 12 samples and achieve 1300x enrichment of targeted regions. In addition, we achieved an even representation of on-target rate of 70% across the 12 barcoded genomic DNA samples.   .

3178W

A comprehensive system for monitoring and analyzing human genome sequencing data. *S. Germer, J. Carter, J. Spencer, R. Musunuri, B.-J. Chen, H. Swaerdlow.* New York Genome Center, New York, NY.

The number of human genomes being sequenced has increased dramatically since the introduction of the Illumina HiSeq X instrument. In the past year, New York Genome Center has sequenced >12,000 whole human genomes at >30x. To ensure sustained production of high quality data generation across multiple library preparation protocols and versions of the sequencing chemistry, we have developed a comprehensive system for monitoring and analyzing the performance of the HiSeq sequencing instruments and the quality of raw and analyzed sequencing data. Instrument performance is monitored through a web-based display that emulates the Illumina Sequence Analysis Viewer software across all instruments, and allows both real-time and historical review of key quality metrics. A searchable browser interface displays flow-cell lane-level performance in terms of pass-filter (PF) clusters, the base quality (%>Q30, Reads 1 and 2), the percent of PhiX control template aligned, phasing and pre-phasing, as well as the error rate. Upon completion of a sequencing run, automated de-multiplexing generates a report for each flow-cell containing lane- and sample-level metrics, including percent PF reads identified and assigned to a sample, the read quality, as well as estimated duplication and adaptor contamination rates. This report flags metrics that fail to meet specified criteria, and holds the corresponding FastQ files from automatically progressing to downstream analysis until reviewed by a dedicated sequencing analyst. The data captured is organized in a database which allows us to track performance over time, by querying, for instance, PF rates, quality, or duplication rates by instrument, time period, library prep type, etc. Once a sample has progressed through the bioinformatics analysis pipeline, a separate web-based tool displays graphically or in a table format more than 25 per-sample metrics (related to alignment, coverage and coverage uniformity, contamination estimates, variant call quality, and more). The efficient and interactive interface allows analysts to quickly review project level data, to ascertain quality and identify QC failures, prior to data release for tertiary analysis. We will illustrate how the ability to track metrics at several levels and across all sequencing runs and samples allows us to identify outliers for further investigation, and to maintain consistently high quality performance across several large scale sequencing projects.

3179T

Whole genome mapping makes vital contributions to human reference assembly curation. *J.W. Torrance, W. Chow, K. Howe.* Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.

Whole Genome Mapping is a technology developed by BioNano Genomics for establishing the large-scale structure of a genome by visualising the spacing of fluorescently labelled sequence motifs on single molecules of DNA that are hundreds of kilobases in size. The resulting genome maps have been successfully employed for scaffolding high quality genome assemblies. The Genome Reference Consortium (GRC) are applying genome maps to a new purpose: the evaluation and improvement of the human reference genome assembly. The genome maps we have consulted include publicly available data generated by BioNano Genomics and the Genome in a Bottle Consortium, covering a trio from the CEPH Utah Reference Collection including NA12878, a trio of Ashkenazi Jewish ancestry, the YH genome, and a Han Chinese genome, as well as a CHM1 haploid hydatidiform mole genome map from BioNano. These genome maps have been vital in correcting the primary reference assembly by allowing more accurate estimation of the sizes of gaps and tandem repeat regions, and by aiding the identification of erroneous deletions. They have also helped to identify and confirm regions of haplotypic variation consisting of indels or copy number variation, providing valuable supporting evidence for the addition of alternate loci to the assembly. One of the examples where the genome maps confirmed the need for a sequence change to the primary assembly involved an erroneous deletion affecting two exons of *CDK11B*. In another case, the CHM1 map identified a 99 kb insertion relative to the primary assembly affecting the pseudogene *GUSBP1*, resulting in the addition of an alternate locus to the assembly. The GRC makes genome map alignments available via their gEVAL browser (geval.sanger.ac.uk), for both GRCh37 and GRCh38. To our knowledge, gEVAL is the only genome browser that provides genome map alignments for human reference assemblies. It thus constitutes a resource for discovering whether large-scale variation or RFLPs exist in a region of interest. These genome map alignments can be viewed in concert with the host of other data sources which gEVAL provides, including cDNA alignments, clone end-pair alignments, alignments with other human genome assemblies, and information on regions under review by the GRC.

3180F

Quality assessment of stranded mRNA libraries from low input RNA material using Illumina NeoPrep digital microfluidics automation. *C. Viollet[1,2], G. Sukumar[1,2], N. Lott[1,2], C.L. Dalgard[1,2].* 1) Uniformed Services University, Bethesda, MD, USA; 2) Collaborative Health Initiative Research Program, Bethesda, MD, USA.

Statement of purpose: Sequencing library preparation may be automated by digital microfluidics. The Illumina NeoPrep system utilizes electrowetting for liquid handling in a microfluidics card and offers complete workflows for amplification-based whole genome and transcriptome sequencing library preparation. For the stranded mRNA workflow, an input of 25–100 ng total RNA is specified. As the workflow utilizes a closed system with low reaction volumes, we aimed to determine the lower limit of input for library generation and assess metrics for yield, primer dimer fraction, and size distribution, as well as sequencing metrics for duplicate rate, alignment quality and concordance of transcript expression across starting inputs. The purpose of this study is to provide guidance for digital microfluidics library preparation of low or even below quantitation samples for mRNA-seq. **Methods:** We used two reference total RNA samples, a universal human reference (UHR) and control iPSC line, as input into the NeoPrep TruSeq Stranded mRNA Library Prep Kit at ranges between 0.3 to 210 ng. Library size distribution was assess using automated capillary electrophoresis (Fragment Analyzer) and yield was measured by qPCR (KAPA Quant on LightCycler 480). Libraries were pooled and sequenced on a NextSeq 500 with paired-end 75 bp reads. Quality filtered output was analysed using TopHat alignment and Cufflinks. **Summary of results:** All generated libraries were above 6 nM yield, homogeneous for ~320 bp peak size. Yields were associated with input and ranged from 6–370 nM. Recommended input of 25 ng yielded ~200 nM libraries with negligible (<0.5%) adapter dimer peak. In comparison, below specification input of 0.3 ng yielded ~6nM libraries with 7% adapter dimer peak. Pooled sequencing resulted in efficient multiplexing with no difference in %Q30 between libraries by input. A decrease in percentage of total reads aligned was observed with decreasing input RNA amounts. Most significantly, the duplicate rate increased as a function of decreased input, ranging from ~5% for high input to ~87% for the lowest input (0.3 ng). However, we observed that transcript expression values were highly correlated ($R^2 > 0.98$) between recommended and low input RNA libraries after normalization. Thus, we have defined a lower range of RNA input for acceptable NeoPrep workflow and how output metrics for library diversity, sequencing efficiency and transcript expression analysis are associated with low input.

3181W

GWAS data generation using heparinized red blood cells. *S. Hall[1], S. Paciga[1], C. Tow-Keogh[1], W. He[1], B. Zhang[2].* 1) WW Research & Dev, Pfizer, Groton, CT; 2) WW Research & Dev, Pfizer, Cambridge, MA.

The Pfizer BioBank manages collections that have been obtained from a subset of clinical trials where participants have consented to provide biospecimens for broad, exploratory research. Occasionally, suboptimal specimens are received and may be the only source available for experimental purposes. In this study, 123 heparinized red blood cells were received for a genome-wide association study (GWAS). Heparinized blood is not generally considered a viable source for DNA extraction and there was no supportive literature available to indicate its suitability for GWAS analysis. Agencourt®Genfind™ v2 magnetic bead DNA extraction chemistry was used to purify DNA from these specimens. The DNA concentration and yields were low due to the presence of heparin but quality measures were satisfactory. We were able to generate GWAS data from all of the specimens using Illumina's OmniExpressExome BeadChip array containing >950,000 SNP markers. Most of the specimens had a genotyping call rate >95% or better.

3182T

Ins and outs of current nucleic acid quantification and a solution towards standardization. *T. Martens, K. Plasman, T. Duthoit, T. Montoye.* R&D, Trinean, gentbrugge, east flanders, Belgium.

Failure of many molecular assays can be attributed to the quality of the nucleic acids used. Therefore, it is of utmost importance to thoroughly investigate both the quantity and purity of nucleic acids before proceeding to downstream assays. Most commonly-used technologies include UV/VIS droplet spectroscopy (i.e. Nanodrop) or fluorescence based assays (i.e. Qubit). These however suffer from some inherent drawbacks like labor-intensity and low accuracy. In an attempt to increase specificity while preserving ease of use and processing speed, UV/VIS droplet spectroscopy and an intelligent algorithm (DropSense16 + cDrop) were combined allowing more accurate nucleic acid quantification. Here, we use experimental data using different types of nucleic acids to demonstrate the advantages and shortcomings of each technology and propose ways to standardize nucleic acid QC ultimately resulting in a higher success-rate of downstream assays.

3183F

Unprecedented characterization of a human trio for new genomic reference materials. *J.M. Zook[1], J. McDaniel[1], M. Salit[2,3].* 1) Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD; 2) Material Measurement Laboratory, National Institute of Standards and Technology, Palo Alto, CA; 3) Department of Pathology, Stanford University, Stanford, CA.

The Genome in a Bottle Consortium (GIAB) has developed unprecedented data for a human trio, which is a candidate NIST Reference Material. These well-characterized genomes will underpin advances in genome sequencing and its applications in basic science, clinical medicine, and engineering of living matter. We are developing assemblies and phased variant calls of all sizes for as much of the genome as possible using 12 publicly available datasets from one trio, with no publication embargo and with publicly disclosed analyses. These data include BioNano Genomics, Complete Genomics paired-end and LFR, Ion Proton exome, Oxford Nanopore, Pacific Biosciences, SOLiD, 10X Genomics, and Illumina paired-end, mate-pair, exome, and synthetic long reads. We are using these data and developing new integrative methods to characterize increasingly difficult types of variants and regions of the genome. We have now expanded our high-confidence benchmark regions from 78% of the pilot genome NA12878 to 88-90% of this genome and a new genome. To further expand these benchmark regions, GIAB is collaboratively using these diverse data and integrating methods to create benchmark variant and homozygous reference calls in difficult-to-map regions and for larger, more challenging types of variation. 20 groups have contributed analyses of 12 different datasets for our trio. We have developed several de novo assemblies from long reads, resulting in a contig N50 >4 Mb and genome size >3 Gb. These assemblies have been used along with mapped short and long reads to generate structural variant calls using at least 20 algorithms. The resulting SV calls include deletions, novel and mobile element insertions, copy number variants, and inversions. High-confidence calls are being formed by integrating and evaluating the candidate calls from all methods. Preliminary integrated results for deletions in the son include 5300 deletions >50bp for which at least 2 different technologies have concordant size predictions and no other technology contradicts. This community effort is continuing to develop high-confidence benchmark calls for difficult variants in challenging regions of the genome - enabling optimization and development of technologies and algorithms and translation to clinical applications.

3184W

Genome-wide sequencing of STRs illuminates the haplotype structure of linked microsatellite-SNPs. *G. Shin[1], H. Lee[1], S. Grimes[2], B. Lau[2], H.P. Ji[1,2].* 1) Division of Oncology, Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center, Stanford University, Palo Alto, CA.

Microsatellites, also referred to as short tandem repeats (STRs) are multiallelic in terms of germline variation as a result of higher mutation rate than SNPs. As a result, STR genotyping is highly informative when characterizing genetic populations. However, the linkage disequilibrium structure of microsatellites and SNPs is largely unknown and there is little informative genetic analysis of the haplotype structure involving these variants. This knowledge gap mainly due to the absence of high-throughput genotyping technology for these loci. In this study, we used a novel targeted sequencing technology (STR-Seq), which can simultaneously genotype more than 2,000 microsatellite loci and their proximal SNPs. Particularly, STR-Seq uses paired-end sequencing reads to physically link STR and SNP genotypes. Unlike previous targeted sequencings, STR-Seq provides extraordinary efficiency in capturing the informative DNA molecules that span the entire repetition as well as the flanking sequences using targeted in vitro CRISPR/Cas9 fragmentation. Moreover, the technology eliminates PCR amplification and thus eliminates artifact insertions and deletions that obscure true genetic variation. We examined 2,191 microsatellite regions where a proximal SNP within 100bp has been reported. To increase the number of potential STR-SNP haplotypes, we considered the number of SNPs within the proximity to the microsatellite and genotype frequency of individual SNPs. In our initial analysis, we identified among a set of eight individuals, an average of STR-SNP haplotypes at 714. These haplotypes proved highly useful for a number of genetic application such as determining the composition of genetic mixtures. Currently, we are using this approach to define the microsatellite haplotype structure of over 1,000 individuals, by which allelic diversity, frequency spectra, and population differences of STR-SNP markers will be characterized.

3185T

Shining a light in the "dark matter" of the exome: Using linked reads and exome bait capture to map NGS short reads to genes and their paralogous segments. *L. Tian, C. Kao, F. Mafra, R. Pellegrino, H. Hakonarson.* The Children's Hospital of Philadelphia.

Mapping the repetitive regions of the genome remains challenging with current sequencing methods due to the reliance on short reads of ~200-400bp. About 10% (1,998 genes) of the exome contain repeat that is at least 250 bp in length and exhibits >98% identity to a paralogous segments located elsewhere in the genome. Of genes involved, 286 of them belong to the Medical Exome. It was further narrowed down to 73 genes that 1) have low coverage in ExAC, 2) are non-unique region by SiRen, and 3) are non-unique region by mappability files generated by GA4GH and the Genome in a Bottle Consortium. Here, we present a methodology to systematically resolve these repetitive regions using the 10x Genomics Chromium platform coupled with Agilent's SureSelect target enrichment with optimized baits. The optimized exome contains "bridging baits" that 1) are spaced within intronic and intergenic regions to preserve long-range info following capture, and 2) can capture uniquely mappable short reads to provide unambiguous reference genome location. The 10x's Lariat, a linked reads aware short reads aligner, was used to resolve the multi-reads in the repetitive regions. We show that 37 genes are completely resolvable (MAPQ≥20) when pre-processed via Chromium + optimized baits. Nine genes are partially resolvable, but can be fully resolved with improved bait design. 14 genes have paralogs that are too close (<50kb) that require improved bait design and lower MW DNA input. Nine genes have high G/C content that may improve by optimizing/increasing baits. Five genes have short repetitive sequences within the genes/exons and likely are not resolvable with linked reads, whereas the rest of them are upon optimization.

3186F

Untargeted metabolomics provides an unbiased approach to diagnosis of rare inborn errors of metabolism and further insight into rare disorders like transaldolase deficiency. *L. Hubert[1], L.H. Rodan[2], T. Donti[1], V.R. Sutton[1], G.T. Berry[2], Q. Sun[1], S. Elsea[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA.

Diagnosis of inborn errors of metabolism typically requires knowledge and experience with these rare disease entities and an understanding of the spectrum of the phenotype, such that targeted testing can be undertaken. Transaldolase deficiency is a rare inborn error of the pentose phosphate pathway caused by mutations in *TALDO1*. Clinical presentation includes coagulopathy, liver dysfunction, dysmorphic features, nephropathy, cutis laxa, splenomegaly, hemolytic anemia and genitourinary malformations. Biochemically, transaldolase deficiency results in the accumulation of erythritol, D-arabitol, ribitol and sedoheptulose in the body fluids of affected patients, and diagnosis is made when targeted analysis reveals elevations of these compounds, reduced/deficient *in vitro* TALDO activity, and/or molecular analysis of *TALDO1*. Despite these established methods of detection, diagnosis of transaldolase deficiency may be delayed due to its broad clinical spectrum, which can result in unfocused investigational efforts. To address the need for rapid and accurate testing for transaldolase deficiency and other inborn errors of metabolism, we developed a large scale analytical scheme capable of generating semi-quantitative z-score values for over 900 unique compounds from the analysis of a single specimen using untargeted metabolomic profiling. Applying this technique to analyze over 500 clinical samples, we were able to make an initial diagnosis in ~5% of cases, including transaldolase deficiency, where elevations of ribitol, erythronate, ribonate, arabitol were uniquely observed. Confirmation of this diagnosis was made by a GC/MS targeted analytical method, where elevated concentrations of several polyols and pentose phosphate pathway metabolites were detected. For example, plasma sedoheptulose was ~100-fold over normal in a previously published patient with homozygosity for the c.574C>T (p.Arg192Cys) *TALDO1* mutation (Rodan and Berry, JIMD Rep. 2016 Apr 30. [Epub ahead of print]). Compared to traditional methods which are time consuming and costly, the ability to identify a disorder with a single test in a large cohort of patients presenting with ambiguous clinical indications highlights the utility of untargeted metabolomic profiling. Thus, metabolomics may offer a route for more efficient diagnosis and earlier treatment of inborn errors of metabolism.

3187W

Combining enzymatic DNA fragmentation with NGS library construction results in high quality, high yield, libraries. *L. Apone, V. Panchapakesa, K. Duggan, T. Shtatland, B. Langhorst, J. Murdoch, C. Sumner, C. Rozzi, P. Liu, K. Krishnan, D. Rodriguez, J. Bybee, D. Rivizzigno, L. Mazzola, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, Inc., Ipswich, MA.

The use of Next Generation Sequencing (NGS) data has been instrumental in advancing our understanding of human genetics and identifying the molecular events that contribute to human disease. Continued advancement relies on overcoming the limitations and bottlenecks associated with NGS. We have focused on NGS library preparation, where the requirement for numerous steps and expensive equipment, can lead to sample loss, errors, and limited throughput. More specifically, we have developed a library construction method that combines enzymatic DNA fragmentation with end repair and dA-tailing in a single step. Integrating these reactions eliminates the need for costly equipment to fragment DNA and reduces the number of sample transfers and losses. Adaptor ligation is carried out in the same tube, after which a single cleanup step is performed. For low input samples, PCR amplification is performed prior to sequencing. This method is compatible with a broad range of DNA inputs and insert sizes. Libraries generated using this streamlined method with inputs ranging from 0.5ng to 500ng of intact DNA show no significant difference in coverage uniformity or sequence quality metrics, compared to libraries generated with mechanically sheared DNA. Likewise, libraries generated to contain insert sizes that range from 150bp to 1kb, display no significant difference in sequence quality from each other or those generated with mechanically sheared DNA. Finally, this streamlined method generates libraries of substantially higher yields than those generated using mechanically fragmented DNA, allowing the use of lower DNA inputs. The ability to generate high quality NGS libraries from intact DNA without the need for costly equipment and numerous cleanup or liquid transfer steps substantially reduces the time, cost and errors associated with library construction. In addition, these advances permit greater use and adoption of NGS technologies.

3188T

Comparison of synthetic long read sequencing methods and optical mapping for *de novo* **genome assembly.** *D.W. Mohr[1], A.F. Scott[1], A. Naguib[2], N.I. Weisenfeld[3], V. Kumar[3], P. Shah[3], D. Church[3], D.B. Jaffe[3].* 1) McKusick Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore MD 21287; 2) BioNano Genomics, Inc. , 9640 Towne Centre Drive, Suite 100, San Diego, CA 92121; 3) 10X Genomics, Inc., 7068 Koll Center Pkwy #401, Pleasanton, CA 94566.

We have been studying methods for assembling high quality genomes from non-human mammals. Current short-read methods have come to dominate genome sequencing because they are cost-effective, rapid, and accurate. However, short reads are most applicable when data can be aligned to a known reference, generally assembled with more traditional methods. Because such methods are time-consuming, costly, and inefficient, we have explored approaches to *de novo* genome assembly using several novel approaches for sequencing and assembly. In particular, we evaluated the Illumina 384-index (aka, "Moleculo") method, the 250 bp paired-end DISCOVAR sequencing protocol, the 10X Genomics Chromium Linked-Read sequencing method with ~1M-indexes, together with the BioNanoGenomics (BNG) optical mapping and hybrid assembly approach. We show that the Chromium method, assembled with SuperNova, provided the longest sequence blocks and, when used for de novo assembly, produced scaffolds with an N50 (>10 kb) of ~15 Mb with the longest individual scaffold >60 Mb. When combined with BNG optical maps the scaffold N50 was >20 Mb and the longest individual scaffold was >70 Mb. Because both BNG and 10X technologies interrogate single DNA molecules they can also be used to construct haplotypes and to detect larger scale structural variants between parental chromatids. Combining the two orthogonal methods of Chromium Linked-reads with BNG optical maps is likely to make the assembly of high quality genomes routine and significantly improve our understanding of comparative genome biology.

3189F

Genome-wide poly(A) tail length profiling using novel Tail End Displacement sequencing (TEDi-seq). *Y. Woo, J.T. Lis, H. Kwak.* Molecular biology and genetics, Cornell University, Ithaca, NY.

Poly(A) tails play a critical role in mRNA stability and are closely involved in the regulation of gene expression in early development and cell cycle. Earlier biochemical studies of poly(A) tails showed that they can be synthesized up to the length of ~250 nucleotides (nt). Recently, sequencing-based techniques such as TAIL-seq and PAL-seq have provided genome-wide measurements of poly(A) tail lengths with the median of ~50 nt in mammalian cells. However, it was still difficult to identify the association between translation efficiency and poly(A) tail length from these results, and unclear why there was a discrepancy between observed poly(A) tail lengths (~50 nt) and early biochemical studies. Furthermore, both sequencing techniques require extensive repurposing of sequencing devices which may limit the feasibility of the assay and produce unexpected biases.Therefore, we developed a novel method for more accessible genome-wide profiling of poly(A) tail length named Tail End Displacement sequencing (TEDi-seq). Here, we applied a precise size selection of hydrolyzed 3' mRNA fragments including the poly(A) tail and sequenced the RNA fragments. In TEDi-seq, the 5' ends of the sequence reads displaced by the fragment size reflect the poly(A) tail length. As a result, we found a median poly(A) tail length of ~ 50 nt in human lymphoblastoid cell lines, and a possible portion of mRNAs with longer poly(A) tail length. Our data suggest that a subpopulation of mRNAs with longer poly(A) tail length exists and may be differentially regulated in mammalian cells. Finally, our novel system may be a potential tool to measure poly(A) tail efficiently and accurately, and contribute to elucidate the dynamic roles of mRNA stability and translational control dependent on poly(A) tail length.

3190W

Leapfrogging genome assemblies by sequencing single cells. *M. Hills[1], E. Falconer[1], K. O'Neil[2], A. Sanders[1,3], P.M. Lansdorp[1,4,5].* 1) Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada; 2) Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada; 3) Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; 4) Division of Hematology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada; 5) European Research Institute for the Biology of Ageing, University of Groningen, The Netherlands.

   The mouse and human genome assemblies have revolutionized the way we conduct molecular research, rapidly expanding our understanding of the biology of these organisms and the diseases associated with them. These references are of high quality but still contain fragments which have not been localized to the genome and large regions that are incorrectly orientated with respect to the rest of the chromosome. Using Strand-seq, a single cell next-generation sequencing technique, we identified errors in the latest assemblies of mouse and humans with the goal of improving and refining these genomes. Strand-seq involves only sequencing parental DNA template strands in single daughter cells, preserving the directionality of DNA. Across multiple single cells, orientation errors and chimeric assembly fragments were readily detected as changes in the strand directionality across a chromosome, while unmapped fragments were located to specific regions. The reference assemblies of most other model organisms that are important for studies of human biology and disease are far less complete. For example, the ferret (*Mustela putorius furo*) is a model for the study of human respiratory diseases, including influenza infection and transmission, but the assembly is still in the draft stage, consisting of ~8,000 unplaced contigs. The pig (*Sus scrofa*) is important medically, representing a model for human pathologies and infectious diseases, and as a surrogate model for human wound healing and gastrointestinal tracts. In contrast to the ferret, the pig has scaffolds ordered into its 20 chromosomes, but still contains ~5,000 unplaced scaffolds and almost 75,000 gaps. More common model organisms also have important implications in human health, but still lack high-quality assemblies, with multiple gaps and unplaced fragments (for example *Danio rerio*), or are at earlier scaffold stages (for example *Xenopus tropicalis* and *Cavia pocellus*). We therefore used Strand-seq to rapidly and inexpensively create more complete genome assemblies for these five model organisms. We identified extensive orientation errors (for example, ~30% of the pig assembly was misoriented), and were able to cluster the thousands of fragments present in scaffold-stage assemblies into individual, ordered chromosomes. We demonstrate how Strand-seq allows us to create accurate reference assemblies rapidly and efficiently, which in turn aids research into human disease, pathologies and evolution.

3191T

Long range applications with short read sequencing. *S.T.K. Garcia, B.D. Galvin, H. Heaton, P. Hardenbol, C. Jabara, D.B. Jaffe, S. Kyriazopoulou-Panagiotopoulou, V. Kumar, P. Marks, H. Ordonez, M. Pratt, M. Schnall-Levin, P. Shah, M. Sooknah, A.W. Xu, N.I. Weisenfeld, I. Yousif, D.M. Church.* 10X Genomics Pleasanton, CA.

   High-throughput sequencing has revolutionized genome analysis. However, it is clear that traditional short read methods provide an incomplete view of the genome and result in an incomplete understanding of the clinical and biological complexity present. In particular, the lack of long-range information combined with inherent limitations in the mapping of short reads severely limits the robust identification of structural variants, haplotypes, and variants in difficult regions of the genome. To address these problems, we developed a technology that retains long-range information while maintaining the power, accuracy, and scalability of short read sequencing. At its core, haplotype-level dilution of long input molecules into >1 million barcoded partitions creates a novel data type referred to as 'Linked-Reads' that enables high-resolution genome analysis with minimal DNA input (~1 ng). Coupling Linked-Reads with novel algorithms that take advantage of these linkages allows for improved individual genotype reconstruction without sacrificing variant calling accuracy. We obtain 99% SNP sensitivity with greater than 99% PPV in confident regions for genome and exome on the NA12878 sample. We see phase blocks with an N50 of 4.1 Mb for genomes and 222 Kb for exome. We also show improved performance in regions of the genome typically inaccessible due to the presence of paralogous sequence/highly repetitive regions. Highly homologous sequences traditionally leading to ambiguous mapping can now be analyzed when associated with distinct barcodes. We estimate that we can rescue ~50 Mb of previously inaccessible sequence. We demonstrate that clinically-relevant variants can be identified in these previously inaccessible regions in genes including *CYP2D6*, *SMN1*, and *STRC*. Finally, retention of long range information facilitates the identification of copy number variants, copy neutral inversions, inter- and intra-chromosomal events, and more complex structural rearrangements. For a representative NA12878 genome 17 large SV calls were made, 16 of which had evidence from orthogonal methods, assembly, or raw reads. Importantly, 10 of the 17 events were fully or partially phased with single nucleotide variants. The utilization of Linked-Reads significantly improves our ability to reconstruct individual haplotypes and provides confidence in variant calls by providing an additional data type- phasing.

3192F

Ultra-sensitive mosaic mutation detection for clinical applications.
A. Hoischen[1], M. Steehouwer[1], M. Kwint[1], R. Acuna Hidalgo[1], E. Mersy[2], A. Paulussen[2], M. van de Vorst[1], L.E.L.M. Vissers[1], M. Nelen[1], C. Gilissen[1], J.A. Veltman[1,2], J. Shendure[3], A. Goriely[4]. 1) Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands; 3) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; 4) Weatherall Institute of Molecular Medicine, University of Oxford, UK.

Despite the great advances in the next generation sequencing field there is still room for improved targeted re-sequencing assays that combine high throughput with ultra-high sensitivity. We have now further optimized a single-molecule molecular inversion probe (smMIP) based targeted re-sequencing approach[1]. Single-molecule tracing is enabled using up to 4^{10} (1,048,576) molecular tags. Consensus calling of respective PCR-duplicates allows correction for PCR and sequencing errors. The improved assay allows low-frequency or sub-clonal variant detection with variant levels of <0.05%. This assay provides very robust genotyping accuracy, high throughput, fast turnaround and cost-effectiveness. We anticipate that this or similar assays allow novel applications in which mutations are present in very low relative abundance in any given DNA sample with important new applications beyond cancer genetics. Here we present first successful applications that include: 1.) Accurate determination of the fraction of mutated alleles for post-zygotic *de novo* mutations[2]. 2.) Detection of previously unrecognized mosaic disease causing mutations for rare clinical syndromes. 3.) Detection of known 'paternal age effect disorders' causing mutations as small clonal events in dissected testis material[3]. 4.) First evidence for presence or absence of parental alleles in cell free DNA from plasma of pregnant women.   References: 1) Hiatt JB, et al. Genome Res. 2013 May;23(5):843-54. 2) Acuna-Hidalgo R, et al. Am J Hum Genet. 2015 Jul 2;97(1):67-74. 3) Goriely A, Wilkie AO. Am J Hum Genet. 2012 Feb 10;90(2):175-200.

3193W

Challenges and opportunities in implementing pharmacogenomic clinical decision support tools within diverse Electronic Medical Record (EMR) systems. *J. Ross, D. Bailey, A. Solomon.* OneOme, LLC, Minneapolis, MN.

There is growing evidence that pharmacogenomic testing can significantly reduce a patient's overall healthcare costs and greatly improve their outcomes.[1,2] Recently, the cost of pharmacogenomic testing has fallen under $250 for multi-gene panels providing increased accessibility to a broader range of patients and providers. However, results are often returned as a static report. These reports are manually scanned into the EMR and reviewed initially but then are treated similarly as other temporal based results quickly relegated to the EMR's archives. This provides very limited value to the patients because the report is only used to address the patient's present situation and doesn't leverage the knowledge in a data build that can then be used to address future prescription complexities and changes. In addition, referencing the document within the EMR system can be challenging, often significantly reducing the utility of the test results for the patient and provider. In order to provide value and maximize effectiveness of pharmacogenomic testing, the tests should be easy to order and results must be integrated into the EMR system as structured data, readily available, thus allowing clinical decision support tools the ability to continually access the information which may be clinically relevant and potentially actionable for managing clinical care for the lifetime of the patient.We will demonstrate how structured pharmacogenomic information integrated with EMR systems such as EPIC and Cerner impact the utility and accessibility, and how OneOme's clinical decision support tools, co-developed and exclusively licensed from Mayo Clinic, can provide ongoing, relevant and actionable information designed to improve patient outcomes.   [1] D. Brixner, E. Biltaji, A. Bress, S. Unni, X. Ye, T. Mamiya, K. Ashcraft & J. Biskupiak (2016) The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy, Journal of Medical Economics, 19:3, 213-228, DOI: 10.3111/13696998.2015.1110160 [2] O Alagoz, D Durham and K Kasirajan Cost-effectiveness of one-time genetic testing to minimize lifetime adverse drug reactions The Pharmacogenomics Journal (2016) 16, 129–136.

3194T

Quantification of genome-wide nascent mRNA by click chemistry and RNA-seq. *W. Wang[1], R. Song[2,3], M. MacGibeny[2,3], J. Wiggins[1], J. Miller[1], L. Enquist[2,3].* 1) Institute for Integrative Genomics, Princeton University, Princeton, NJ; 2) Department of Molecular Biology, Princeton University, Princeton, NJ; 3) Princeton Neuroscience Institute, Princeton University, Princeton, NJ.

   RNA-seq has been widely used to quantify transcript abundance; however, in most cases it is applied directly on all RNA transcripts existing in the samples at certain time points to measure the steady-state transcript levels. Partitioning the whole RNA sample into newly synthesized RNA over a period of time and pre-existing RNA would enable characterization of the dynamics of the transcriptome. Such RNA partition can be achieved by incubating live cells with an alkyne-modified nucleoside 5-ethynyl uridine (EU) to label the nascent RNA, followed by click reaction with an azide-modified biotin to capture the EU-containing nascent RNA with streptavidin magnetic beads. In the current study, SMART-Seq2 method, originally developed to amplify and sequence poly-A containing RNA transcripts from single eukaryotic cells, was applied on the small amount of EU-labeled bead-captured RNA from primary cell culture to characterize the nascent mRNA synthesis.   Compared to the input RNA before capture, amplified cDNA from EU-labeled RNA demonstrated shorter fragment length, likely due to RNA degradation during the lengthy hybridization capture process. Linear correlation was clearly observed in the transcript abundance on log scale between EU-labeled RNA and input RNA, while the spread was wide. Control RNA captured from cells without EU treatment (unlabeled RNA) showed much lower cDNA amplification yield, which was barely detectable. However, its transcript abundance still demonstrated correlation to the input RNA, especially for genes with high expression levels, revealing the non-specific capture of unlabeled RNA by streptavidin beads. In conclusion, it is feasible to quantify genome-wide nascent mRNA abundance by combing Click chemistry with mRNA amplification and sequencing analysis. There is still room for improvement in the bead capture specificity and efficiency.

3195F

A highly efficient, low-bias and strand-specific RNA-seq library preparation method for transcriptome profiling. *E. Yigit, K. Karaca, K. Krishnan, B. Langhorst, T. Shtatland, D. Munafo, D. Rodriguez, P. Liu, L. Apone, V. Panchapakesa, K. Duggan, C. Sumner, C. Rozzi, F. Stewart, L. Mazzola, J. Bybee, D. Rivizzigno, E. Dimalanta, T. Davis*
. New England Biolabs, Ipswich, MA.

   RNA-seq (RNA sequencing) has undoubtedly become the most popular method for transcriptome analysis. It is widely used for gene expression analysis including detection of mutations, fusion transcripts, alternative splicing, and post-transcriptional modifications. Recent improvements in next generation sequencing technologies (NGS) and sample barcoding strategies allow analysis of multiple samples in parallel in a cost effective manner. As RNA-seq is increasingly adopted for molecular diagnostics, the quality and reproducibility of library preparation methods become more important. In addition, demand for library preparation methods that produce successful NGS libraries from low input RNA or precious clinical samples is increasing. To overcome these challenges, we have developed a strand-specific RNA-seq library preparation method that retains information about which strand of DNA is transcribed. Strand specificity is important for the correct annotation of genes, identification of antisense transcripts with potential regulatory roles, and for correct determination of gene expression levels in the presence of antisense transcripts. Our method can be used for a wide range of input RNA (10-1000 ng total RNA) without any modifications to the protocol, making it a convenient method for RNA-seq library preparation. More importantly, GC content analysis, gene body coverage and gene expression correlation show that there are no significant differences between libraries with varying inputs, even though input amounts varied by two orders of magnitudes. As a result, our method has increased sensitivity and specificity, especially for low-abundance transcripts, and reduced PCR duplicates and sequence bias, delivering high quality strand-specific data. Our method is compatible with both poly A-tail enriched and ribosomal RNA depleted samples, and is amenable to large-scale library construction and automation.

3196W

The Korean Biobank Array project. *Y. Kim, S. Moon, M. Hwang, B. Han, B. Kim.* Center for Genome Science, National Institute of Health, Heungdeok-gu, Chungcheongbuk-do, South Korea.

Recently, customized genotyping chips with low cost have gathered increasing attention to conduct a large-scale population based genomic study. Generally, customized chips are designed to contain variants for a specific research purpose or variants selected from the discovered variants of sequencing data. For example, biobank arrays such as a custom array for Kaiser permanente/UCSF genotyping project, UK biobank array, Japonica array, and China Kadoorie biobank array have been introduced to perform genome-wide association study on complex traits by genotyping more than tens of thousands of samples. In 2014, Korea National Institute of Health initiated the Korean biobank array(Korean chip) project. The Korean chip project aims to study hidden genetics of various diseases by genotyping more than 100,000 samples from three prospective cohorts. At the first phase of the Korean chip project(2014-2015), Korean chip was designed and genotyping of 35,000 samples using the Korean chip was completed in 2015. Korean chip contains about 833K markers including genome-wide taggers and about 200K of functional variants. In the preliminary analysis using about 7,000 samples, Korean chip showed better genomic coverage over widely used commercial arrays (95% for MAF ≥ 5%, 73% for MAF 1-5%). Also Korean chip discovered novel associations for liver enzymes and lipid traits and numerous nonsynonymous variants within the previously reported GWAS loci.

3197T

Universal sample preparation for omics–based applications in personalized medicine. *L. Poveda, J. Tracy, L. Opitz, N. Selevsek, R. Schlapbach.* Functional Genomics Center Zurich, ETH Zurich/University Zurich, Switzerland.

Combining "Omics" information from the analysis of diverse biomolecules, such as nucleic acids and proteins, opens up new opportunities in the field of personalized/precision medicine where patients can be monitored across multiple molecular readouts to improve disease diagnosis and treatment. Ideally, the data should be collected from the same sample using a universal sample preparation method, which enables the extraction of the different classes of biomolecules for sensitive analysis. Recently, the pressure cycling technology (PCT) has emerged as an efficient sample preparation strategy, enabling the extraction of proteins from a small amount of samples such as needle tissue biopsies. Here, we performed the lysis of small amounts of porcine tissue biopsies using the PCT approach, enabling the downstream analysis of proteins, RNA and DNA. In total, more than 300 molecular readouts were generated from 7 different tissue types (pig liver, heart, aorta, spleen, lung, aorta, brain) processed in four replicates and at four different tissue weights (0.2, 0.5, 1 and 3 mg) each. Estimated protein yields resulting from the PCT workflow were directly proportional to the tissue weights and demonstrated high reproducibility between replicates and organs. For nucleic acids, the PCT workflow was combined with standard DNA/RNA isolation kits enabling the purification of genomic DNA and total RNA from the samples. For both, high DIN and RIN numbers as measures for the integrity and quality of the nucleic acids were obtained across almost all tissue weights and organs, reflecting low degradation of the biomolecules during sample processing. After isolation, all tested samples were further processed for proteome and transcriptome profiling using mass spectrometry and Next Generation Sequencing analysis, respectively. As expected, a high correlation was identified between protein and transcript abundances for the same organ, but a lower correlation was measured across organs. Pathway enrichment analysis revealed several activated molecular pathways that were found to be expressed strictly organ and tissue-specific. The presented robust and sensitive method for the isolation of multiple molecular species from minimal samples provides the basis for an integration of complex data sets from proteomics and genomics for a comprehensive view on cells or organs under defined physiological states, a prerequisite in system biology and personalized medicine.

3198F

Comparison of polyA-selected, ribo-minus, and exon capture RNA-sequencing in splicing events detection. *S. Peng[1], M. Seiler[1], J. Wu[1], J. Zou[2], R. Darman[2], P. Kumar[2], C. MacKenzie[2], J. Palacino[2], L. Yu[1].* 1) Data Science and IT, H3 Biomedicine Inc., Cambridge, MA; 2) H3 Biomedicine Inc., Cambridge, MA.

Defects in pre-mRNA splicing cause or contribute to various human diseases. Large scale cancer genome projects also identified mutations in spliceosome components, e.g. SF3B1, U2AF1, or SRSF2, initially in myeloid-lineage malignancies and, more recently, in breast cancer and melanoma. These mutations result in frequent selective aberrant mRNA splicing events, including exon skipping / inclusion (SRSF2 & U2AF1) or alternative 3' splice site (SF3B1). RNA-seq is used to profile transcriptomes and identify novel transcripts and alternative splicing events. RNA-seq libraries can be prepared from total RNA using different protocols, such as polyA-selection, ribo-minus, and exon-capture. Current publications comparing different RNA-seq library preparation methods focus on the gene expression profiles, but neglected to compare detection power of different methods for splice junctions, which is critical when studying mutational or interventional splicing modulation.To compare splicing detection capacities of different RNA-seq protocols, we obtained data with polyA (pA), ribo-minus (RZ) and TruSeq RNA Access (ExCap) in K052 cells, which carries SRSF2.p.P95H mutation.

	pA(+)	RZ	ExCap
% Unique Reads	90.8	72.6	84.6
% Splice Junction Reads	32	18	35
# Unique Junctions	215000	198000	154000

The overlapping percentages of splice junctions detected when comparing between two methods range from 63 to 87%. When limiting the comparison to the exon-capture target regions, the overlapping percentages improve, and range from 79% to 90%. In ribo-minus sample, a large number of reads are mapped to non-polyA RNAs, such as histone mRNAs, lncRNAs and snoRNAs, which decrease the detection power of splice junctions. The conclusion holds true when we examined several reported aberrant splicing events associated with SRSF2.p.P95H. Overall, PolyA-selected RNA-seq offers the broadest splice junction coverage, including identification of novel junctions. Ribo-minus RNA-seq provides information on non-polyA transcripts, but needs higher sequencing coverage to reach similar detection power for splice junctions. Exon-capture based RNA-seq (RNAaccess) is not recommended for global splicing profiling and discovery, however, it can be a viable alternative to profile known splicing events.

3199W

RD-Connect: Data sharing and analysis for rare disease research within the integrated platform and through GA4GH Beacon and GA4GH/IRDiRC Matchmaker Exchange. *S. Beltran[1,2], D. Piscia[1,2], S. Laurie[1,2], J. Protasio[1,2], A. Papakonstantinou[1,2], A. Cañada[3,4], J.M. Fernández[3,4], M. Thompson[5], R. Kaliyaperumal[5], S. Lair[6], P. Sernadela[7], T. Katsila[8], G.P. Patrinos[8], O.J. Buske[9], M. Girdea[9], M. Brudno[9], A. Blavier[6], J. Dawson[10], R. Thompson[10], H. Lochmüller[10], M. Bellgard[11], D. Spalding[12], M. Roos[5], P.A.C. 't Hoen[5], A. Valencia[3,4], D. Salgado[13,14], C. Béroud[13,14,15], I. Gut[1,2], RD-Connect Consortium.* 1) Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain; 4) Instituto Nacional de Bioinformática (INB), Spain; 5) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 6) Interactive Biosoftware, Rouen, France; 7) DETI/IEETA, University of Aveiro, Portugal; 8) Department of Pharmacy, School of Health Sciences, University of Patras, University Campus, Rion, Patras, Greece; 9) Centre for Computational Medicine, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada; 10) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK; 11) Centre for Comparative Genomics, Murdoch University, Perth, Western Australia; 12) European Molecular Biology Laboratory , European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom; 13) Aix-Marseille Université, Marseille, France; 14) Inserm, UMR_S 910, Marseille, France; 15) APHM, Hôpital TIMONE Enfants, Laboratoire de Génétique Moléculaire, Marseille, France.

RD-Connect (rd-connect.eu) is a platform connecting omics data, registries, biobanks and clinical bioinformatics for rare disease research (platform. rd-connect.eu). RD-Connect aims to contribute to the IRDiRC objectives for 2020 and to narrow the gaps in rare disease research, where patient populations, clinical expertise and research communities are small in number and highly fragmented. The RD-Connect platform securely integrates -omics data with biosample and clinical information, providing not only a centralised data repository but also a sophisticated and user-friendly online analysis system. Raw Whole Genome, Exome or gene panel NGS data are deposited at the European Genome-phenome archive and processed with RD-Connect's standardised analysis and annotation pipeline. The corresponding clinical information is recorded in a connected PhenoTips instance using the Human Phenotype Ontology, OMIM and Orpha codes. The results are made available to authorised users through the highly configurable and efficient platform, which runs on a Hadoop cluster and uses ElasticSearch. The user-friendly interface enables filtering and prioritization of variants using the most common quality, genomic location, effect, pathogenicity and population frequency annotations. Additional tools, such as Exomiser, DiseaseCard, Alamut Functional Annotation (ALFA), ePGA (epga.gr) and UMD Predictor (umd-predictor.eu) are integrated at several levels. To share data beyond the platform, RD-Connect has lit a GA4GH Beacon within the Beacon Network (beacon-network.org) and is actively involved in the development of the GA4GH/IRDiRC Matchmaker Exchange (matchmakerexchange.org, MME) 1-sided matching approach. The development of the MME API will enable more complex queries while still ensuring that no personal sensitive information is shared. There are over 2200 NGS samples already available in RD-Connect. The number is quickly increasing and we welcome new submissions.

3200T

Pan-microbial detection using Axiom® Microbiome Array. *P. Rack, M. Wu, L. Le, C. Sheppy, M. Mittmann, A. Pirani, M. Shapero.* Affymetrix, part of Thermo Fisher Scientific, Santa Clara, CA.

The human microbiome is defined by a diverse set of microbes, including eukaryotes, archaea, bacteria, and viruses, that can differ across body sites. These microbes play a key role in both human health and disease. As sequences of microbial genomes continue to expand, microarrays are well positioned to capitalize on this information content. Here we present an overview of Axiom® Microbiome Array, which enables researchers to detect all known microorganisms in a sample with a single assayAxiom Microbiome Array contains approximately 1.3 million probe sequences targeted towards more than 11,000 microbial organisms. The array design enables species- and strain-level resolution from individual samples and uses Axiom® assay biochemistry to interrogate non-polymorphic sequences in both family-conserved and target-specific regions from NCBI database sequences. Axiom™ Microbial Detection Analysis Software (MiDAS) powers the analysis workflow for Axiom Microbiome Array. Axiom™ MiDAS is built on a Composite Likelihood Maximization (CLiMax) algorithm developed at Lawrence Livermore National Laboratory, which is used to predict the identities of targets in an unknown sample.Array performance has been evaluated on both known complex mixtures and *bona fide* biological samples (stool) against stringent quality control metrics including positive predictive value (PPV), true positive rate (TPR), limit of detection (LOD), and reproducibility. TPR and PPV were evaluated on 222 samples of known composition with complexity varying from 1 to 22 strains per sample. LOD was determined using a broad log dilution series of known microbial genomes in the presence or absence of background human genomic DNA (gDNA). The results indicate a higher overall accuracy at the species level than 16S rRNA gene sequencing technology.In summary, Axiom Microbiome Array offers a solution for the identification and enumeration of microbial entities comprising complex biological samples. The scalable sample throughput enabled by 24- and 96-array formats, with the potential to expand to a 384-array format, coupled with laboratory automation allows processing of hundreds to thousands of samples per week with minimal manual intervention. This feature makes Axiom Microbiome Array a cost-effective means for detection and examination of microbiomes.

3201F

Saliva is a good alternative DNA source for whole genome and/or whole exome sequencing. *A. Thrasher[1], Z. Wang[1], H. Mulder[1], D. Hedges[1], T. Chang[1], C. Rosencrance[1], J. Easton[1], L. Robison[2], J. Zhang[1].* 1) Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN; 2) Department of Epidemiology & Cancer Control, St. Jude Children's Research Hospital, Memphis, TN.

The increased accessibility and decreased cost of next generation sequencing (NGS) has enabled an increasing number of genomic studies. Blood is the preferred DNA source for germline studies using NGS. However, drawing blood is an invasive process requiring trained personnel to interact directly with the sample donor. DNA from saliva is a more convenient and cost-effective alternative. The primary concern with DNA isolated from saliva has been the quality and suitability of this DNA for the generation of high quality non-array data, such as that obtained by next generation sequencing. To address this, matching blood and saliva samples were collected from five healthy subjects (age range 36-64). A series of DNA isolation optimization experiments were performed on the saliva samples, with phenol/chloroform isolated DNA from blood serving as the reference. Whole genome (WGS) and whole exome (WES) sequencing was performed on samples from optimized saliva DNA isolation and DNA from blood. The average coverage for WGS and WES was 30X and 100X respectively with >70% of the coding exons at >20X in both data sets. Genetic variations, including single nucleotide variation (SNV), small insertion/deletion, copy number variation (CNV) and structural variation (SV) are highly concordant between the two source materials with only 1 saliva-specific coding SNV across the five pairs. Comparison of the somatic variants identified exclusively in one source detected *TCRA* and *TARP* rearrangements in blood, as expected. Further, we performed metagenomic analysis of unmapped reads from both blood and saliva samples. Blood samples had 5-6% unmapped reads compared to 7-18% in saliva. Evaluation of source of unmapped reads found that 0.54% mapped to viral species in both sample types. In contrast, 10.3% of unmapped reads from saliva samples mapped to bacterial species, compared with only 1% of unmapped reads from blood samples as expected[1]. This analysis represents, to our knowledge, the first comprehensive examination of WGS and WES data generated from blood compared to saliva. Our analysis shows a high level of concordance between the two sample sources, and as expected few somatic variants. This indicates that high quality sequencing data can be derived from saliva samples for germline genetic analyses. [1] C. James, R.M. Iwasiow, H.C. Birnboim (2011). Human genomic DNA content of saliva samples collected with the Oragene self-collection kit. DNA Genotek. PD-WP-011 Issue 4/2011-11.

3202W

A custom mapping and annotation workflow for array-based genotyping applied to a large biobank cohort. *E.A. Tsai[1,2,3], C.-F. Lin[1,2,3], A. Brown[1], S.S. Amr[1,3,4], M.S. Lebo[1,3,4].* 1) Personalized Medicine, Partners HealthCare, Cambridge, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA; 4) Department of Pathology, Brigham and Women's Hospital, Boston, MA.

Background: In recent years, there is an increasing number of SNPs identified through next generation sequencing that are not typically included in genotyping arrays. Given the low cost of genotyping arrays and the clinical importance of these positions, it is valuable to include them on SNP arrays applied to large populations. However, as many of these SNPs are not currently present on the commercial array, they require novel design, additional quality control, and manual curation. There tends to be limited testing of the probe design and probe performance due to the sheer number of novel SNPs included, so this process requires careful quality control. Methods: 1) We created a workflow that remaps the probes from the Illumina manifest data to ensure that they optimally aligned to the targeted region. 2) We designed a custom, curated flat file that provides information about the variant not available in the traditional PED format, including variant level information such as the rsID, annotations to the nearest gene, and transcript-level consequences for coding variants. In addition, the normalized VCF nomenclature and observed call rates of each marker are also included. 3) We identified variants that would be potentially returnable in the ACMG incidental findings list of 56 genes using our annotated SNP file. Other components of the overall workflow include sample phasing, imputation, and easy access to subsets of the sample dataset. Results: We have implemented this workflow for Illumina Multi-Ethnic Genotyping Array (MEGA), Expanded Multi-Ethnic Genotyping Array (MEGA[EX]), and Multi-Ethnic Global (MEG) array. This family of biobank genotyping arrays have 1.8 to 2 million probes including custom-designed exome content (50,000+ loss of function variants) and an increased representation of SNPs present in non-Caucasian cohorts. Using the ACMG incidental findings gene list, we identified more than 1,500 variants genotyped that could potentially be a returnable finding if they were identified in an individual. Summary: We present a custom workflow for accurate mapping, strand determination, and annotation of genotyped data. This workflow is the standard mechanism for data processing and delivery for data generated via the Partners HealthCare Biobank Initiative. It was applied to genotyping results for 10,000 samples, and an additional 15,000 samples are in the queue. Further analysis will determine the number of returnable variants identified among this cohort.

3203T

Method for fetal fraction determination through INDEL targeted genotyping. *N.M. Kiesling, A. Kruger, J. Thompson, K.M. Klusman, B.M. Hanlon, T.P. Mann.* Progenity Inc, Ann Arbor, MI.

Noninvasive prenatal testing for fetal chromosome abnormalities relies on the presence of fetal DNA circulating in the mother's blood. The sensitivity of the test may be reduced in cases of low fetal fraction, the percentage of fetal DNA to maternal DNA, resulting in a false negative test outcome. The amount of fetal DNA present in the mother's blood is highly variable and dependent on factors such as gestational age, maternal BMI and smoking. NIPT results can be put into clearer context by accounting for the fetal fraction within the sample. The most widely accepted method for ascertaining the fetal contribution to the total DNA is through maternal and fetal genotyping, i.e. detection of paternally inherited alleles. The quantity of fetal DNA present in the sample has a direct relationship to the sensitivity and accuracy of the genotyping methodologies, thus enrichment techniques such as whole genome amplification (WGA) are often done to increase the quantity of DNA. However, genotyping amplified product can be complicated by several factors including the high probability of WGA and sequencing error to introduce point mutations into the product. These point mutations, when occurring in specific locations, can appear as alternate alleles and result in an erroneous genotype. Here we describe a method for fetal fraction determination through genotyping using WGA material that is designed expressly to eliminate sequence errors and reduce noise in allele ratios. This methodology begins by querying INDELs rather than SNPs for genotype determination as INDELs are much less likely to be affected by WGA and sequencing induced mutations than SNPs. Amplification and sequencing errors are further reduced through condensation of PCR duplicates into a consensus sequence. Finally, an accurate count of reference and alternate alleles is obtained through an alignment-free counting analysis, and fetal fraction is quantified by an expectation-maximization approach. We have demonstrated that INDEL targeted genotyping with an alignment-free analysis can accurately determine the proportion of fetal DNA in a pregnant plasma sample that has undergone WGA.

3204F

Size distribution, yield, and genomic mapping characteristics of cell-free DNA extracted from different donors, different body fluids, and at different time-points. *M. Toloue, M. Goldrick.* Bioo Scientific, Austin, TX.

Purpose: There has been a recent surge of interest in analysis of fragments of DNA recovered from cell-free biological fluids as they may allow non-invasive monitoring of malignant disease and other pathological conditions. Clinical studies have shown the feasibility of detecting genetic variants associated with recurrence of cancer after treatment, in cell-free DNA (cfDNA) extracted from blood plasma and urine, and there are many reported cases of earlier detection of metastatic disease by analysis of cfDNA, compared to conventional imaging tests. However, there have been relatively few cfDNA-based assays ("liquid biopsies") validated for clinical use, partly due to incomplete understanding of the biological processes that lead to release of DNA into biofluids. A substantial fraction of cfDNA is recovered as fragments of ~170 bp and multimers thereof, which is thought to reflect its association with the histone proteins that comprise nucleosomes. Larger fragments may also be recovered, even after removal of contaminating cells. To better define the characteristics of cfDNA, we carried out studies to compare size distribution, concentration, and genomic origin of cfDNA extracted from plasma and urine from a diverse group of healthy subjects. Methods: Platelet-poor plasma and clean-catch centrifuged urine samples were obtained from healthy donors representing different ages, genders, and ethnic groups. For some donors, multiple samples were obtained over a time-course. cfDNA was extracted using magnetic-bead-based kits (Bioo Scientific) and analyzed on an Agilent 2100 HS DNA chip to determine concentration and size distribution. A subset of samples was converted into whole-genome libraries and used for paired-end sequencing on an Illumina MiSeq. Sequencing reads were binned according to size of the cfDNA from which they originated, and mapping characteristics (genomic coverage and %GC) were determined for larger and smaller fragments. Results: Differences in yield and size distribution of cfDNA were observed for different biofluids, for different donors, and for sequential samples collected from the same donor. The proportion cfDNA larger than 500 bp could not be attributed to differences in sample collection or preanalytical processing. Variation in genomic mapping characteristics were seen in cfDNA extracted from different donors, from different size ranges of cfDNA, and from different samples from the same donor.

3205W

OneSeq phased exome for structural variant and compound heterozygote detection. *E. Lin[1], F. Useche[1], M. Isaksson[1], B. Galvin[2], H. Ordonez[2], J. Lee[2], A. Wei Xu[2], P. Marks[2], M. Schnall-Levin[2], D. Church[2], S. Garcia[2], M. Sooknah[2], M. Corioni[1], K. Jeong[1], A. Vadapalli[1], G. Amparo[1], M. Ramaiah[1], C. Pabon[1], D. Roberts[1].* 1) Genomics, Agilent Technologies, Santa Clara, CA; 2) 10x Genomics, Pleasanton, CA.

Massively parallel sequencing has been extremely successful in elucidating the genetic etiology of Mendelian disease and complex disorders such as cancer. Target enrichment applications such as whole exome captures have enabled higher sequencing depths to accurately determine single nucleotide variants (SNV), small insertions/deletions (INDEL), and chromosomal copy number changes. However, short-read sequencing provides only a few hundred nucleotides of uninterrupted sequence, preventing the assignment of successive variants along the chromosome to a specific allele. Moreover, structural variants can only be identified if reads are mapped across a breakpoint. This is difficult especially if the breakpoint occurs in repetitive regions. To address these limitations, we have developed the OneSeq Phased Exome Target Enrichment System for use with the Chromium™ linked-read technology from 10X Genomics. The phased exon capture is aided by selected probes targeting intronic regions. Multiple reads derived from a single long DNA molecule are linked together and aligned, providing a phased scaffold across the exons. This allows segmentation of variants into long haplotype blocks and the phasing of entire genes. Linked reads also reduce alignment ambiguities, thus improving coverage in regions that are traditionally difficult to map. Further, linked reads can identify breakpoints and chromosomal translocations. Herein we describe the wetlab, instrumentation, and informatics workflows for combined 10X library prep and exome capture. We describe phasing performance and demonstrate applicability for determining both structural variants and compound heterozygotes. The addition of phasing and structural variant determination to traditional exome analysis will aid in the elucidation of complex genotypes and drive research and understanding of disease causality.

3206T

Effect of coverage depth and haplotype phasing on structural variant detection in human genomes from PacBio long reads. *A.M. Wenger, S. Kujawa, L. Hickey, C.S. Chin, J. Korlach.* Pacific Biosciences, Menlo Park, CA.

Each human genome has thousands of structural variants compared to the reference assembly, up to 85% of which are difficult or impossible to detect with Illumina short reads and are only visible with long, multi-kilobase reads. The PacBio RS II and Sequel single molecule, real-time (SMRT) sequencing platforms have made it practical to generate long reads at high throughput. These platforms enable the discovery of structural variants just as short-read platforms did for single nucleotide variants. Numerous software algorithms call structural variants effectively from PacBio long reads, but algorithm sensitivity is lower for insertion variants and all heterozygous variants. Furthermore, the impact of coverage depth and read lengths on sensitivity is not fully characterized.   To quantify how zygosity, coverage depth, and read lengths impact the sensitivity of structural variant detection, we obtained high coverage PacBio sequences for three human samples: haploid CHM1, diploid NA12878, and diploid SK-BR-3. For each dataset, reads were randomly subsampled to titrate coverage from 0.5- to 50-fold. The structural variants detected at each coverage were compared to the set at "full" 50-fold coverage. For the diploid samples, additional titrations were performed with reads first partitioned by phase using single nucleotide variants for essentially haploid structural variant discovery.   Even at low coverages (1- to 5-fold), PacBio long reads reveal hundreds of structural variants that are not seen in deep 50-fold Illumina whole genome sequences. At moderate 10-fold PacBio coverage, a majority of structural variants are detected. Sensitivity begins to level off at around 40-fold coverage, though it does not fully saturate before 50-fold. Phasing improves sensitivity for all variant types, especially at moderate 10- to 20-fold coverage. Long reads are an effective tool to identify and phase structural variants in the human genome. The majority of variants are detected at moderate 10-fold coverage, and even extremely low long-read coverage (1- to 5-fold) reveals variants that are invisible to short-read sequencing. Performance will continue to improve with better software and longer reads, which will empower studies to connect structural variants to healthy and disease traits in the human population.

3207F

Rapid, high-throughput clinical sequencing and reporting for personalized medicine. *D. Muzny[1], K. Walker[1], J. Hu[1], V. Korchina[1], R. Raj[1], Y. Ding[1], H. Doddapaneni[1], Q. Wang[1], J. De La Cruz[1], Y. Yang[2,3], M. Leduc[2,3], C. Eng[2,3], D. Ames[4], A. Carroll[4], M. Murugan[1], C. Kovar[1], W. Salerno[1], E. Boerwinkle[1,5], E. Venner[1], R. Gibbs[1].* 1) Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX 77030; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX 77030; 3) Baylor Miraca Genetics Laboratories, 2450 Holcombe Blvd, Houston TX 77021; 4) DNAnexus, Mountain View, CA 94040; 5) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77225.

Advancement of NGS applications in the clinical diagnostic arena requires methods for rapid, robust delivery of high-quality sequencing data covering clinically relevant genes and HIPPA compliant cloud-based tools for sample accessioning and reporting along with robust analysis pipelines for clinical applications. These combined laboratory and analytical methods enable rapid data generation, variant annotation, diagnostic interpretation and generation of clinical reports for large studies (>10,000 patient samples) using capture panels (~500kb) within a CAP/CLIA environment.   Our pipeline utilizes a '*lightning capture*' process to deliver variant calls in 5-7 days after sample intake. The '*lighting capture*' process includes: quick enrichment library preparation, (5-6 hours) and capture enrichment (approximately 14 hours); rapid sequencing (Illumina HiSeq2500) and data analysis via the HGSC-developed clinical pipeline - Neptune. Methods employ a cost-effective 47plex co-capture format for regional capture hybridization followed by sequencing 94 samples (2 capture pools) per HiSeq 2500 lane for data generation. To ensure integrity of the end-to-end process, we have implemented the Fluidigm SNPTrace assay by Fluidigm to confirm sample identity and detect potential sample cross-contamination.   Application of this pipeline includes a capture panel (CAREseq, 535kb) of 109 genes and ~1500 unique single nucleotide variant positions with known common genetic variants of clinical relevance for the National Institutes of Health eMERGE network. This clinical capture design and process has been validated utilizing 647 control samples. High enrichment efficiency was observed (>79% reads on target and buffer) and superior coverage metrics (99.7% of target bases ≥20x coverage, and 347x average coverage). A detailed analysis of the design performance using 42 de-identified samples found that known mutations were correctly identified with high confidence (100%), including CNV mutations. Coverage analysis of the 647 validation test samples, found only 14 exons with low coverage using the criteria of ≥10% of samples with coverage <20X, demonstrating the high quality of the capture reagent and process. This pipeline, including secure cloud-based sample accessioning and automated clinical reports for review will be used to process >12,500 samples for the eMERGE program. .

3208W

MTTE: An innovative strategy for the evaluation of targeted/exome enrichment efficiency. *K. Klonowska, L. Handschuh, A. Swiercz, M. Figlerowicz, P. Kozlowski.* European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Recently, next-generation sequencing (NGS) has become the leading method for analyzing the architecture of human genomes. Although currently available strategies for the preparation of exome-enriched libraries are well established, a final validation of the libraries in terms of exome enrichment efficiency prior to the sequencing step is of considerable importance. Here, we present a strategy for the evaluation of exome enrichment, i.e., the Multipoint Test for Targeted-enrichment Efficiency (MTTE), based on the well-validated protocol of multiplex ligation-dependent probe amplification (MLPA) method. Our assay for post-capture exome enrichment validation is composed of 20 MLPA probes, including 10 probes located in targeted genomic regions (mostly exons of protein coding genes and one region overlapping miRNA sequence), 9 probes located in non-targeted genomic regions (introns and intergenic regions), and one probe located in flank of the targeted regions. The MLPA probe set was designed according to a strategy developed previously in our group, allowing easy design and generation of assay for the analysis of almost any region of interest. We used the MTTE assay for the evaluation of enrichment efficiency at five consecutive steps of the Illumina TruSeq Exome Enrichment procedure, including libraries after the first and second steps of enrichment. The calculated values of enrichment-associated parameters (i.e., relative enrichment, relative clearance, overall clearance, and fold enrichment) and the comparison of MTTE results with the actual enrichment calculated based on the fraction of NGS reads mapping to the targeted regions revealed the high reliability of our assay. Additionally, we took advantage of our MTTE results to determine the potential sequence-associated features that may confer bias in the enrichment of different capture targets. The identified features (e.g., content of purines R=0.70, p=0.02) add to the current knowledge regarding factors that influence exome enrichment efficiency. Importantly, the MTTE strategy can be easily adapted to the region of interest important for a particular project, e.g., if particular genes are sequenced, target-specific probes may be located in each exon of the gene. The additional advantage of MTTE is its low cost (~5 USD per sample). Thus, the MTTE strategy is attractive for post-capture validation in a variety of targeted/exome enrichment NGS projects. Acknowledgements: NCN 2015/17/B/NZ2/01182.

3209T

Advances in small RNA NGS: Low input samples and automation. *S. Shore[1], J.M. Henderson[1], J.F. Quintana[2], F.W.N Chow[2], K. Chhatbar[2], G. Zon[1], A.P. McCaffrey[1], A. Buck[2], R.I. Hogrefe[1].* 1) Research and Development, TriLink BioTechnologies, San Diego, CA; 2) Infection and Evolution, University of Edinburgh, Edinburgh, UK.

Next-generation sequencing (NGS) is revolutionizing nucleic acid analysis. To obtain high quality sequencing data it is vital to minimize undesired products such as adapter dimers. These can occupy valuable flow cell space and result in unproductive reads. Many biological samples contain limited genetic material, thus it is important to use a library preparation method that yields highly specific final products from low input amounts. One of the more challenging sample types to prepare for sequencing is small RNA (sRNA). sRNAs are non-coding short regulatory RNA sequences, such as miRNA. Due to the inherently short lengths, a sRNA library is difficult to size select away from an empty library (known as adapter dimer). Therefore we developed a technology specifically targeted to reduce adapter dimer formation during small RNA library preparation using chemically modified CleanTag™ adapters. We found that without adapter dimer formation during the NGS sample preparation, small RNA libraries can be purified by an automatable bead-based size selection process eliminating the need for gel purification. In addition, reduction of adapter dimer formation allows for successful NGS sample prep with significantly lower total RNA inputs. For example with CleanTag™ modified adapters, gel purified libraries generated from 10 and 1 nanogram (ng) total human brain RNA input libraries resulted in 86% of total reads mapped to miRNA and less than 1% of total reads attributed to adapter dimer. At 10 ng input gel purified libraries from a commercially available kit resulted in only 40% of total reads mapped to miRNA while 50% of total reads attributed to adapter dimer. While commercial kits do not produce usable data at 1 ng, replacing gel purification with magnetic bead purification at 1 ng input, CleanTag™ libraries resulted in sufficient miRNA reads and less than 10% adapter dimer. We have successfully sequenced low RNA input biological samples such as extracellular vesicles, plasma, urine, FACS sorted and immunoprecipitated samples that have previously proven difficult to sequence. Sequencing results at 10 picogram total human brain RNA input show the potential to sequence small RNA at the single cell level using modified adapters. These results demonstrate the necessity of adapter dimer suppression and utility of CleanTag™ technology for both low input and high throughput library preparation. Currently, we are adapting CleanTag™ for use in long RNA sequencing.

3210F

Monitoring guide RNA synthesis for CRISPR/Cas9 genome editing workflow using the Agilent 2100 Bioanalyzer System. *M. Liu[1], A. Padmanaban[2], I. Pechtl[3].* 1) Agilent Technologies, Inc., La Jolla, CA; 2) Agilent Technologies, Bangalore, India; 3) Agilent Technologies, Waldbronn, Germany.

Bacterial clustered regularly interspaced short palindromic repeats (CRISPR) - associated protein 9 (Cas9) system has increased in popularity as a genome editing tool for targeted mutations, insertions, deletions and gene knockout studies. CRISPR genome editing has also proved superior to Zinc-finger nucleases (ZFN) and transcription activator- like effector nucleases (TALENs) due to its simplicity and easy programmability. In CRISPR, a guide RNA (gRNA) is used to recognize and introduce a double standard break (DSB) in a target DNA. The DNA repair mechanism triggered after the break is then exploited to introduce an insertion/deletion (indel) in the case of non-homologous end joining (NHEJ), or precise genetic modification if a homology-directed repair (HDR) pathway is triggered.A critical part of the CRISPR/Cas9 tool is the design and synthesis of the gRNA that comprises T7 promoter sequence, target sequence, and protospacer adjacent motifs (PAM). Monitoring the transcription of the gRNA is critical to the workflow to ensure successful gene editing.Here we present an automated electrophoresis approach for monitoring the synthesis of gRNA. The study involves design and synthesis of gRNA against two candidate genes by PCR amplification. The resulting amplicon is expected to contain all necessary components in gRNA to induce DSB. The synthesized gRNA is then purified to remove contaminating DNA oligos and checked for size and purity using the electrophoresis system.

3211W

Automation of genome gel-bead and library preparation kit for 10X Genomics Chromium Controller workflow. *M. Benway[1], P. Kocjan[1], A. Price[2], B. Gerwe[1].* 1) PerkinElmer, Inc. Life Sciences & Technology, Hopkinton, MA. USA; 2) 10X Genomics, Inc., Pleasanton, CA. USA.

The field of genomics is shifting towards longer-read sequencing than those typically obtained by current next-generation sequencing (NGS) technologies, in order to capture the full complexity of genomes. Current short-read sequencing technologies keep costs relatively low but the short-reads miss critical information including phasing, structural variants and the ability to map highly repetitive regions. Without this information, only a partial structure of the genome is realized with many mutations and variants being missed.   10x Genomics' Chromium™ platform is able to preserve this information over distances >150kb by linking the short-reads to a larger DNA fragment by use of a barcode. In this process, ~1 ng of DNA is partitioned into millions of nanoliter droplets via a microfluidic chip, each containing a unique barcode introduced by a co-partitioned gel bead. The generated libraries are currently compatible with all short-read sequencers, enabling this hidden information to be accessed without the need to invest in a new sequencing infrastructure. The use of finely-tuned liquid-handling systems becomes advantageous and ensures optimal performance when manipulating the novel 10x Genomics gel bead and partitioning oil reagents, as well as, preserving the integrity of high-molecular weight input DNA during routine handling.   Here, we describe and demonstrate how the uniquely flexible capabilities of the PerkinElmer Sciclone® NGSx Workstation interface with the 10x Genomics chip to successfully handle these specific requirements. We show the preservation of intact genomic DNA during automated library preparation and demonstrate that the libraries made using the Sciclone are of comparable quality to manual preparation. Automation is critical to the reproducible scale-up of high-throughput projects by removing manual variability, increasing efficiency, and enabling LIMs tracking. Combining the Sciclone with 10x Genomics' unique partitioning technology has the potential to unlock the promise of NGS in these large-scale studies.

3210F

3211W

3212T

Automated, high-throughput NGS library preparation from low input FFPE and cell-free DNA sample material. *B. Gerwe[1], V. Vanessa Kelchner[2], L. Cook[3], C. Fronick[3], V. Magrini[3], M. Benway[1], J. Laliberte[2], S. Sandhu[2], C. Schumacher[2], L. Kurihara[2], V. Makarov[2], T. Harkins[2], R. Fulton[3].* 1) Perkin-Elmer, Inc. Life Sciences & Technology, Hopkinton, MA; 2) Swift Biosciences, Inc., Ann Arbor, MI; 3) McDonnell Genome Institute at Washington University School of Medicine, St. Louis, MO.

While technical advances in Next-Generation Sequencing (NGS) have facilitated lower input sample volumes and material, the challenges of library bias and library complexity affecting data quality remain. Accel-NGS® 2S DNA Library kits (Swift Biosciences) have been designed to overcome these challenges providing uniform and more complete coverage, resulting in improved sequencing efficiency and reduced costs. Additional cost and time savings can be found in automating library preparation, leading to increased throughput while maintaining high-quality data and reproducibility with little to no variation across wells. In this study, we describe and demonstrate high-throughput automated NGS library preparation for targeted capture and WGS from 10ng FFPE and 10ng cfDNA on the PerkinElmer Sciclone® NGSx with Accel-NGS 2S DNA Library kits. An Alu 247/115 repeat assay was used to assess contamination of high molecular weight DNA, quantity and integrity of cfDNA and FFPE prior to library preparation. Reproducibility and quality were determined by qPCR, LabChip® GX analysis and subsequent sequencing on the Illumina HiSeq X10. As DNA extracted from FFPE exhibits varying degrees of damage, the Alu repeat assay is an accurate way to assess and quantify usable DNA. Our results show a DNA integrity score (ratio of ng/µl Alu 247/115) of ≥0.4, improving as fixation time decreases. For cfDNA, the expected integrity score of ≥0.29 was achieved. DNA quantity was sufficient for 10ng input into the prep. Library yields were sufficient for sequencing, and LabChip analysis validated the targeted insert size of 350bp for FFPE and 165bp for cfDNA. Sequencing of manual versus automated library preparations yielded similar results: ultra-low duplication, no adapter dimer formation, median insert size in-line with the LabChip results and library complexity as expected for 10ng human gDNA. Accel-NGS 2S DNA Library kits on the Sciclone NGSx provide sequence-ready libraries, with equal or better reproducibility and quality to manually prepared libraries. Highly efficient library preparation driven by end repair of both 3' and 5' DNA termini delivers a more complex library requiring less sequencing, enabling comprehensive analysis of low input DNA samples such as FFPE and cfDNA. Significant cost savings and improvement in lab efficiencies were achieved combining the use of PerkinElmer automation for higher throughput library prep with Swift's highly efficient library preparation.

3213F

Single molecule targeted sequencing for cancer gene mutation detection. *J. He[1,2].* 1) Department of Biology, South University of Science and Technology of China, Shenzhen, Guangdong, China; 2) Direct Genomics Ltd. Shenzhen, Guangdong, China.

With the rapid decline in cost of sequencing, it is now affordable to examine multiple genes in a single disease-targeted clinical test using next generation sequencing. Current targeted sequencing methods require a separate step of targeted capture enrichment during sample preparation before sequencing. Although there are fast sample preparation methods available in market, the library preparation process is still relatively complicated for physicians to use routinely. Here, we introduced an amplification-free Single Molecule Targeted Sequencing (SMTS) technology, which combined targeted capture and sequencing in one step. We demonstrated that this technology can detect low-frequency mutations using artificially synthesized DNA sample. SMTS has several potential advantages, including simple sample preparation thus no biases and errors are introduced by PCR reaction. SMTS has the potential to be an easy and quick sequencing technology for clinical diagnosis such as cancer gene mutation detection, infectious disease detection, inherited condition screening and noninvasive prenatal diagnosis.

3214W

Improvements to chromatin immunoprecipitation applications through more robust library preparation methods. *A. Blattler[1], K. Cunningham[2], K. Chen[1], L. Kurihara[2], B. Egan[1], V. Makarov[2], T. Harkins[2], T. Kelly[1].* 1) Active Motif, Inc., Carlsbad, CA, USA; 2) Swift Biosciences, Inc., Ann Arbor, MI, USA.

Chromatin immunoprecipitation followed by Next Generation Sequencing (ChIP-seq) has allowed for the genome wide identification of transcription factor binding sites and histone modifications, providing insight into key mechanisms of gene regulation in human disease. One limitation in ChIP-seq is the requirement for millions of cells, making it difficult to perform the assay in primary cells and tissues. Recent advancements by Active Motif enabled high-quality ChIP from lower cell numbers and from difficult formalin-fixed, paraffin-embedded (FFPE) tissues. However, due to low-abundance or damaged DNA, standard library preparation methods limited the ability for these methods to be applied genome-wide. The Swift Accel-NGS 2S Plus DNA Library Kit dramatically improved the sequencing of these samples due to high efficiency library conversion at low input. By combining both 5' and 3' end repair with optimized ligation that requires no adapter titration at low input, libraries representing the target complexity of these samples could be generated. Here we present data demonstrating how the combination of Swift library technology and Active Motif's ChIP technology has enabled high quality sequencing of several ChIP-based applications. CTCF ChIP-Seq using 2S Plus from 50,000 cells resulted in peaks comparable to ENCODE data for CTCF from 20,000,000 cells. Additionally, Active Motif and Swift have enabled ChIP-seq from as little as 1,000 cells for histone modifications and increased resolution from FFPE ChIP-sequencing experiments.

3215T

CHDbase: A genomic variation data warehouse for congenital heart disease. *J. Fitch[1], D. Gordon[1], S. Fitzgerald-Butt[2], V. Garg[2], K. McBride[2], P. White[1].* 1) The Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 2) Center for Cardiovascular Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH.

In a simple case, the genetic cause of a disease can be determined by sequencing the DNA from an affected individual, comparing that sequence to a known reference, and then looking through the locations where the sequences differ for a variation that is known to cause the phenotype in question. However, since an average person has 4 - 5 million sites where their genome differs from the reference, it is impractical to manually evaluate them all. Many different techniques have been developed to automate the reduction of this list by assuming that some characteristics of a variant make it unlikely to cause a rare disease, such as a high frequency in non-diseased populations, being a synonymous change, or occurring in a region of the genome that is not highly conserved. When the straightforward approaches to finding causative genetic mutations fail, a larger cohort study can be performed that attempts to find association between variants that are found in many examples of diseased genomes while not being found in control genomes. One of the main challenges with this approach is that as cohort sizes are increased, to increase the statistical power of the study, computational complexity is increased as well. Further exacerbating the computational challenge is the realization that some of the simplifying assumptions are not valid. For example, causative variants may be non-coding, or caused by rare combinations of more common variants. These factors drive the requirement for a Big Data approach. Hadoop is a popular Big Data distributed storage and computation framework that has been widely used in many different data science fields. The goal of this project is to leverage this technology to facilitate the storage and analysis of genomic variation from large cohort studies in order to find causative mutations. To prove the effectiveness of this approach, we have loaded variant data for 800 whole exome and whole genome sequenced samples in a study to identify the genetic etiologies of congenital heart disease (CHD). The variant warehouse design allows integration of multiple data sources, including variant frequency, damaging prediction criteria, and phenotypic data. Writing applications to process CHD variant data stored in this variant warehouse provides a mechanism to rapidly perform complex data processing steps with a great deal of flexibility and enabling the development of the oligogenic functionality required to look beyond monogenic disease etiologies.

3216F

Accelerating discovery in the undiagnosed: Nationwide Children's Hospital's Research Genomics Strategic Initiative. *B. Kelly[1], T. Mihalic Mosher[1], D. Corsmeier[1], H. Kuck[1], J. Fitch[1], K.L. McBride[2], G.E. Herman[1], P. White[1].* 1) The Institute for Genomic Medicine at Nationwide Children's Hospital, Columbus, OH; 2) The Center for Cardiovascular Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH.

A key component of Nationwide Children's Hospital's strategic plan is to integrate genomics into the everyday practice of pediatrics. Following our participation in the CLARITY Undiagnosed Challenge, we set out to create an interdisciplinary research program to identify new disease genes for pediatric disorders through exome and whole genome sequencing, translational bioinformatics and functional genomics. CLARITY Undiagnosed, launched by Boston Children's and Harvard Medical School, was markedly more difficult than the original CLARITY challenge. Whole genome sequencing data was provided for families for whom extensive clinical and genetic testing had been previously performed with few to no answers. Competitors were asked to provide reports on their findings that could be interpreted by researchers, medical professionals and families. The key to our approach was the diversity of the team that we assembled, with clinical geneticists, genomic researchers, genetic counselors, clinicians with phenotypic expertise and bioinformaticians to review clinical results, assign phenotypes, discuss analyses and agree on results to report. Our team was specifically recognized for its ability to combine research and clinical expertise, "setting a high bar for clinical genomics." Building upon this success, we have established a new Research Genomics Program at Nationwide Children's Hospital with two main goals: 1) to educate clinical faculty on the use of genomics approaches in patient care, treatment, and precision medicine and 2) to discover novel, disease-causing variations in targeted clinical populations. To achieve these goals, we formed a Steering Committee to solicit cases from clinicians within the hospital and a Genomic Review Board to decide case qualification and play an active role in the analysis and review of cases. We are utilizing advanced techniques for sequencing, analysis and annotation, and applying a variety of computational approaches to prioritize candidate genes. To date, we have 20 rare disease families identified and/or enrolled, representing 7 hospital sections (Genetics, Rheumatology, Neonatology, Neurology, Hematology/Oncology, Gastroenterology and Endocrinology). Ten of these families have been consented, sequenced and analyzed with three pathogenic variants discovered in known genes, demonstrating the power of our approach for genomic diagnosis while leading to the discovery of new disease genes. An update on our efforts will be presented.

3217W

Variation in accuracy of genetically determined sex from genome-wide genotyping chip arrays according to low frequency SNP content. *S. Provost[1,2], G. Asselin[1,2], L-P. Lemieux Perreault[1,2], D. Valois[1,2], I. Fillion[1,2], V. Normand[1,2], L. Mailloux[1,2], I. Mongrain[1,2], J-C. Tardif[1,2,3], M-P. Dubé[1,2,3].* 1) Montreal Heart Institute, Montreal, Canada; 2) Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal, Canada; 3) Université de Montréal, Faculty of Medicine, Montreal, Canada.

At the Université de Montréal Beaulieu-Saucier Pharmacogenomic Centre, we have genotyped large numbers of samples from clinical and observational studies. As part of our regular procedures, we perform genomic quality control verification including the concordance between clinically reported sex and genetically determined sex. In the present study, we assessed genetically determined sex using: i) the PLINK F statistics, ii) heterozygosity of the X chromosome, iii) the number of genotype calls on the Y chromosome, and iv) by comparing mean intensities of X and Y chromosome probes using the log R ratio (LRR) and B allele frequency graphics (BAF). Genotype data was generated for over 44,000 DNA samples from 8 randomized clinical studies and 3 large observational studies tested with different Illumina genome-wide BeadChip arrays (Human610quad, ExomeChip, Omni2.5, Omni2.5-exome, HumanCore, and MEGA). Analyses were performed using the pyGenClean data clean up pipeline. We found that the rate of X-chromosome heterozygosity in females varied according to the array's low frequency SNPs content (minor allele frequencies <0.05) and that it had an impact on the performance of metrics to assess the concordance between clinically reported sex and genetically determined sex. In particular, X-chromosome heterozygosity was 30% for arrays with lower low frequency SNP content (610quad, HumanCore), 20% for arrays with intermediate low frequency SNP content (Omni2.5), 10% for arrays with high low frequency SNP content (Mega) and finally, 4% for chips with very high low frequency SNP content (ExomeChip). In order to optimize accuracy of genetically determined sex by GenomeStudio and Beeline, we report on array-specific expected parameters for quality metrics for the determination of sex. Across the different studies, sex discrepancies ranged from 0.12% to 0.97% and originated from different sources. Sex chromosome abnormalities represent a minority of the sex mismatches, and we report on the rates of putative XXX, XXY, partial loss of heterozygosity (LOH) and XO/XX mosaics detected based on to the tested quality metrics. This study highlights the importance of genomic quality control metrics calibration with evolving genotyping chip content.

3218T

Detection and depletion of bacteria contamination in saliva derived DNA samples for human whole genome sequencing. *M. Costello, N. Lennon, S. Gabriel, Broad Data Sciences and Data Engineering Group.* Genomics Platform, Broad Institute, Cambridge, MA.

Obtaining saliva from patients for downstream genomic analysis has become increasingly popular as collection kits are easy to use and are much less invasive than blood draws or biopsies. However, one major drawback of saliva samples is the presence of non-human DNA, mainly due to oral bacterial contamination. In reviewing data from the 2500 saliva samples processed over the past year through our whole genome sequencing (WGS) pipeline, the average non-human contamination in saliva derived WGS samples is about 10% but can range anywhere from just 1% to greater than 70%. This non-human contamination makes achieving the desired human genome coverage quite difficult, incurring extra sequencing costs and often requiring additional filtering steps during data analysis. Here, we present our two pronged approach for dealing with bacterial DNA contamination in human samples. First, we are developing a specific and sensitive QPCR assay to detect and quantify the amount of bacterial contamination in saliva derived DNA samples. Second, we are also developing methods to deplete bacteria DNA from human samples which allow separation of prokaryotic DNA from the CpG-methylated eukaryotic human DNA. Implementation of these methods will allow us to assess % contamination in a high throughput manner and provide methods for enriching for human DNA prior to library preparation and sequencing. Combined, these new methods will allow us to separate out samples with extremely high contamination, move forward with those with low contamination, and rescue samples with moderate contamination via the depletion protocol. Implementation of these processes will reduce sequencing costs, improve usable data yield, and simplify downstream analysis for saliva derived WGS samples.

3219F

Automated, low-cost and low-input RNA-Seq and DNAseq library preps. *S. Vaezeslami[1], S. Srinivasan[3], P. Malihi[2], P. De Hoff[3], T. Herrero[3], T. Cuong[3], J. Jenkins[1], L. Laurent[3], J. Hicks[2], P. Kuhn[2].* 1) TTP Labtech, Cambridge, Ma; 2) University of Southern California, Los Angeles, CA; 3) University of California at San Diego, San Diego, CA.

Significant decrease in sequencing costs, having access to genomics automation platforms, and recent increase of both research and clinical sequencing applications have led to a larger demand for high-throughput, low-cost and low-sample input NGS sample preps. Using low volume liquid handlers, with low dead volume and ability to work with difficult and viscose solutions at sub-microliter volumes offer many advantages in NGS sample prep processes, such as library construction, quantification and normalization/pooling. These advantages include lower cost, higher throughput, better reproducibility, less hands-on time, less sample input, and improved process control. In this poster we will discuss two applications in which the sample prep volumes were reduced to sub microliter volumes in order to significantly lower the cost and amount of sample input. Moreover we discuss a 384-well low-volume Ampure bead clean up performed in less than 1 hour. mosquito liquid handler, based on true-positive displacement technology, has been instrumental in these workflows to provide an accurate, precise and reliable walk-up solution. In the first example the volumes of NEBNext® Ultra™ II DNA Library Prep kit (E7645S) were reduced by 10 times and the amount of sample inputs were reduced to 5, 37 and 60 ng in 6.5 µL total volume, from 185 ng sample input in manual reaction set up at 65 µL. In the second example the reaction volumes and sample inputs of NEBNext® Small RNA Library Prep kit for Illumina® (E7580) were reduced by more than 5 times (down to 1 ng of input RNA). The results of miniaturized volumes sample preps were compared to those of manual set up at the same or larger volumes. Sequencing data show that the results are not affected by lowering the sample input or volumes. The most important factors are having an accurate and precise liquid handler to avoid sample to sample variation when miniaturizing the volumes and sample inputs. Finally, we discuss a magnetic bead clean up protocol in less than 5 uL total volume and in 384 well plate format. The challenges of low volume bead clean up in 384 format are the large tip dimensions and high dead volume of most large volume liquid handlers. By using mosquito HV liquid handler and doing the clean up in less than 5 µL the required sample input and the amount of magnetic beads used were dramatically reduced, which in turn will save on the cost, especially for high throughput applications.

3220W

Ligation-free small RNA sequencing libraries from clinical samples using SMART® technology. *A. Farmer[1], N. Bolduc[1], S. Lee[1], K. Tanriverdi[2], E. Malahias[2], E. Mikhalev[2], J. Freedman[2].* 1) Takara Bio USA, Inc., 1290 Terra Bella Ave, Mountain View, CA 94043, USA; 2) University of Massachusetts Medical School, Department of Medicine, Division of Cardiovascular Medicine, Worcester, MA 01605, USA.

Small non-coding RNAs regulate gene expression via diverse mechanisms and facilitate fundamental cellular processes such as transcript splicing and protein translation. Moreover, small RNAs may act at sites different from their original production via incorporation and transport in bodily fluids. Obtaining an accurate portrait of small RNA expression levels from small sample inputs carries potential both for the fulfillment of basic research objectives and the development of novel therapeutics and clinical diagnostic solutions. Towards this end, we have developed a novel, ligation-free approach for the preparation of small RNA sequencing libraries that leverages 3′ RNA polyadenylation followed by cDNA synthesis and template switching. This approach minimizes sample representation bias and is sensitive enough to accommodate inputs of as little as 1 ng of total RNA. Here we present data demonstrating the accuracy, sensitivity, and reproducibility afforded by the SMARTer® smRNA-Seq Kit for Illumina®. Using a library of synthetic miRNAs, we show that the SMARTer method generates a more accurate representation than the typical adapter ligation method. Furthermore, we successfully created high-quality libraries using 400 pg to 1 ng of RNA extracted from plasma and serum samples, demonstrating the suitability of the SMARTer approach for applications such as biomarker discovery.

3221T

Improved expression profiling and cellular localization of circRNA molecules in human samples. *A. Zaghlool[1], A. Ameur[1], C. Wu[2], J.O. Westholm[1], M. Manivannan[3], K. Bramlett[3], M. Nilsson[2], L. Feuk[1].* 1) Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Sweden; 2) Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; 3) Clinical Sequencing Division, Life Science Solutions Group, Thermo Fisher Scientific, San Francisco, CA, USA.

There is strong evidence that circular RNAs (circRNAs) play an important role in gene regulation and may contribute to the development of various human disorders. However, the global properties of circRNAs such as accurate estimation of expression levels and cellular localization remain elusive. Here, we developed an Ion AmpliSeq™ target-enrichment sequencing panel to provide an accurate estimation of circRNAs expression levels and to increase the sensitivity for the detection of circRNAs expressed at low levels. Moreover, we adapted the previously published PadLock technology to visualize circRNA in situ at subcellular levels. To evaluate the circRNA panel, we performed sequencing on total, cytoplasmic and nuclear RNA extracted from human cell lines, brain, liver, heart, placenta and blood (total number of samples=30) using the Ion Torrent™ system. Our panel included 130 circRNA and 130 linear targets from the same genes and 15 negative controls (coding sequences lacking annotated circRNAs). The circRNA sequencing produced around 8 million reads per sample, of which 97% where mapped to the targets in the panel. CircRNA expression showed high correlation between technical replicates (R=0.98), highlighting the reproducibility of the panel. In agreement with previously published data, we show that the majority of circRNAs are located in the cytoplasm, and that circRNA expression levels vary between tissue types. Our results also show that circRNAs are detectable in human blood and we will further investigate other types of samples including serum and FFPE material. Our panel increased the sequence coverage for circRNA targets by several orders of magnitude as compared to expression levels obtained from total RNA sequencing, thereby providing increased sensitivity for detecting circRNAs expressed at low levels. Moreover, the unique design of the panel allows an accurate estimation of expression ratios between circRNA and their linear counterparts. These ratios were validated using qrtPCR. CircRNAs visualization using FISH has been challenging due to their short sequence length. By adapting the PadLock method on fixed SHSY-5y cells we also demonstrate an efficient visualization of circRNA localization in situ on a subcellular level. In summary, we describe a novel, cost effective and scalable sequencing panel to accurately estimate circRNA expression from various input materials, as wells as an efficient method to determine their subcellular localization.

3222F

Measuring warfarin resistance of every *VKOR* variant to guide dosing.
M.A. Chiasson[1], A.E. Rettie[2], D.M. Fowler[1]. 1) Genome Sciences, University of Washington, Seattle, WA; 2) Medicinal Chemistry, University of Washington, Seattle, WA.

Vitamin K epoxide reductase (VKOR) is the target of warfarin, a commonly prescribed anticoagulant with an extremely narrow therapeutic window. Warfarin-related hemorrhages result in an estimated 21,010 hospitalizations yearly. Approximately 30% of the variation in warfarin dose is driven by sequence variation in *VKOR*. Thus, knowing a patient's *VKOR* sequence before initiating treatment could allow for more precise dosing if we knew its warfarin resistance phenotype. However, only 28 of the 54 known missense variants in *VKOR* have been characterized for warfarin resistance. These 54 are only the tip of the iceberg: as more people are sequenced, additional variants will appear. To address this issue, I am measuring the warfarin resistance of all 3,260 VKOR missense variants using deep mutational scanning, a method that combines high throughput mutagenesis with deep sequencing. I created a reporter cell line with the endogenous *VKOR* gene knocked out and am engineering a library of these cells to contain all VKOR missense variants. I will use a flow cytometric assay to sort variants into bins according to VKOR activity across a range of warfarin doses. Deep sequencing of each bin will enable me to calculate an IC_{50} value for every variant. I will validate IC_{50} values for a subset of variants using traditional assays and clinical data. These IC_{50} data will comprise a *VKOR* sequence-resistance map that reveals how each singly mutated variant we could observe in a patient affects warfarin resistance. My ultimate goal is to develop a *VKOR* sequence-based dosing algorithm to help clinicians start patients on the appropriate dose of warfarin.

3223W

MinION nanopore sequencing of human leukocyte antigen-B region.
K.N.T. Ton, S. Cree, A.L. Miller, M.A. Kennedy. Department of Pathology, University of Otago, Christchurch, Canterbury, New Zealand.

The human leukocyte antigen (HLA) system is a gene family that encodes the human major histocompatibility complex (MHC). The HLA complex plays a key role in the human immune system by presenting peptide antigens to T-cells. *HLA-B* is the most polymorphic genes in MHC class I, leading to over 4000 *HLA-B* alleles (IPD-IMGT/HLA Database Release 3.24). Many *HLA-B* alleles have been associated with adverse drug reactions and disease risks. *HLA-B* alleles are routinely identified by PCR-sequence-based typing (PCR-SBT) but the use of Sanger sequence data and genotypes can lead to difficulties resolving allele ambiguities. MinION is a nanopore sequencing device developed by Oxford Nanopore Technologies (Oxford, UK). This pocket-sized device allows generation of long reads on single molecules. A fragment of 1,704 bp and a sub-region of 943 bp covering *HLA-B* exon 2 and 3 from six DNA samples were amplified by nested-PCR. Both PCR amplicons of each sample were tagged with sequencing adapters and two different barcodes. All 12 amplicons were pooled and purified using Genomic DNA Sequencing kit (ONT). The library was loaded into the R7.3 flow cell and run for 48 hours. Raw sequence data was uploaded for base-calling using Metrichor software (ONT) and followed marginAlign pipeline (Jain et al., 2015) for mapping and variant calling. Data were compared with sequences obtained from PCR-SBT method. Using the MinION, we have sequenced *HLA-B* gene for 6 samples with two different read lengths. A total of 1,090 reads were parsed using uniquely identifying barcode, of which 100% were correctly aligned to the reference sequence. The mean of read length was 1,357 bases and average of read depth was 80X. There was a high consistency of variant calls when comparing within each sample pairs indicating there was no difference in sequencing quality between two amplicon sizes. Comparing with data from Sanger sequencing, we found that the discordant call rate was high, but good quality data could still be discriminated. Our conclusion is that the MinION is capable of sequencing *HLA-B* amplicons of multiple barcoded samples, and variants can be detected. However, the fact that MinION reads from the R7.3 chemistry are relatively error-prone, combined with HLA complexity, make accurate interpretation of haplotypes difficult. We will shortly repeat these analyses with the new MinION R9 flow cells which should lead to increased accuracy and read depths.

3224T

A highly sensitive cDNA kit for single cell or low input RNA-seq workflow. *Y. Bei, N. Guan, T. Shtatland, B.W. Langhorst, E. Yigit, N.M. Nichols.* New England Biolabs, Ipswich, MA.

RNA sequencing (RNA-seq) based transcriptome studies have become an important tool for gene expression profiling from diverse tissues, cell types, developmental stages and diseases. Until recently, most of the RNA-seq studies have required relatively high sample inputs, which can limit the study of low abundance samples such as relied on input from bulk samples, which can pose challenges for the study of low abundance samples, such as stem cells, circulating tumor cells (CTCs) (CTCs) or other scarce clinical samples. More recently, single cell RNA-seq has emerged as a powerful approach to characterize the gene expression heterogeneity of gene expression within phenotypically identical cell populations. Findings from such single cell RNA-seqthese studies have important implications in both basic research and clinical settings, including the study of tumor microevolution to develop potential targeted therapies.   To address support the growing needs in the area of single-cell transcriptomics research, we have developed a simple and robust single cell or low input (10 pg to 200 ng) RNA-seq workflow in which full- length cDNAs are first generated directly from cell lysis in a single tube**.** that transcribes and amplifies full length cDNAs in a single tube from either single cells or low input total RNA ranging from 10pg to 200ng. The resulting cDNA libararies are subsequently converted into sequencing-ready Illumina libraries using the NEBNext® Ultra™ II DNA Library Prep Kit, or alternatively can be converted into libraries for other major NGS platforms. can be easily further converted into NGS libraries that are compatible with any major sequencing platforms. In this work,Using this method and other commercially available alternatives, we present collected RNA-seq data from single HEK 293 cells and low input Universal Human Reference RNA (UHR) with an ERCC reference RNA spike-in., Compared to other methods, we observed using our cDNA Kit in conjunction with NEBNext® Ultra™ II DNA Library Prep Kit. For comparison, we also prepared libraries using SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing from Clontech. hHigher library yields and a reduction in bias against transcript 5' ends with our low-input cDNA synthesis approach. were observed in samples prepared with our method described in this work. In addition, In addition, our workflow identified significantly more transcripts from both samples examined in this study.

3225F

Post RNA-seq analysis GUI. *J. Chung[1,2], G. Jun[1,8], L.A. Farrer[1,2,3,4,5,6,7].* 1) Biomedical Genetics, Boston University Medical Campus, Boston, MA; 2) Bioinformatics Graduate Program, Boston University, Boston, MA; 3) Department of Medicine, Boston University Medical Campus, Boston, MA; 4) Department of Biostatistics, Boston University Medical Campus, Boston, MA; 5) Department of Neurology, Boston University Medical Campus, Boston, MA; 6) Department of Ophthalmology, Boston University Medical Campus, Boston, MA; 7) Department of Epidemiology, Boston University Medical Campus, Boston, MA; 8) Eisai Inc, Andover Product Creation Innovation System, Andover, MA.

Most current RNA-Seq data analysis tools were developed primarily for the purposes of differential expression gene (DEG) analysis and quality control (read mapping). We developed an interactive software program, *Post-RNA-Seq graphical user interface (GUI),* which facilitates exploration of the association of genetic variants with expression of individual transcripts and corresponding amino acids in protein structures. *Post-RNA-Seq GUI* software is designed to integrate 1) isoform-specific expression levels resulting from the RNA-Seq analysis tool, RSEM (*Li B et. al.* 2011) or MISOPY (*Katz Y et. al.* 2010); 2) genotype and DNA sequence data; and 3) protein structure data. Users can explore specific transcripts (or exons) of their interest in published genetic association studies (e.g. quickly determine whether any genetic loci been detected near the transcripts) and compare isoform expression levels between cases and controls and detect significant cis-acting eQTLs for the transcripts. Finally, users can view the amino acids encoded by specific variants in 3-dimensional (3D) protein structures of the transcripts. In summary, *Post-RNA-Seq GUI* allows exploration of diverse types of data resources to search for other important information related to RNA-Seq findings using a graphical interface. This software saves time and effort which might otherwise be expended on manually exploring many datasets.

3226W

RNASeq 2.0 @ Broad Genomics: Advances in RNA sequencing analysis.
C. Cibulskis, M. Shand, K. Bergin, A. McBeth, G. Vicente, Y. Farjoun, K. Tib-betts, N. Lennon. Broad Institute of MIT and Harvard, Cambridge, MA.

   RNA sequencing has the potential for broad reaching applications in clinical research and clinical diagnostics[1]. Approaches to the generation and analysis of RNA sequencing data has numerous potential endpoints including, but not limited to, the detection of variants, gene fusions, alternative splicing, and expression profiling. The ability to generate high quality RNA sequencing data and perform sensitive and efficient analyses is critical to realizing the full value of transcriptome data in clinical research and clinical practice. Broad Genomics and our collaborating research teams have a long history of generating and analyzing RNA sequencing data for both cancer[3,4] and germline applications[2]. Analytic tools currently available for RNA sequencing analysis are many and varied in terms of application, practicality and efficiency. To better enable our research teams to gain optimal utility from RNA based studies, we complete-ly reassessed the data and analyses that we are delivering. We then built and launched an updated RNASeq 2.0 pipeline that includes optimized read alignment, fusion, expression, and variant detection. Broad RNASeq 2.0 will be made available in our cloud-based analysis workbench (FireCloud) and will facilitate scalable and rapid analysis runs, RNA/DNA analysis integration and ease of data sharing. We will describe the details and merits of Broad RNASeq 2.0 as well as other important improvements in our laboratory processing.   1: Byron SA, Et al. Translating RNA sequencing into clinical diagnostics: oppor-tunities and challenges. Nat Rev Genet. 2016 May;17(5):257-71. doi:10.1038/nrg.2016.10. 2: GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May 8;348(6235):648-60. doi: 10.1126/science.1262110. 3: Van Allen EM, Et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med. 2014 Jun;20(6):682-8. doi: 10.1038/nm.3559. 4: Hacohen N, Et al. Getting personal with neoantigen-based therapeutic cancer vaccines. Cancer Immunol Res. 2013 Jul;1(1):11-5. doi: 10.1158/2326-6066.CIR-13-0022.

3227T

Sensitive capture of full-length transcript information with targeted RNA-seq. *S. Gandlur, R.N. Fish, S. Zhang, Y. Gindin, S. Taylor, A. Farmer.* Takara Bio USA, Mountain View, CA.

   While recent advances in technology have greatly improved research in whole transcriptome RNA-seq, several challenges still remain, stemming from the complexity inherent in such large-scale sequencing. The large dynamic range of the transcriptome often means that a few highly abundant transcripts account for the majority of sequencing reads while less-abundant transcripts (representing a majority of RNAs) account for only a small percentage of se-quencing reads. Targeted RNA-seq aims to overcome this problem by improv-ing sequence coverage of transcripts of interest that may be present in low amounts, saving costs, and simplifying analysis. Sensitive targeted enrichment of RNA enables the capture of information about transcripts that would other-wise be missed or would require a much greater number of sequencing reads to be detected, including chimeric gene fusions, transcript isoforms, and splice variants. We combined a streamlined method for direct capture of full-length transcripts from total RNA with the cDNA synthesis technology of the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing to develop a target-specific protocol with high sensitivity and low background. The resulting RNA capture method demonstrates consistent enrichment and coverage from 1 ng to 1 µg of a variety of total RNA inputs. By capturing full-length transcripts, the protocol enables the detection of structural variation in expressed RNAs. For example, we identified a gene fusion event present at a frequency lower than 0.5% in a sequencing library of fewer than 2 million reads while targeting only one partner of the gene fusion. Additionally, we were able to maintain relative expression levels for targeted genes post-enrichment, providing confidence in differential expression analysis of transcripts of interest. Furthermore, the substantial increase in coverage of only a subset of genes of interest results in a lower sequencing depth required, thereby reducing analysis time and experimental costs.

3228F

Parsing nature's modification strategies in defining new therapeutic targets for Marfan syndrome. *J.P. Habashi[1], E.G. MacFarland[2], N. Huso[2], R. Bagirzadeh[2], Y. Chen[2], D. Bedja[2], H.C. Dietz[2,3].* 1) Dept Pediatric Cardiology, Johns Hopkins Univ, Baltimore, MD; 2) Dept Genetics, Johns Hopkins Univ, Baltimore, MD; 3) Howard Hughes Medical Institute, Bethesda, MD.

Marfan syndrome (MFS) is an autosomal dominant disorder caused by mutations in FBN1. The leading cause of mortality is aortic root dilation and dissection. Women with MFS show an enhanced risk of accelerated aortic dilatation or tear late in pregnancy or in the early postpartum period. We previously identified oxytocin as a critical mediator in mice homozygous for a hypomorphic Fbn1 allele (mgR/mgR). 91% of mgR/mgR females die from aortic dissection in the immediate postpartum period; prevention of lactation or treatment with a highly specific oxytocin receptor antagonist diminished ascending aortic growth rate (0.54 ± 0.43 and 0.44 ± 0.49 vs. 1.02 ± 0.61mm/7weeks in placebo treated mice, respectively) and resulted in a dramatic decrease in death from dissection (26% and 7%, respectively). Curiously, the deleterious gene-by-environment interaction imposed by pregnancy closely mimicked that observed upon exposure of MFS mouse models to calcium channel blockers (CCBs); both showed acceleration of growth and a unique predisposition for tear in the distal ascending aorta (AscAo) – an aortic segment typically spared in both people and mice with MFS. Using a genetic modifier screen and a candidate-based approach, we previously showed that CCB-induced aortic disease in MFS relates to a PLC/IP3/MAP3K4/MEK/ERK axis of activation in aortic vascular smooth muscle cells. Based upon phenotypic concordance, we reasoned that this same pathway might be relevant to pregnancy-induced disease; this hypothesis was strengthened by the knowledge that oxytocin activates peripheral tissues such as the myometrium through PLC activation. We observed that hydralazine, an antihypertensive agent that is FDA-approved in pregnancy and inhibits IP3-mediated events, achieved a near-complete rescue from dissection (5%) and abrogation of pathologic growth of the AscAo (0.23 ± 0.32mm/7wks; $p \leq 0.0001$). Hydralazine clearly outperformed a hemodynamically equivalent dose of propranolol that lowers blood pressure but does not address aberrant signaling (43% dissection and 0.59 ± 0.28mm/7wks; $p \leq 0.05$). In keeping with hypothesis, we also found that the recently FDA-approved MEK antagonist trametinib reduced both postnatal death due to aortic dissection (15%, $p \leq 0.0001$) and AscAo growth (0.64 ± 0.55mm; $p \leq 0.05$) in MFS mice. All outcomes showed direct correlation with the extent of ERK phosphorylation in the aortic wall. These data illustrate the power of modifier studies in defining therapeutic opportunity.

3229W

What to expect when she's expecting: The development of multidisciplinary obstetrical and neonatal management guidelines for Long QT syndrome types 1 and 2 in British Columbia. *S.F. McIntosh[1], J. Hathaway[2], J. Grewal[3], V. Rychel[4], H. Bos[5], E. Sherwin[6], A. Krahn[3], L.T. Arbour[1].* 1) Department of Medical Genetics, University of British Columbia (Island Medical Program), Victoria, BC, Canada; 2) BC Inherited Arrhythmia Program, St. Paul's Hospital, Vancouver, BC, Canada; 3) Division of Cardiology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada; 4) Department of Obstetrics and Gynaecology, St Paul's Hospital, Vancouver, BC, Canada; 5) Perinatology, Victoria General Hospital, Victoria, BC, Canada; 6) Division of Pediatrics, Department of Cardiology, University of British Columbia, Vancouver, BC, Canada.

Background: Although the majority of women with Long QT syndrome (LQTS) type 1 or 2 have uneventful pregnancies, extra monitoring of the mother, fetus and newborn from conception to the postpartum period is warranted to prevent adverse outcomes. Maternal issues include choice and titration of beta-blocker and monitoring for cardiac events, especially in the higher risk postpartum period. Fetal growth surveillance is required due to the association between beta-blockers and intrauterine growth restriction (IUGR). The newborn must be followed carefully due to the possibility of adverse effect of beta-blocker exposure in utero or through breast milk. To date, no comprehensive guidelines exist to guide clinicians on best practices for obstetrical and postpartum management of LQTS. Frequent clinical dilemmas prompted the BC Inherited Arrhythmia Program to form an expert group to develop Provincial consensus guidelines. The expert team, comprised of an obstetrician and cardiologist from a high-risk cardiac obstetric clinic, pediatric and adult electrophysiologists, geneticists, and a perinatologist, reviewed the literature and held meetings over a 2-year period to gather input, build consensus, and invite input from primary care providers. **Recommendations:** These consensus guidelines provide specific expert-level recommendations including: *Maternal-*Titrate beta-blocker up through the latter stages of pregnancy due to increasing blood volume and excretion, and back down to pre-pregnancy levels in the postpartum period -Schedule routine ECGs until 12 weeks postpartum to monitor/optimize beta-blockade -During labour, anaesthesia, and delivery: avoid potential triggers for ventricular arrhythmia such as QT-prolonging medications, electrolyte imbalance, excessive bleeding and prolonged 2nd stage pushing *Fetal-*Monitor fetal growth by ultrasound at 18, 28 and 36 weeks, with follow-up scan in two weeks and referral to obstetrician if any growth concerns identified *Neonatal-*Monitor for signs of beta-blockade (drowsiness, hypotension, and hypoglycemia) in the newborn and breastfeeding infant -Perform ECGs at 24 hrs and 3 weeks of age -Discuss genetic testing for LQTS (if known mutation in family) **Conclusion:** These multidisciplinary obstetrical and neonatal management guidelines for LQTS types 1 and 2 provide a foundation for further national and international practice guidelines. Evaluation from patient and provider perspectives will determine next steps.

3230T

Polymorphisms in CYP2C9 are associated with response to indomethacin among neonates with patent ductus arteriosus. *C.J. Smith[1], A.M. Momany[2], K.K. Ryckman[1,2], J.M. Dagle[2].* 1) Department of Epidemiology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, Division of Neonatology, University of Iowa, Iowa City, IA.

Background: Indomethacin is a non-steroidal anti-inflammatory drug (NSAID) used to treat patent ductus arteriosus (PDA). In general, approximately two-thirds of neonates experience closure of the PDA in response to indomethacin, while approximately one-third will require surgical ligation to close the ductus. Evidence suggests that response to indomethacin is highly heritable. We investigated the association between single nucleotide polymorphisms (SNPs) and PDA closure in response to indomethacin. **Methods**: Candidate SNPs were identified in the literature. A case-control analysis was performed among neonates who responded to indomethacin and neonates who required surgical ligation (56 cases, 104 controls). Independent transmission disequilibrium tests (TDT) were performed among parent-child trios of neonates who responded to indomethacin (102 probands) and neonates who required surgical ligation (50 probands). **Results**: rs2153628 was associated with response to indomethacin in the case-control analysis (OR: 0.521, 95% CI: 0.287, 0.947, p=0.0307). rs2153628 and rs1799853 were associated with response to indomethacin in the TDT analysis (OR: 2.357, 95% CI: 1.262, 4.404, p=0.00558 and OR: 2.50, 95% CI: 1.101, 5.676, p=0.0233, respectively). Haplotype analyses did not reveal any significant haplotype associations. **Conclusion**: We identified an association between two SNPs, rs2153628 and rs1799853, and response to indomethacin for the treatment of PDA. The G allele of rs2153628 conferred a protective effect against requiring surgery. The T allele of rs1799853 conferred a protective against requiring surgery. Both SNPs are located in *CYP2C9*, a member of the cytochrome P450 family of enzymes and a known metabolizer of NSAIDs. These findings suggest that response to indomethacin in the closure of PDA may in some cases be due to polymorphisms associated with less efficient metabolism of indomethacin.

3231F

Metabolomic study of amino acid, carnitine and acyl-carnitine profiles in Chinese preterm infants. *X. Zhao[1,2,3], X. Tao[1,2], P. Wang[1,2], Y. Tian[3], N. Zhong[1,2,4].* 1) Department of Medical Genetics, Peking University Health Science Center, Beijing, China; 2) Maternal & Children's Hospital of Lianyungang City, China; 3) Center of Translational Medicine, PLA General Hospital, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA.

Objective: Preterm birth (PTB) was defined as the delivery of an infant of less than 37 complete weeks of gestation. Due to the immaturity of the neonatal liver, premature infants may have a higher rate of out-of-range results [1]. Distinct metabolic patterns have been shown to be associated with different classes of gestational age in population of preterm and term infants [2]. The current study is to determine the impact of gestational age to metabolic profiles. **Material and Methods:** A total of 10,087 neonatal dried blood spot samples, including 300 of preterm and 9,787 of full-term infants, were collected between the 4th and 7th day post-delivery in South-western China. None of the infants was diagnosed with an inborn error of metabolism. The samples were prepared and analyzed by LC-MS-MS that screens 11 amino acids (Ala, Arg, Cit, Gly, Leu, Ile, Pro, Orn, Phe, Tyr, Val) and 32 carnitines. The result was processed by SPSS 19.0. For comparison of two groups of samples, the 2-tailed Wilcoxon Signed Ranks or Mann–Whitney test was used in suitable circumstances. **Results:** Compared to the term group, the concentration of Ala, Leu+Ile+Pro-OH, Pro,Val, C4DC+C5OH, C14, C16, C16:1OH, C18, and C18:1 among the 11 tested amino acids and 32 carnitines were lower in the preterm group (p<0.01), but the serum concentrations of Arg, C5, and C18:2 were higher in preterm group (p<0.01). **Discussion:** We reported here that the preterm infants were low for various amino acids and acyl-carnitines. Contrary to the findings of Mandout et al., who found the serum concentrations of Tyr, Phe, Met, Leu and Ala were higher in Indian preterm infants comparing with local term infants [3]. The low serum amino acid concentration maybe due to the reduced digestive activity of preterm infant, as Henderson et al reported that pepsin activity and postprandial output are significantly lower for preterm infants[4]. Our results demonstrate the complexity of understanding the impact of immaturity on metabolic profiles, and may lead to the development of specific amino acid, carnitine supplement for Chinese premature infants. 1. Pediatr Res 2001; **49**:125-9. 2. Clin Chem 2005; **51**:745-52. 3. Indian J Pediatr 2013; **80**:736-44. 4. Pediatr Res, 1998; **43**(4 Pt 2):1A-420A. .

3232W

Peripheral blood vitamin D-related microRNAs and mRNAs in early pregnancy of women with preeclampsia. *H. Mirzakhani[1,2], S. Liu[1], M. Santolini[1,2], A. Litonjua[1,2], A. Sharma[1,2], S. Weiss[1,2].* 1) Channing Division, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School.

Preeclampsia (PE) is a major cause of maternal and perinatal morbidity and mortality, complicating approximately 5% of pregnancies. Identification of circulating biomarkers could provide insight into the pathophysiology of PE and suggestion of preventive measures. Maternal vitamin D (25OHD) deficiency is associated with the development of PE. Using the data from Vitamin D Antenatal Asthma Reduction Trial (VDAART), we conducted a nested case control study of 157 pregnant women (47 with PE) to explore the integration of peripheral blood vitamin D-associated miRNA and mRNA expression profiles in association with PE at 10-18 weeks of gestation. We previously found 348 vitamin D-associated genes (158 up-regulated) demonstrating differential expression between the women who developed PE and controls (FDR < 0.05 in VDAART; p<0.05 in a replication OMEGA cohort). Functional enrichment analysis and interactome mapping of this gene set suggested several highly functional modules related to systematic inflammatory and immune responses including some known nodes with high degrees of connectivity (e.g. *IL8, MX1, HLA-DP1, HLA-DBQ1, MMP9, RPL31, ACTA2, IFII44L* and *SERPING1,* all previously reported in association with PE). Applying the same method for the mRNA expression profiling, we investigated the differently expressed miRNAs between the above 157 cases and controls. Furthermore, we used mirDip miRNA target prediction database (after filtering high confidence interactions with standardized scores >80, range 0-100) to investigate the potential regulatory miRNAs of the replicated PE gene signatures. Among a total of 602 expressed miRNAs (Ct<30), 467 miRNAs were analyzed. 68 differentially expressed miRNAs (41 downregulated) in women who developed PE compared to the controls were identified (FDR<0.05). Of these, 18 (14 downregulated) were vitamin D-associated miRNAs (FDR<0.05). Accordingly, miR-597 and miR-29a targets demonstrated significant enrichment in the gene set as potential regulators of the differentially expressed genes (p<0.05). miR-29a, a previously investigated miRNA in association with PE, was also among 18 vitamin D-associated miRNAs. Our data suggests expressed plasma miRNAs in early pregnancy might play a role in the regulation of distinctive genes involved in immunomodulatory responses and subsequent PE ocurrence, and might be potential biomarkers for diagnosis or prevention of PE. RT-PCR and functional validation of miR-597 and miR-29a are undergoing.

3233T

Clinical, pathologic, cytogenetic, and molecular profiling in self-identified Black women with uterine leiomyomata. *C. Gallagher[1,2], M. Hayden[3], Z. Ordulu[1,3], B. Quade[1,4], R. Anchan[1,3], N. Middleton[1,3], S. Srouji[1,3], E. Stewart[5], C. Morton[1,3,4,6,7].* 1) Harvard Medical School, Boston, MA; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA; 4) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 5) Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, MN; 6) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 7) University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

Uterine leiomyomata (UL), also known as fibroids, are the most common tumor of the female reproductive tract and are the primary indication for hysterectomy in the U.S. Black women are disproportionately affected by UL compared to other racial groups, having a greater lifetime incidence of developing fibroids and an earlier age of diagnosis. In order to elucidate molecular and genetic mechanisms responsible for the increased prevalence and morbidity associated with UL in black women, clinical, pathologic, cytogenetic, and select molecular profiling (*MED12* mutation analysis) of 75 self-reported black women undergoing surgical treatment for UL was performed. Mean age at time of surgical treatment was 39.5 years with 41 women undergoing myomectomy and 34 hysterectomy. The primary indication for medical or surgical intervention was menorrhagia with 53.3% of the cohort reporting this symptom. Our cohort demonstrated a more severe clinical presentation than previously conducted studies, specifically in regard to the number of fibroids per patient (18.2 vs. 9.9) and uterine weight (591.9 g vs 477 g) in the group undergoing hysterectomy. However, our cohort broadly resembles previous cytogenetic studies of UL tumors: Karyotypically abnormal tumors were detected in 30.7% of women and 17.9% of analyzed tumors. Therefore, data from our cohort are not supportive of an association between race and increased occurrence of cytogenetic abnormalities that may contribute to any population-specific morbidity or prevalence rate. In addition to the very complex cytogenetic abnormalities present in our cohort (24.1% of abnormal tumors), there is a notable decrease of t(12;14) (10.3% of abnormal tumors vs. expected 20%) and increased frequency of tumors with deletions of segments of chromosome 7 (51.7% of abnormal tumors vs. expected 17%). Our data on *MED12* mutation analyses (73.2% of tumors harbored a *MED12* mutation) provide additional support for a significant role of *MED12* in tumorigenesis. Although the effect of *MED12* tumorigenesis appears significant irrespective of race or ethnicity, downstream events such as karyotypic abnormalities appear different in black women. While more detailed analyses of individual cases and the inclusion of more comprehensively profiled cases are warranted, the case series presented herein demonstrates that certain genetic and molecular characteristics of UL alone do not explain the increased prevalence and morbidity of UL in black women.

3234F

Rare mutations in spermatogenesis genes underlie a significant fraction of idiopathic male infertility. *K.A. Fakhro[1], M. Arafa[2], H. El Bardisi[2], A. Robay[3], J.L. Rodriguez-Flores[4], J.G. Mezey[4], C. Abi Khalil[3], J. Malek[3], A. Al Ansari[2], S. Said[2], R.G. Crystal[4].* 1) Translational Medicine, Sidra Medical and Research Center, Doha, Qatar, Qatar; 2) Urology, Hamad Medical Corporation, Doha, Qatar; 3) Genetic Medicine, Weill Cornell Medical College - Qatar, Doha, Qatar; 4) Genetic Medicine, Weill Cornell Medical College, New York, NY USA.

Non-obstructive Azoospermia (NOA), the complete absence of sperm production, is the most severe form of male infertility, affecting up to 1% of the male population worldwide. Despite state-of-the-art clinical assessment, most NOA patients remain idiopathic and incur a significant socioeconomic burden between fertility treatments and surgical sperm retrieval procedures, both with limited success. The genetics of NOA are poorly understood primarily due to its sporadic nature coupled with the difficulty of isolating a genetic signal from a variety of environmental exposures in this adult-onset disease. To approach this challenge, we used whole exome sequencing (WES) to assess patients with idiopathic NOA. First, to increase the number of genes associated with causes for NOA, we identified 10 multiplex families and sequenced all available members (including fertile siblings). In all 10 families, we discovered rare mutations altering highly conserved residues segregating with disease. Seven of these occur in genes with previous evidence of having critical roles in spermatogenesis from mouse models and two are related to human infertility, but none were previously reported in isolated NOA. We combined our 10 new genes with 56 genes reported in the literature as associated with human spermatogenic failure to create an "NOA gene panel." Using WES, we assessed 66 additional unrelated subjects with "idiopathic" NOA and identified that 15 (23%) had 16 rare mutations altering highly evolutionarily conserved residues in 14 of these NOA panel genes. To determine if this was expected by chance, WES analysis was carried out on a population-matched group of 64 confirmed fathers. Of these, only 3 had mutations in the NOA gene panel, 2 of which were only heterozygous in recessive genes (p=0.0002 compared to the idiopathic NOA group). Altogether, our data indicates that for millions of idiopathic NOA men worldwide, WES could significantly improve the diagnostic yield if introduced as part of the routine patient work-up.

3235W

Amniotic fluid metagenomics in pregnancies affected with gastroschisis. *M.S. Patel[1], H. Wang[1], G.X. Yang[1], Y. Hu[2], A. Swenerton[1], D. Siciliano[1], K. Sangha[1], R.R. Miller[3], P. Tang[4].* 1) Dept. of Medical Genetics and Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 2) Department of Obstetrics and Gynaecology and Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 3) BC Centre for Disease Control and University of British Columbia, Vancouver, BC, Canada; 4) Department of Pathology, Sidra Medical and Research Center, Doha, Qatar.

We developed a method to use massively parallel sequencing followed by pathogen identification to ask the question - does gastroschisis more commonly affect pregnancies in younger mothers because they have more limited immunity to pathogens? That is, could some cases of gastroschisis be caused by a previously undetected infectious agent? To study this question, we obtained amniotic fluid at diagnostic amniocentesis from 12 pregnancies affected by gastroschisis and 24 maternal age matched, gestational age matched and season matched controls undergoing diagnostic amniocentesis for elevated maternal serum screen results. To start, we first determined which nucleic acid isolation column had the best performance using amniotic fluid spiked with different viruses and bacteria, with the QIAamp MinElute and PureLink columns found to work best, depending on the virus tested. We then optimized a metagenomics protocol to allow detection through massively parallel sequencing on an Illumina HiSeq 2500 instrument of RNA and DNA viruses regardless of size, with as little as 200 microliters of amniotic fluid or 15 pg starting material. Between 30 and 75 million reads were obtained from each sequencing library. Bioinformatics processing of the raw read files was performed to remove random primers, trim low quality bases, remove ribosomal sequences, remove duplicate reads and filter out human sequences, resulting in 4-8% of reads remaining. We assess the performance of this method by spiking control amniotic fluid with various infectious agents, performing sequencing and then using a variety of pathogen identification algorithms and a variety of taxonomic assignment algorithms as well as developing our own software. We describe the microbiome of midgestation human amniotic fluid from healthy pregnancies and also report our findings in pregnancies affected by gastroschisis.

3236T

A multi-stage genome-wide association study of uterine fibroids in African Americans. *J.N. Hellwege[1], J. Jeff[2], L.A. Wise[3], C.S. Gallagher[4], M. Wellons[1], K.E. Hartmann[1], S. Jones[1], E.S. Torstenson[1], S. Dickinson[5], E.A. Ruiz-Narváez[3], N. Rohland[4], A. Allen[4], D. Reich[4,6,7], B. Pasaniuc[8], N. Mancuso[8], H.K. Im[5], D.A. Hinds[9], J. Palmer[3], L. Rosenberg[3], J.C. Denny[1], D.M. Roden[1], E.A. Stewart[10], C.C. Morton[4,7,11], E. Kenny[2], T.L. Edwards[1], D.R. Velez Edwards[1].* 1) Vanderbilt University Medical Center, Nashville, TN; 2) Icahn School of Medicine at Mount Sinai, New York City, NY; 3) Boston University, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) University of Chicago, Chicago, IL; 6) Howard Hughes Medical Institute, Chevy Chase, MD; 7) Broad Institute of Harvard and MIT, Cambridge, MA; 8) University of California at Los Angeles, Los Angeles, CA; 9) 23andMe, Inc., Mountain View, CA; 10) Mayo Clinic, Rochester, MN; 11) University of Manchester Academic Health Science Centre, Manchester, England.

Uterine fibroids (UF) are common, benign tumors of the uterus that affect 77% of women by menopause, are the leading indication for hysterectomy, and account for costs of up to $34 billion annually in the United States. African American (AA) race and increasing age up to menopause are the strongest known risk factors. AA women are more likely to suffer from UF than European American women, and on average have earlier onset as well as larger and more numerous fibroids. We conducted a multi-stage genome-wide association study (GWAS) of UF risk among AAs followed by *in silico* genetically predicted gene expression profiling of top hits. In Stage 1, UF AA cases and controls (1273 cases, 1379 controls) were confirmed by pelvic imaging, either from a validated algorithm applied to electronic health records (PPV 98%, NPV 96%) at eMERGE Network sites or study ultrasounds in the CARDIA Women's Study. Confirmation of UF status by imaging reduces rates of misclassification of cases with subclinical UFs. Stages 2 and 3 used *in silico* replication from the imputed 23andMe, Inc. survey-based data (1354 AA cases, 1419 AA controls) and low-pass sequencing data from the Black Women's Health Study (2511 AA self-reported cases, 2231 AA self-reported controls). Stage 1 was genotyped by GWAS arrays and was imputed to the 1KGP reference haplotypes. Genetic UF risk was modeled by logistic regression adjusted for principal components of ancestry, and was followed by a fixed effects meta-analysis of results from all three stages. We observed a genome-wide significant association among 5528 AA cases and 6513 AA controls at the African ancestry-specific SNP rs10939126 (minor allele frequency=0.07) in *TBC1D19* on chromosome 4 (OR=1.69, 95% CI 1.47–1.91, p-value=3.21×10^{-9}, p_{het}=0.46). *TBC1D19* (TBC1 domain family member 19) is most highly expressed in uterine tissue from GTEx, supporting its role in uterine biology. Evaluation of the genetic association results with MetaXcan identified nominally significant genetically predicted gene expression of *TBC1D19* in breast tissue (p-value=0.0087). This suggests tentative support for a mechanism underlying the genetic result, as both uterine and breast tissues are hormone-sensitive. Overall, we performed the first multi-stage GWAS for UF among AA women, and identified a novel African-specific locus within *TBC1D19* with potential biological relevance for UF through genetically predicted gene expression. .

3237F

Admixture mapping of uterine fibroid characteristics using African American women from BioVU. *M.J. Bray[1], T.L. Edwards[1,2,3,4], M.F. Wellons[2], K.E. Hartmann[2,3,5], D.R. Velez Edwards[1,3,4,5].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN; 3) Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 4) Institute of Medicine and Public Health, Vanderbilt University, Nashville, TN; 5) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN.

Uterine fibroids are benign tumors affecting up to 77% of women by menopause and cost the US up to 9.4 billion dollars annually in healthcare expenditures. A strong risk factor for fibroids is African ancestry, where African Americans (AAs) have larger and more numerous fibroids than European Americans (EAs). It is possible that differences in fibroid characteristics between AAs and EAs are driven by genetic loci that differ in frequency between AAs and EAs. To address this, we performed admixture mapping (AM) analyses on fibroid tumor characteristics including volume, largest dimension, and number in AA women. Using BioVU, a clinical cohort within the Vanderbilt University Medical Center, we abstracted information regarding fibroid size and number from the medical records of 439 AA women. Using existing genome-wide association studies (GWAS) data, we conducted AM analyses using Local Ancestry in adMixed Populations-ancestral (LAMP-ANC) to call ancestry at each single nucleotide polymorphism (SNP) and subsequently performed regression analyses on $\log_{10}$ transformed volume of all fibroids, $\log_{10}$ transformed largest dimension of all fibroids, and number (1 vs >1). Afterwards, we conducted single SNP association analyses on imputed genetic data under the strongest mapping peaks. We limited all analyses to common SNPs (minor allele frequency >5%) and adjusted for age, body mass index, and five principal components for all analyses. After performing AM analyses, preliminary results highlight AM peaks on chromosomes 2p24.1 and 10q24.1 for volume, on chromosome 14q32.11 for largest dimension, and on chromosomes 7q21.11 and 14q24.3 for number. The strongest associations in the AM peak 10q24.1 for volume were in the gene slit guidance ligand 1 (*SLIT1*). After performing single SNP association analyses on suggestive AM peaks, single SNP association analyses for max dimension on chromosome 14q24.3 highlighted coiled-coil domain containing 88C (*CCDC88C*) (rs11160001, Beta = 0.18, 95% confidence interval (CI) 0.09-0.26, p = 5.79×10^{-5}). A recent study evaluating the associations of breast cancer risk variants in AA women from previous GWAS found that *CCDC88C* was associated with ER+ breast cancer in AA women. Future studies involve validating our AM study findings for tumor characteristics using an independent AA cohort. This work will lay the foundation for understanding the relationship between the genetic risk factors and phenotypic heterogeneity of fibroids.

3238W

The importance of clinical genetics during pregnancy via a hypophosphatasia family due to a novel mutation. *I. Boga[1], A. Bisgin[1], C. Cetin[2], S. Buyukkurt[2].* 1) Department of Medical Genetics, Balcali Clinics and Hospital,Cukurova University Faculty of Medicine, Adana, Turkey; 2) Department of Obstetrics And Gynecology, Perinatology Unit, Balcali Clinics and Hospital, Cukurova University Faculty of Medicine, Adana, Turkey.

   Objective: Hypophosphatasia is an autosomal recessive rare genetic disease with profound skeletal hypomineralization and respiratory compromise, caused by mutations in the Alkaline Phosphatase *(ALPL)* gene. We here report an infantile form of hypophosphatasia confirmed by mutational analysis and discuss the importance of diagnosing a rare disease for proper genetic counselling that include differential diagnosis, clinical genetics assessment and prenatal molecular testing. **Methods:** The pregnant women referred to the Medical Genetics Clinic due to having a child history with severe convulsions and rickets/osteomalacia. The genetic analysis performed in the *ALPL* gene by Sanger sequencing. Chorionic villus samples were obtained for prenatal diagnosis. **Results:** The history taking and clinical assessment revealed a possible diagnosis of hypophosphatasia. A novel p.I218S (c.653T>G) homozygous mutation was detected in a child. While the child had been found to have hypophosphatasia on the basis of both clinical genetic and molecular analysis, only the mother was carrier with the heterozygous mutation in the same position and the father does not have any mutation in the *ALPL* gene. We could not get the permission to the SNP array to confirm paternity. Obviously, the family was counselled to provide accurate information and possibilities. As a result, the family chose to undergo prenatal diagnosis that revealed no mutation in the foetus. **Conclusion:** Clinical genetic assessment and prenatal genetic counselling is thus considered an importance for confirming a diagnosis of rare disorders such as hypophosphatasia.

3239T

Validation of targeted next-generation sequencing for preimplantation genetic diagnosis of heritable connective tissue diseases. *S. Symoens[1], P. Van Acker[1], E. Van Holm[1], O. Vanakker[1], F. Malfait[1], C. Christodoulou[2], B. Heindryckx[2], P. De Sutter[2], E. Van den Abbeel[2], B. Callewaert[1], J. De Backer[1], A. De Paepe[1], S. Janssens[1], P. Coucke[1].* 1) Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Oost-Vlaanderen, Belgium; 2) Ghent Fertility and Stem Cell Team (G-FaST), Department of Reproductive Medicine, Ghent University Hospital, 9000 Ghent, Belgium.

   Preimplantation genetic diagnosis (PGD) is a diagnostic procedure that allows testing of preimplanted embryos for couples who are at-risk for a genetically defined monogenic disorder or a chromosomal abnormality. PGD for monogenic disorders has benefitted from the introduction of next-generation sequencing (NGS), as this technique enables the detection of a single mutation on a few (embryonic) cells in a more accurate manner. The Center for Medical Genetics Ghent functions as a national and international reference center for heritable connective tissue diseases (HCTD), including Marfan syndrome and other familial thoracic aortic aneurysm syndromes, osteogenesis imperfecta, and Ehlers-Danlos syndromes. These syndromes comprise a heterogeneous group of monogenic disorders that affect the normal development and homeostasis of the hard (bone, cartilage) and soft (skin, blood vessels, tendon and ligaments, ocular structures) connective tissues. Many of these conditions are associated with important morbidity and increased mortality, which justifies the use of prenatal diagnosis and PGD. We applied NGS to implement PGD for patients with these conditions in our center. Couples with a genetically defined HCTD and interested in PGD are counseled by a clinical geneticist and written informed consent is obtained. First, an optimization experiment is performed in order to validate the detection of the familial mutation on a limited number of cells. In total, for each proband 32 aliquots are isolated from an EDTA-blood sample each comprising 2-8 lymphocytes. Subsequently, lymphocyte aliquots are lysed and a PCR reaction encompassing the familial mutation is performed *in duplo* (PCR-A and PCR-B). Both PCR products are NGS sequenced on two separate MiSeq runs. Following this strategy, possible allele-drop-out and PCR performance on a minimal amount of DNA is evaluated. Hitherto, 3 *FBN1*, 2 *SMAD3* and one *TNNT2* at-risk couples have undergone this PGD validation phase, which has been successful for all. In a next step, trophectoderm cells are biopsied from day 5-6 embryos and the underlying familial mutation is examined by NGS. These results demonstrate that PGD through direct mutation detection by NGS is a powerful prenatal approach. Further optimization and validation for a larger cohort of HCTD patients followed by pre- and/or post-natal confirmation of the familial mutation is ongoing to calculate the success rate.

3240F

High-throughput screening for *SMN1* "2+0" copy number status by next-generation sequencing. *E.D. Boyden, K.A. Robinson, D.K. Neitzel, N.E. Faulkner, G.J. Porreca, M.A. Umbarger.* Good Start Genetics, Cambridge, MA.

Spinal muscular atrophy (SMA) is a lethal autosomal recessive neuromuscular disorder caused by functional loss of the *SMN1* gene, with a carrier rate of ~1 in 40. Most mutations involve *SMN1* copy number (CN) loss, produced by either deletion of part or all of *SMN1*, or more commonly by conversion of *SMN1* to *SMN2,* a linked paralog that encodes an identical protein but is poorly expressed due to a silent coding variant that disrupts proper splicing. We previously developed a protocol by which the CN status of *SMN1* and *SMN2* can be measured directly by next-generation sequencing (NGS). This method, which uses molecular inversion probes to capture loci specific to *SMN1* and *SMN2*, shows sensitivity and specificity for detection of *SMN1* CN loss similar to that of multiplex ligation-dependent probe amplification (MLPA), but has the advantage of being compatible with automated high-throughput screening. However, both methods typically determine CN status without the benefit of haplotype information, which renders them unable to identify "2+0" carriers. These individuals, who may represent between 4 and 27 percent of all SMA carriers depending on ethnicity, have 2 functional *SMN1* alleles on one chromosome and 0 on the other. Accordingly, they appear to be noncarriers by CN status, yet are in fact carriers as they may still transmit to offspring a chromosome that lacks a functional copy of *SMN1*. Recently, Luo et al. (Genet Med 2014) discovered markers that are commonly linked to *cis* duplications of *SMN1*; the presence of these markers in individuals with 2 copies of *SMN1* may indicate "2+0" carriers. Our NGS-based SMA screening method can reliably detect these markers, and so our protocol can be adapted to identify potential "2+0" carriers. Hence, this modification provides an avenue to improve the sensitivity of our method relative to that of other SMA carrier screening approaches.

3241W

Multigene molecular testing provides significant diagnostic yield for prenatal holoprosencephaly. *S. Duffer, J. Meck, L. Vincent, E. Kramer Dugan, V. Zvereff, G. Richard.* GeneDx, Gaithersburg, MD.

Holoprosencephaly (HPE), the most common malformation of the forebrain, occurs in 1/250 conceptuses and is readily detected by prenatal ultrasound. Karyotype abnormalities account for HPE in 40-50% of fetuses, but clinical sensitivity information for prenatal molecular testing of the four major HPE-associated genes *SHH, ZIC2, SIX3, TGIF* is limited. Here we report the detection rate for pathogenic and likely pathogenic variants (PV/LPV) identified by sequencing and copy number analysis in these genes in prenatal cases. **Method:** Data was reviewed for 176 fetuses with sonographic features of HPE referred for testing from 2003-2015. Of these cases, 55% reported negative karyotype and/or microarray results prior to molecular testing. Sanger sequencing analysis of *SHH, ZIC2, SIX3,* and *TGIF* was complemented by Multiple Ligation Probe Amplification (MLPA) for deletion/duplication analysis of these genes. **Results:** PV/LPV were identified in 41/176 (23%) prenatal HPE cases, including distinct variants in *ZIC2* (17), *SHH* (13), *SIX3* (8), and *TGIF* (3). More than half (51%; 20/39) of the PV/LPV sequencing variants resulted in loss of function. Only three copy number variants were detected, including whole gene deletions involving either *SHH* or *SIX3*, and a very rare, apparently *de novo* duplication of *SHH*. Additionally, three (1.7%) fetuses were heterozygous for a sequence variant of uncertain significance in the *SHH, ZIC2,* or *SIX3* gene, respectively. Parental testing in 24 families determined that nine variants (37%) occurred apparently *de novo* in the fetus, while twelve (63%) were inherited. Clinical parental information was available for nine of the inherited variants, with seven parents reporting mild clinical indications of HPE (e.g. hypotelorism) and/or a family history of previous pregnancies with HPE. **Conclusions:** In our large cohort, molecular testing of the major HPE genes determined the cause of HPE in 23% (41/176) of cases, validating results of a previous study (Bendavid et al., 2006). In both studies, *ZIC2* presented as the highest molecular contributor to prenatally detected HPE. The high frequency of inherited PV/LPV highlights the known intra-familial variability of symptoms, the importance of parental testing, and the need for careful clinical examination for minor features of HPE. We conclude that the combination of chromosome analysis and HPE panel testing will identify a causative genetic abnormality in over 70% of fetuses with prenatally detected HPE.

3242T

Research on maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting) in mothers and their children with schizophrenia. *K. Lee[1], J. Park[1,2], B. Lee[1,2], Y. Lee[1,2], E. Moon[1,2], H. Jeong[1].* 1) Department of Psychiatry, Pusan National University Hospital, Busan, Republic of Korea; 2) Medical Research Institute, Pusan National University Hospital, Busan, Republic of Korea.

   Background : Schizophrenia emerges during early stages of life and it has various effects such as causing considerable familial and social burden, as well as increasing health expenses. Also, it is disabling disease associated with heterogeneous psychiatric phenotypes. It is assumed that this disease is caused by a combination of genetic and/or environmental factors during critical periods of brain development increase the risk for this illness. **Introduction :** Various genetic effects like epigenetic effects might be operated by the inter-uterine environment mechanisms. The effects of parent-of-origin (imprinting), maternal genotype, and maternal-fetal genotype interactions could be investigated by collecting DNA from affected offspring and their mothers (case/mother duos) and comparing this with an appropriate control sample. Another method uses data from cases and both parents (case/parent trios) but does not require controls. In this study, we describe a novel implementation of a multinomial modeling approach that allows the estimation of such genetic effects using either case/mother duos or case/parent trios. We investigate the effects of estimation of maternal, imprinting and interaction effects using multimodal modeling using parents and their offspring with schizophrenia in Korean population. **Subjects and Methods:** We have recruited 27 probands(with schizophrenia) with their parents and siblings whenever possible. We analyzed 96 SNPs(Single Nucleotide Polymorphisms) of 17 functionally only relevant genes and 21 neuronal genes relevant to schizophrenia for DNA samples that was checked for the data quality and genotype error. We used EMIM analysis program for the estimation of maternal, imprinting and interaction effects using multimodal modeling.**Results:** Of analyzed 96 SNPs functionally and positionally relevant to schizophrenia, very significant SNP (rs 324420) was suggested in EMIM analysis for child genetics effects. (p= 1.5 □ 10⁻⁴) rs 324420 was also significant for child genetic effects allowing for maternal genetic effects (p = 5.3 □ 10⁻⁴) with very stringent multiple comparison Bonferroni correction. Additionally, no significant SNP was presented in the analysis results for maternal genetic effects allowing for child genetic effects.

3243F

Clinical consequences of whole-genome sequencing vs. SNP methods of noninvasive prenatal screening. *C.G. Artieri, C. Havery, K.A. Beauchamp, I.S. Haque, D. Muzzey.* Counsyl Inc., South San Francisco, CA.

   Noninvasive prenatal screening (NIPS) for aneuploidies via next-generation sequencing of cell-free DNA has gained wide clinical adoption due to its high accuracy. The most widely used NIPS offerings employ either whole-genome sequencing (WGS) or SNP-based methods to determine whether patterns deviate from those expected in a euploid pregnancy. Although published outcome studies and clinical guidelines indicate that both approaches have comparable performance [1], claims regarding the limitations of WGS methods persist, causing confusion among clinicians and patients. Here we directly and quantitatively compare the algorithmic requirements, performance, and clinical consequences of WGS vs. SNP-based NIPS. To enable fair comparison, we used published algorithms and assay parameters [3,4] to implement computational simulations for each method, allowing us to compare the methods in silico on a limitless number of euploid and aneuploid pregnancies. We show that the SNP-based method is sensitive to noise in sampling of allele frequencies, which is well represented by a beta-binomial distribution. In contrast, the depth-based WGS method is sensitive to the negative-binomial noise commonly observed in sequencing counts. By simulating data over a range of parameters congruent with published data, we show that WGS NIPS outperforms the SNP-based method in terms of both sensitivity and specificity, even when simulating the latter with less sampling noise than indicated by published data. Furthermore, we demonstrate that this sensitivity deficit results from the SNP-based method's dependence on the aneuploidy parent of origin: the deviation in allele frequencies produced by maternally inherited trisomies — up to 90% of Down Syndrome cases — is far less easily detectable as compared to rare paternal trisomies. In contrast, parental origin is irrelevant to WGS-based methodology, as it does not affect the depth signal. Finally, we evaluate the clinical consequences of how each NIPS strategy handles samples with low fetal fraction. Importantly, because only a subset of patients pursue an invasive test after receiving a "no-call" from noninvasive screening [4], clinics offering tests that routinely "no-call" low fetal fraction samples do not ultimately improve aggregate prenatal detection of aneuploidies in their patients. 1. ACOG/SMFM Committee Opinion #640 2015 2. Ryan et al. 2016. Fetal Diagn Ther 3. Lefkowitz et al. 2016. Am J Obstet Gynecol 4. Currier et al. 2012. Prenat Diagn.

3244W

De novo mtDNA point mutations are common and have a low recurrence risk. *C.E.M. de Die-Smulders[1,2], H.J.M. Smeets[1,2,3], A.T.M. Hendrickx[1], D.M.E.I. Hellebrekers[1], I.F.M. de Coo[4], C.L. Alston[5], C. Knowles[5], R.W. Taylor[5], R. McFarland[5], S.C.E.H. Sallevelt[1].* 1) Department of Clinical Genetics, Maastricht University Medical Center+ (MUMC+), Maastricht, The Netherlands; 2) Research School for Developmental Biology (GROW), Maastricht University, Maastricht, The Netherlands; 3) Research School for Cardiovascular Diseases in Maastricht, CARIM, Maastricht University, Maastricht, The Netherlands; 4) Department of Neurology, Erasmus MC-Sophia Children's Hospital Rotterdam, Rotterdam, The Netherlands; 5) Wellcome Trust Centre for Mitochondrial Research, Institute of Neuroscience, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom.

   Background Severe, disease-causing germline mitochondrial (mt)DNA mutations are maternally-inherited, or arise *de novo*. Strategies to prevent transmission are generally available, but depend on recurrence risks, ranging from high/unpredictable for many familial mtDNA point mutations to very low for sporadic, large-scale single mtDNA deletions. Comprehensive data are lacking for *de novo* mtDNA point mutations, often leading to misconceptions and incorrect counseling regarding recurrence risk and reproductive options. We aim to study the relevance and recurrence risk of apparently *de novo* mtDNA point mutations. **Methods** Systematic study of prenatal diagnosis (PND) and recurrence of mtDNA point mutations in families with *de novo* cases, including new and published data. "*De novo*" based on the absence of the mutation in multiple (postmitotic) maternal tissues is preferred, but mutations absent in maternal blood only were also included. **Results** In our series of 105 index-patients (33 children and 72 adults) with (likely) pathogenic mtDNA point mutations, the *de novo* frequency was 24.6%, the majority being pediatric. PND was performed in subsequent pregnancies of mothers of 4 *de novo* cases. A fifth mother opted for Preimplantation Genetic Diagnosis (PGD) because of a coexisting Mendelian genetic disorder. The mtDNA mutation was absent in all 4 prenatal samples and all 11 oocytes/embryos tested. A literature survey revealed 137 *de novo* cases, but PND was only performed for 9 (including 1 unpublished) mothers. In one, recurrence occurred in 2 subsequent pregnancies, presumably due to germline mosaicism. **Conclusions** *De novo* mtDNA point mutations are a common cause of mtDNA disease. Recurrence risk is low. This is relevant for genetic counseling, particularly for reproductive options. PND can be offered for reassurance.

3245T

The role of the mitochondrial bottleneck in mtDNA disease risk and embryogenesis in zebrafish (non-)germline cells and human preimplantation embryos. *A. Otten[1], T. Theunissen[1], S. Sallevelt[1], J. Dreesen[1], M. Adriaens[2], J. Derhaag[3], M. Gerards[2], J. Vanoevelen[1], D. Samuels[4], A. Stassen[1], M. Muller[5], C. de Die[1], H. Smeets[1].* 1) Genetics and Cell Biology, GROW, Maastricht University, Maastricht, Limburg, Netherlands; 2) Maastricht Centre for Systems Biology (MACSBio), Maastricht University, Limburg, Netherlands; 3) Department of Obstetrics & Gynaecology, GROW, Maastricht University, Maastricht, Limburg, Netherlands; 4) Vanderbilt University School of Medicine, Nashville, USA; 5) GIGA, Université de Liège, Liège, Belgium.

   The mitochondrial DNA (mtDNA) inherits through a segregational bottleneck: a limited number of the mtDNA molecules from the oocytes pass to the offspring. We studied size, timing and mechanisms underlying this bottleneck in zebrafish embryos and in rest oocytes/embryos from preimplantation genetic diagnosis (PGD) for mtDNA disease. In zebrafish, the mtDNA content in maturing oocytes increases proportionally with the size of the cells and mature oocytes contain ~19.0×10^6 mtDNA molecules with high intra-individual variation (range: 3.3×10^6 – 42.2×10^6). During embryogenesis, the mtDNA copy number decreases to ~170 for FACS-isolated primordial germ cells (PGCs), a number similar to mammals, and to ~50 for non-PGCs. As the lowest mtDNA copy numbers occur at a fixed developmental stage, variation in the oocyte copy number dictates the mtDNA amount at the bottom of the bottleneck. To test whether a low copy number at the bottom of the bottleneck renders an oocyte vulnerable for *de novo* mtDNA mutations reaching significant heteroplasmy levels, we performed NGS of the mtDNA in 103 individual oocytes from eight zebrafish. Randomly scattered *de novo* mtDNA mutations were detected in ~20% of the oocytes with heteroplasmy levels above the detection limit of 1.5%. This is in line with our observation that ~20% of the germ cells have an mtDNA copy number ≤73 and would lead to detectable mutation loads. As the frequency of *de novo* mutations is close to reported error rate of the mitochondrial replication enzyme POLG, this seems to be the predominating factor. Moreover, this process could also explain the high frequency of *de novo* mtDNA mutations (~25%) observed in human patients. The mechanisms underlying the mtDNA bottleneck were further characterized from the mutation load distribution in human oocytes/embryos of m.3243A>G or m.8993T>G mutation carriers. In addition to genetic drift selection occurred on OXPHOS capacity and mitochondrial membrane potential, explaining the absence of m.3243A>G loads above 80% and the presence of m.8993T>G loads close to 100%. We calculated that proper embryonic development would require at least 50,000 copies of mtDNA, a figure similar to mice. In conclusion, complementary studies in zebrafish and human embryos have generated novel insights for the mtDNA bottleneck and for its role in mtDNA disease initiation and transmission and in fertility.

3246F

Prenatal diagnosis of non-viable anomalous fetuses by exome sequencing. *N. Vora[1], B. Powell[2], K. Foreman[2], E. Hardisty[1], K. Gilmore[1], A. Brandt[2], D. Marchuk[2], A. Lyerly[3], C. Rini[4], D. Skinner[5], C. Powell[2], K. Weck[2], J. Evans[2], J. Berg[2].* 1) Department of Obstetrics & Gynecology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Social Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Gillings School of Global Public Health; UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) FPG Child Development Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Our objective was to investigate the diagnostic performance of exome sequencing (ES) in non-viable fetuses with sonographic abnormalities who had normal microarray analysis. Given the complexities of analysis and return of results, we also sought to understand patient perspectives and knowledge of ES. ES was performed on DNA extracted from chorionic villi, amniocytes, or cord blood in 7 non-viable fetuses and from peripheral blood from their parents (12 additional trios are currently being sequenced). Trios were sequenced simultaneously. Prenatal-specific gene lists and bioinformatics pipelines specific to trio analysis were developed. Variant curation was performed manually and results were presented at a multidisciplinary meeting. Parents were consented for the return of 1) diagnostic results in the fetus; 2) medically actionable findings in the parents; 3) carrier status for significant autosomal recessive conditions in which both parents are carriers. Reported variants were confirmed by Sanger sequencing. A mixed-methods assessment using questionnaires and semi-structured interviews with 19 mother-father dyads was completed. In 4/7 (57%), ES provided a diagnosis or possible diagnosis of the following disorders: osteogenesis imperfecta type 3 (*COL1A1*), fetal akinesia sequence (*MUSK*), scalp-ear nipple syndrome (*KCTD1*), primordial microcephaly-dwarfism syndrome (*RTTN*). In two other cases, a single mutation in an autosomal recessive gene was identified. Incomplete sequencing coverage and the possibility of undetected deletions precluded exclusion of a second mutation. One medically actionable finding in a parent (familial hypercholesterolemia) was reported. None of the couples had significant carrier results to report. The average perceived likelihood (5.4/10) that this technique would explain the results was consistent with pre-test counseling. Most families (16/19) verbalized hope that sequencing would provide an answer. Women in the highest income levels scored highest on the pre-sequencing genetics literacy assessment. ES offers a useful technique to obtain a prenatal diagnosis in a cohort of non-viable fetuses with sonographic abnormalities and normal microarrays. Use of ES in this group may improve clinical care, reproductive decision-making, and refine phenotypes for known genes. Challenges related to genetics literacy, diagnostic capability, and variant interpretation must be addressed by highly tailored pre- and post-test genetic counseling.

3247W

Enhanced screening performance of a SNP-based NIPT for five clinically significant microdeletions in a large clinical cohort. *K. Martin, A. Kalyan, C. Lan, M. Stosic, K. Kobara, H. Ravi, T. Truong, S. Iyengar, A. Ryan, Z. Demko.* Natera, San Carlos, CA.

Objectives: To report on: (1) the clinical performance of a single-nucleotide polymorphism (SNP)-based noninvasive prenatal test (NIPT) for screening 5 clinically significant microdeletions [at the 22q11.2, 1p36, Cri-du-chat (CdCS), and Prader-Willi (PWS)/Angelman syndrome (AS) regions]; (2) the test's performance using an updated approach involving increasing the confidence threshold of the algorithm, and reflex sequencing of high-risk calls at a higher depth of read. **Method:** 58,501 consecutive clinical samples received for 22q11.2 del testing from Aug 2014–Feb 2015, and 42,318 samples received for 1p36, CdCS, PWS, and AS del testing from Feb 2014–Feb 2015, were analyzed at a CAP-accredited, CLIA-certified lab using a SNP-based NIPT. The same samples were retrospectively reviewed using the updated approach. Follow-up information from confirmatory diagnostic testing for the targeted regions and details of any ultrasound findings was collected for the high-risk cases to determine the positive predictive value (PPV) and false-positive rate (FPR) of the test. **Results:** The commercial test reported 403 cases as high-risk for a fetal microdeletion, approximately 80% of which became low-risk using the updated approach, primarily due to reduction in the number of FP calls. Follow-up information revealed PPVs ranging from 1.5%–20% with the original approach, depending on the condition screened, each of which increased 2.5–7x with the updated approach **(Table 1)**. For subcohorts with average *a priori* risk, the PPVs for selected microdeletions with the updated method were: 22q11.2, 12.5%; 1p36, 50.0%; and CdCS, 50.0%. **Conclusions:** An increased confidence threshold and reflexing at higher-depth led to marked improvements, with minimal loss in sensitivity, in the performance of a SNP-based NIPT for 5 clinically significant microdeletions, most notably due to a reduction in the FPR. Early intervention in children affected with these deletion syndromes can lead to improved outcomes and an enhanced quality of life. Population-wide implementation of this updated SNP-based NIPT for microdeletions has the potential to realize this outcome.

Table 1: Performance of the SNP-based NIPT

	22q11.2 del	1p36 del	CdCS del	PWS	AS
High-risk call-original approach, (N)	188	25	75	4	111
High risk calls-updated approach, (N)	54	7	7	4	17
PPV-original approach, (%)	14.1	20	8.9	-	1.5
PPV-updated approach, (%)	38.7	50	66.7	-	10
FPR-original approach, (%)	0.31	-	-	-	-
combined					0.58
FPR-updated approach, (%)	0.06	-	-	-	-
combined					0.09

3248T

Development of a novel targeted assay for non-invasive prenatal testing of fetal trisomies and sub-chromosomal abnormalities exhibits diagnostic level accuracy. *P.C. Patsalis[1,2], E. Kypri[1,2], K. Tsangaras[2], A. Achilleos[2], P. Mina[2], M. Neofytou[1], E.A. Papageorgiou[1,2], M. Ioannides[2], G. Koumbaris[1,2].* 1) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) NIPD Genetics Ltd, Nicosia, Cyprus.

Purpose There is a great need for the development of highly accurate, cost effective technologies, which can facilitate the widespread adoption of Non Invasive Prenatal Testing (NIPT). We hereby present a novel cost effective assay of unparalleled accuracy which overcomes the limitations of current technologies. *Methods* This method enables the targeted analysis of selected genomic regions at very high sequencing depth and allows highly accurate fetal fraction determination to ensure extremely accurate aneuploidy detection. The analytical performance of the assay was evaluated using a proprietary bioinformatics pipeline that combines the results of four statistical models, including cell free DNA (cfDNA) fragment size information. A validation study, which included 611 samples derived from pregnancies of at least 10 weeks of gestation, was performed. Furthermore, in a proof of principle study, we applied this technology for the detection of microdeletion/microduplication syndromes using samples from normal pregnancies and spike-in samples simulating 5%, 10% and 20% fetal fraction. *Results* The validation study exhibited 100% sensitivity and specificity and correctly classified 52/52 (95% CI: 93.2-100%) cases of trisomy 21, 16/16 (95% CI: 79.4-100%) cases of trisomy 18, 5/5 (95% CI: 47.8-100%) cases of trisomy 13, and 538/538 (95% CI: 99.3-100%) normal cases. The test also correctly identified fetal sex in all cases (95% CI: 99.4-100%). In an additional proof of principle study, the test correctly detected the microdeletions and microduplications in all abnormal samples exhibiting 100% sensitivity and 100% specificity. *Conclusion* The clinical impact of cfDNA testing has been significant as indicated by its quick adoption in prenatal care. This novel technology overcomes limitations of current technologies and exhibits diagnostic level performance. This method enables accurate and cost-effective non-invasive fetal aneuploidy detection of trisomy 21, 18 and 13 as well as sub-chromosomal abnormalities which is critical for wide-spread adoption of NIPT.

3249F

Non-invasive prenatal testing for *de novo* **dominant monogenic diseases using maternal cell-free DNA.** *J. Zhang[1], J. Li[2], Y. Feng[2], H. Dai[2], G. Wang[2], X. Ge[1], X. Wang[1], A. Purgason[2], H. Mei[1], A. Breman[1], I. Van den Veyver[1,3], A. Beaudet[1], L. Wong[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Departments of Gynecology, Baylor College of Medicine, Houston, TX.

Background : It is estimated that *de novo* dominant mutations in ~500 genes cause severe intellectual disability in 3-5 per 1000 newborns which is 3-5 times more frequent than Down syndrome. Although prenatal non-invasive testing using maternal blood containing cell free fetal DNA (cffDNA) has been demonstrated to have better performance over conventional serum screening for women at increased risk of a pregnancy with aneuploidy for chromosome 21, 13 or 18, availability of non-invasive prenatal screening for severe fetal *de novo* Mendelian disease is currently limited. In this study, we aimed to develop a new non-invasive prenatal screening method using cffDNA to detect pathogenic variants in 43 genes associated with common dominant monogenic diseases. **Methods:** After molecular barcoding of ~ 6-10 ng input DNA by adaptor ligation and hybridization-based target enrichment, next-generation sequencing (NGS) on an Illumina platform was used to generate sequencing data with average read-depth of >2,000X in the targeted region. DNA fragments harboring the informative paternal alleles were counted after applying a mathematical algorithm involving a de-duplication step based on unique molecular barcodes and error correction for PCR, library construction and/or sequencing artefacts. Fragmented genomic DNAs with similar size to cffDNA from mother/child pairs, plasma cell free DNA admixture from female/male donors, Horizon cell free reference DNA as well as plasma cell free DNA from pregnant women at 10-18 week gestational age were used for the analytical and clinical validation. **Results:** Approximately >99% of the targeted region has a minimum coverage of 400X in order to call *de novo* mutations in the fetus. Among ~ 180 highly polymorphic loci included in this panel, all informative paternal alleles were evaluated and the analytical sensitivity was >97% in four cffDNA specimens with fetal fraction ranging 3.1-14%. The analytical specificity is > 98.5%. A pathogenic variant in *PTPN11* associated with Noonan syndrome was successfully identified in one spike-in control sample. **Conclusion:** We developed a highly sensitive and specific NIPT method to detect *de novo* mutations in the cffDNA. Clinical validation studies on larger numbers of samples from pregnant women are in progress to evaluate this new test which has a potential to be offered as a population-based prenatal screening for single gene Mendelian diseases.

3250W

Microarrays in prenatal diagnosis – some interesting cases during validation using the Affymetrix CytoScan 750K SNP chip. *L. Gole[1], S. Chan[1], C. Chua[1], S.P. Chua[1], L.L. Su[2], A. Biswas[2].* 1) Department of Laboratory Medicine, National University Hospital, Singapore; 2) Department of Obstetrics and Gynaecology, National University Hospital, Singapore.

Chromosomal microarray analysis has emerged as a primary diagnostic tool for the evaluation of developmental delay and structural malformations in children. For prenatal diagnosis, this technology has not been accepted very rapidly for fear of obtaining excessive information. We were apprehensive as we started our validations for this application, but were surprised by the results we obtained, which actually helped decision making for parents. We would like to share our experiences in this area by presenting a few cases. Only patients with high risk ultrasound findings were included in this study. DNA from cultured amniotic fluid was run on the Affymetrix CytoScan 750K SNP chip, with karyotyping being done simultaneously. Analysis was done using the Affymetrix ChAS software with the following cutoffs. DNA copy number loss of ≥ 1Mb or gain of ≥ 2Mb with UPD of ≥ 15Mb interstitially or ≥ 10Mb telomerically. Here we report 6 interesting cases where microarray provided additional and helpful information to the patients.

No	Clinical indication	Karyotype	Microarray karyotype	Results
1	Abnormal ultrasound findings A=18	46,XY,del(5)(p13p15.1).ish 5p15.2(D5S23,D5S721)x2	arr[hg19] 5p14.3p13.2(21,057,088-37,749,526)x1	Cri-du Chat not involved, but *NIPBL* deleted, known to cause Cornelia De Lange syndrome 1
2	Fetus with cardiac echogenic foci and echogenic bowels. A=22+6	47,XY,+mar. ish(15q11.2q13)(D15S11-,D15Z1-)	arr[hg19](1-22)x2,(XY)x1	Marker had no significant genomic content
3	Missed abortion	46,XX,add(6)(p24)	arr[hg19] 3p26.3p24.2 (61,891-24,670,223)x3,6p25.3p25.1 (222,200-5,277,366)x1	Microarray refined the abnormality as dup(3p) and del(6p)
4	Dandy Walker A=23+1	46,XX	arr[hg19] 4p16.3p16.1(2,931,120-6,494,394)x1	Del 4p results in Weyers acrofacial dysostosis or Ellis-van-Creveld syndrome
5	Fetus showed symmetrical IUGR A=21+3	46,XX,del(13)(q12q14)	arr[hg19] 13q12.3q14.13(29,116,876-47,290,217)x1	Does not involve *RB1*, but results in development delay and other significant developmental or morphological phenotypes. *BRCA2* also deleted
6	Twins. Lower fetus with IUGR, CVS,VSD A=19+3	46,XX,der(21) ins(21;?)(q11.2;?)	arr[hg19](1-22,X)x2	Surprisingly normal.

In conclusion, the microarray testing in prenatal samples of cultured amniotic fluid was very useful to parents in decision making. As long as genetic counselling is done clearly and clear guidelines are set with regards to whether the parents would like to know incidental findings involving (a) cancer genes e.g. *BRCA 1* and *2*, and (b) Infertility.

3251T

Frequency of prenatal diagnosis of most common chromosomal abnormalities in 289,076 births in Bogota and Cali, Colombia: 2011-2015. *P. Hurtado[1,2], S. Victoria Mojica[2], A. Puerto Niño[3], G. Gracia[4], I. Zarante[3].* 1) Ciencias Basicas de la Salud, Pontificia Univ Javeriana, Cali, Colombia; 2) Pontificia Universidad Javeriana Cali, Congenital Birth Defects Surveillance Programme of Cali; 3) Pontificia Universidad Javeriana Bogota, Congenital Malformations Surveillance Programme of Bogota; 4) Secretaria de Salud de Bogota, Congenital Malformations Surveillance Programme of Bogota.

Introduction: Chromosomal abnormalities (CA) have been reported in ~50% spontaneous abortion. CA is expected in ~10% of children. In Colombia, the CA rate per 10.000 births, for 1998 to 2000 years, was 3.52 for trisomy 18 (T18), 1.76 for trisomy 13 (T13) and 10.21 for trisomy 21 (T21). Nowadays exists primary, secondary and terciary prevention strategies for CA. Currently, the diagnosis based on ultrasound, biochemical and cytogenetic tests increases the likelihood of diagnosing CA during and after pregnancy. The gold standard, karyotyping, requires amniocentesis or chorionic villus sampling, which are invasive, and can cause abortions and anxiety in parents. Other new techniques, as non invasive prenatal diagnosis, are not are not available for the entire population.The aim of this study was to determine the frequency of prenatal diagnosis of most common CA (autosomal) in Bogota and Cali between 2011-2015.**Materials and Methods:** An observational study, based cohort 289,076 births in 9 hospitals from Bogota and Cali, Colombia, using Latin American Collaborative Study of Congenital Malformations ECLAMC methodology. Cases and controls were analysed looking at prenatal diagnosis of chromosomal abnormalities.**Results**: We found 409 patients with clinical syndromes of T21, T18 and T13. Suggestive ultrasound findings were found in 23.9% and confirmation of diagnosis with cytogenetic prenatal studies was done in 7.5%. Low birth weight was detected in 45.2% of patients. The diagnosis was most frequent in males with 54.2%. The frequency of T21 was 12.8 per 10,000 births (CI: 11.56-14.21), of whom suggestive prenatal ultrasound was 22% (n=83) and 5.9% cytogenetic prenatal studies (n=22). The frequency of T18 was 0.83 per 10,000 births (CI: 0.53-1.24), of whom sugestive prenatal ultrasound was 39.1% (n=9) and 30.4% cytogenetic prenatal studies (n=7). The frequency of T13 was 0.48 per 10,000 births (CI: 0.26-0.81), of whom sugestive prenatal ultrasound was 42.8% (n=6) and 14.2% cytogenetic prenatal studies (n=2).**Conclusions:** Prenatal diagnosis is essential to provide timely care and interdisciplinary approach to the patient and determine the level of care complexity for the delivery. Unfortunately, prenatal and cytogenetic diagnosis still is far below expectations and continues using the clinical diagnosis. The challenge is can use minimally invasive test with low morbidity and to make a timely diagnosis to improve the quality of life.

3252F

Diagnostic utility of chromosomal microarray analysis in neonates: A comprehensive multi-year study. *T. Sahoo, M.N. Strecker, S.B. Commander, M.K. Travis, N. Dzidic, K. Hovanes.* CombiMatrix, Irvine, CA.

   INTRODUCTION: Genomic imbalances are a frequent cause of congenital anomalies in neonates, and result in considerable morbidity and mortality. Although multiple, large studies have confirmed a major role for chromosomal microarray analysis (CMA) in pediatric evaluation, much less data exists specifically for the neonatal population. We undertook a comprehensive genotype-phenotype correlation analysis of all neonatal patients (≤29 days old) referred to our laboratory for CMA. **METHODS**: This study included 1351 patients over a 4 year period who were evaluated by oligonucleotide-aCGH (N=386) or single nucleotide polymorphism (N=965) array. The reasons for referral and any additional clinical data were codified into phenotypic categories and enumerated. **RESULTS**: Overall, clinically significant abnormalities were identified in 245/1351 patients (18%). Whole chromosome aneuploidy was identified in 80 cases (6%), segmental abnormalities in 163 cases (12%), and uniparental disomy (UPD) of an imprinted chromosome in 2 cases (0.1%). Variants of uncertain clinical significance (VOUS) were identified in 120 cases (9%), and multiple regions of homozygosity were reported in 46 cases (3%). In terms of genotype-phenotype correlations, 45% of aneuploidy cases were referred for a suspected chromosome abnormality, but only 20 cases (25%) correctly specified the aneuploidy of concern based on clinical features. Of the 165 abnormal cases (segmental abnormalities or UPD), the most frequent clinical indications included cardiac defects/suspected DiGeorge syndrome (N=50, 30%), suspected chromosome abnormality (N=39, 24%), central nervous system anomalies (N=32, 19%), multiple congenital anomalies (N=29, 18%) and dysmorphic features (N=26, 16%). Many patients had broader referral indications that overlapped two or more of the above categories. There were 13 cases in which the abnormality was mosaic (aneuploidy=6; segmental abnormality=6; UPD=1). We identified 74 cases with non-mosaic aneuploidies and 49 cases with segmental aneuploidies >10 Mb that may have been amenable to detection by karyotype. **CONCLUSIONS**: Our data identifies common referring indications for CMA in neonates and demonstrates the immense value of CMA in identifying the broad spectrum of associated genomic imbalances. Precise and rapid molecular diagnosis by CMA enables appropriate management, making CMA an optimal diagnostic tool for evaluating newborns with both syndromic and non-syndromic birth defects.

3253W

Pregnancy and neonatal outcome associated with genome-wide paternal uniparental disomy. *L. Wallace[1], E. McCready[2], M. Huggins[3], S. Winsor[4], S. McDonald[5], C. Li[6].* 1) Department of Obstetrics and Gynecology, Hamilton Health Sciences, Hamilton, Canada; 2) Department of Laboratory Medicine, McMaster University; 3) Obstetrics and Gynecology, Hamilton Health Sciences, Hamilton, Canada; 4) Obstetrics and Gynecology, McMaster University, Canada; 5) Department of Obstetrics and Gynecology, McMaster University, Canada; 6) Pediatrics, McMaster University, Canada.

   Mosaic genome-wide paternal uniparental disomy (GWpUPD) is rare, and associated clinical features are variable, presumably due to the proportion and tissue distribution of the normal and abnormal cells. We report a case that was referred to our center for an abnormally large and cystic placenta. Ultrasound at 14 weeks of gestation revealed a thick, multi-cystic, and ball-like placenta and an intact abdominal wall. The differential diagnosis included partial molar pregnancy, mosaic triploidy and mesenchymal dysplasia. Repeat scans at 16, 18 and 20 weeks showed persistently thick placenta with multiple small, non-vascular cysts throughout. Color-flow Doppler assessment was not consistent with a molar pregnancy. Uterine artery Doppler studies showed bilateral notching at 18 weeks. Amniotic fluid levels were high normal to borderline polyhydramnios. Fetal growth was appropriate for dates and gross structural fetal anomalies were not seen. Amniocentesis was done at 15 weeks of gestation. SNP oligonucleotide microarray revealed abnormal allele peak patterns, suggestive of two possibilities: mosaic GWpUPD (mos arr(1-22,X)X2 hmz pat) or mosaic triploidy (arr(1-22,X)X2~3). However QFPCR was consistent with a normal disomic complement for chromosomes 13, 18, and 21 in a female fetus, no triploid cells were seen. Testing for UPD 11p15 on cultured amniocytes was inconclusive. Paternal UPD of specific chromosomes or regions is known to be associated with a number of recognizable genetic conditions: paternal UPD6 with neonatal transient diabetes; UPD11 with Beckwith-Wiedemann syndrome; UPD 14 with Kagami-Ogata syndrome; UPD15 with Angelman syndrome; and UPD 20 with hypoparathyroidism. GWpUPD is associated with variable phenotypes presenting with complete or partial features of some or all of the above imprinting disorders, a combination of them or a normal outcome, presumably reflecting the proportion and tissue distribution of the abnormal cells. The pregnancy is ongoing at 28 weeks and we await to assess the baby postnatally. Our case offers a unique opportunity to gain further insight into the phenotypic spectrum of prenatally detected mosaic GWpUPD.

3254T

Genomic abnormalities in products of conception in the CMA era. *H. Yonath[1,3], D. Bar-Avin Dayan[2], J. Shamash[1], E. Guetta[1], H. Carp[2,3], D. Seidman[2,3], E. Pras[1,3], S. Reinstein[1].* 1) Inst Human Gen, Sheba Med Ctr, Ramat Gan, Israel; 2) Gynecology and Maternity Center, Sheba Med Ctr, Ramat Gan, Israel; 3) Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

BACKGROUD Approximately 15% of clinically confirmed pregnancies abort spontaneously in the first trimester and about 1.5% of pregnancies have severe malformations. About half of these early abortions can be explained by chromosomal abnormalities. Chromosomal microarray analysis (CMA) is transforming clinical cytogenetic practice and might have better success and detection rates. AIM: To demonstrate the usefulness and robustness of an array based method for the elucidation of genetic causes underlying pregnancy loss and termination of pregnancies for medical reasons. METHODS: 117 samples: from RPL (recurrent pregnancy loss) and TOP (termination of pregnancies) specimens underwent cytogenomic analysis and/or chromosomal microarray analysis. Chromosomal microarray analysis was performed using a oligonucleotide-based microarray platform. Results were evaluated at the cytogenetic and microscopic (greater than 10 Mb) and submicroscopic (between 3Mb to 10 Mb) levels. Maternal cell contamination was assessed in each fetal sample. RESULTS: Maternal cell contamination was identified in 5% of specimens. The remaining specimens were considered to be of true fetal origin. A result by CMA can be obtained from 6 weeks of gestation (range 6-35) A result was obtained in roughly 96% of RPL-CMA group, significantly higher than in the RPL-KARYOTYPE group (63%) (p=0.000173) Chromosomal aberrations were detected by CMA analysis in 28.8% of the RPL cases, and 38.5% of the TOP cases. Of these, 53% showed classical cytogenetic abnormalities (most of them in the RPL group) In 15.6% of cases considered normal at the cytogenetic level, chromosomal microarray analysis revealed a pathogenic copy number change (5.1% of the RPL cases and 26.3% of the TOP cases ,p=0.009221) . CONCLUSIONS: Chromosomal microarray analysis of products-of-conception specimens yields a high diagnostic rate. Moreover, compared with karyotyping, there appears to be an increased detection rate of chromosomal abnormalities when CMA is used to analyze the products of conception especially in the group of TOP. Identification of these abnormalities helps to estimate recurrence risks in future pregnancies.

3255F

Case Report: Prenatal diagnosis of X-linked Optiz G/BBB Syndrome by chromosomal microarray in a male fetus presenting with fetal hydrops, increased nuchal fold and absent nasal bone. *N. Martin[1], G. Nimmo[2], N. Watkins[3], J. Keunen[4], E. Kolomietz[3], M. Roifman[1,2].* 1) Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, Toronto, Ontario, Canada; 2) Division of Clinical & Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, Ontario, Canada.

X-linked Opitz G/BBB syndrome (XLOS) is a multiple congenital anomaly disorder typically characterized by ocular hypertelorism, laryngotracheoesophageal defects, and genitourinary abnormalities caused by mutations in the *MID1* gene. Prenatal presentations of XLOS are rare. We report a G4P1A2L1 29 year-old woman who presented at 19 weeks gestation with fetal ultrasound findings of bilateral pleural effusions requiring chest shunts in utero. The fetus also had an absent nasal bone, thickened nuchal fold and ascites, the latter of which resolved following fetal intervention. Amniocentesis was performed and microarray analysis revealed a novel likely pathogenic *de novo* 126 kb deletion encompassing the 3' region of the *MID1* gene in a male fetus. Deletions in this region have not been reported in the literature to be associated with XLOS. Postnatal examination of the newborn was consistent with multiple features of XLOS: severe hypertelorism, widow's peak, prominent forehead, and right cryptorchidism. One prior case of prenatal bilateral pleural effusion in a child with presumed XLOS has been reported; however, no molecular diagnosis was made. The current case confirms the finding of bilateral pleural effusions with or without hydrops as a presenting feature of XLOS and highlights the importance of considering a targeted fetal ultrasound for features of *MID1* and/ or *MID1* variant analysis if prenatal chromosome analysis is normal.

3256W

Prenatal CMA findings in urinary system malformations fetuses with normal karyotype. *H. Wang[1,2], Z. Dong[1,2], M. Chen[3], L. Xiong[4,5], Y. Song[3], J. Lou[6], Y. Kwok[1], K. Choy[1,2].* 1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong , Hong Kong, Hong Kong; 2) Shenzhen Research Institute, The Chinese University of Hong Kong; 3) Key Laboratory of Reproductive Medicine of Guangdong Province, Key Laboratory for Major Obstetric Diseases of Guangdong Province, Key Laboratory of Reproduction and Genetics of Guangdong Higher Education Institutes. The Third Affiliated Hospital of Guangz; 4) Bao An Maternal and Child health hospital, Shenzhen, Guangdong, China; 5) Shenzhen Key Laboratory of Birth Defects (Number:ZDSYS201504301707152); 6) Dong Guang Maternal and Child health hospital, DongGuan, Guangdong, China.

　Background: Fetal urinary system malformation, accounts for one third of all congenital malformations detected by routine fetal ultrasonography. Also known as congenital anomalies of kidney and urinary tract (CAKUT) are observed in 0.3 to 0.6% of live births and accounts for 40 to 50% of the etiology of chronic kidney disease in children. Chromosomal microarray analysis (CMA) has been proven to detect submicroscropic genomic imbalances prenatally because of its high-resolution and time-saving. However, prenatal studies on isolated or syndromic CAKUT and its advantage over traditional cytogenetics are limited. In this study, we aimed to investigate the diagnostic performance of CAKUT among prenatal cases with normal karyotype. **Methods:** This study recruited 108 prenatal samples, with detailed documented ultrasound findings and CMA results from four prenatal diagnosis centers. Pathogenic copy-number variants (CNVs) were identified based on the American College of Medical Genetics guideline. **Results:** The overall detection rate was 9.2% (10/108) for pathogenic CNVs by CMA. The extra diagnostic yield compared with karyotype was 1.7% (1/60) for the isolated CAKUT, and 18.7% (9/48) for the syndromic group. The incidence of pathogenic CNV is significantly different (p<0.05) between the isolated group and syndromic group. **Conclusion:** This study shows the improved diagnostic yield by CMA in prenatal CAKUT, compared with conventional karyotyping, with or without syndromic anomalies. Our data supports that CMA should be provided in prenatal diagnosis as a first-tier test for fetus with urinary system malformations. The pathogenic CNVs identified help to provide further understanding of the pathogenesis of the urinary system malformations and CNV spectrum.

3257T

Differentially expressed miRNAs in trisomy-21 placenta. *I. Svobodova[1], M. Korabecna[1], P. Calda[2], M. Brestak[3], E. Pazourkova[1], S. Pospisilova[1], M. Krkavcova[4], M. Novotna[2], A. Horinek[1].* 1) Institute of Biology and Medical Genetics, General University Hospital in Prague, Prague, Czech Republic; 2) Department of Obstetrics and Gynecology of the First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; 3) Screening center, Prague, Czech Republic; 4) GENVIA genetic laboratories, Prague, Czech Republic.

　Background: The Down syndrome (DS) etiology is still unknown. However, all data suggest that the mechanism is more complex than previously supposed. Epigenetic mechanisms, including miRNAs gene expression regulation, are one of potential influencing factors. The aim of this study was to compare miRNAs expressions in placentas with normal and trisomic karyotype and to associate differentially expressed miRNAs with concrete biological pathways.　Materials and Methods: In the pilot study, 754 miRNAs were profiled in 30 chorionic villi samples (CVS) - 14 normal and 16 DS, using real-time PCR technology and TaqMan Human miRNA Array Cards A and B. Twelve differently expressed miRNAs between compared groups of samples were validated on 50 independent samples (25 normal and 25 DS) using TaqMan miRNA Assays. The Mann-Whitney test with a cutoff p-value < 0.05 was used for results evaluation. Functional annotation and diseases association of selected miRNAs were performed using several web-based software tools (miRWalk, DAVID, WebGestalt).　Results: Seven miRNAs were verified as upregulated in DS placentas; three of these miRNAs are located on chromosome 21 (miR-99a, miR-125b and let-7c). Many essential biological processes, transcriptional regulation or apoptosis, were identified as being potentially influenced by altered miRNA levels. Moreover, miRNAs overexpressed in DS placenta potentially regulates genes involved in placenta development (GJA1, CDH11, EGF, ERVW-1, ERVFRD-1, LEP or INHA).　Conclusion: These findings suggest the possible participation of miRNAs in Down syndrome impaired placentation and connected pregnancy pathologies.　Supported by the Ministry of Health of the Czech Republic RVO VFN64165.

3258F

Molecular and functional characterization of a novel mouse mutant of the Planar Cell Polarity gene, *Vangl2. A.R. El-Hassan[1,2], F. Kharfallah[1,2], M.C. Guyot[2], Z. Kibar[1,2].* 1) University of Montreal, Montréal, Quebec, Canada; 2) Chu Saint-Justine, Montréal, Quebec, Canada.

Vangl2 is a core planar cell polarity (PCP) gene that acts in the non-canonical Wnt signaling pathway. The latter is known to be implicated in the regulation of morphogenetic processes of convergent extension during gastrulation and neural tube formation. Studies in mice and humans have demonstrated an important role of the PCP pathway in neural tube defects (NTDs). The mouse mutant of *Vangl2*, *Looptail* (*Lp*), has been used as a model for molecular studies of NTDs in humans. Homozygous *Lp/Lp* embryos suffer from a severe type of NTDs called craniorachischisis where the neural tube remains open along the entire body axis. They also show a PCP defect manifested by misorientation of hair cells in the cochlea of the inner ear. In the present study, we have identified a novel mutant allele of *Vangl2* called the curlybob mouse (Crb). A total of 40% of heterozygotes show a curly tail phenotype and 90% of homozygotes suffer from craniorachischisis. Around 17% of heterozygous females *Crb/+* have an imperforate vagina as compared to 50% in *Lp/+* females. Crb genetically interacts with Lp where 93% of double mutants *Crb/Lp* have craniorachischisis. The orientation of inner ear ciliated cells is more disturbed in *Crb/Lp* (double mutants) than in single Crb/+ or Lp/+ mutants. This novel mouse mutant represents another powerful model to dissect the role of *Vangl2* in PCP signaling in the inner ear and neural tube formation.

3259W

Implementation of non-invasive prenatal testing by semiconductor sequencing in a genetic laboratory. *A. Dheedene[1], T. Sante[1], M. De Smet[1], J.F. Vanbellinghen[2], B. Grisart[3], S. Vergult[1], S. Janssens[1], B. Menten[1].* 1) Center for Medical Genetics Ghent, Ghent University, Ghent University Hospital, Ghent, Belgium; 2) Plateforme de Biologie Moléculaire, Département des Laboratoires, Cliniques Universitaires Saint-Luc, Bruxelles, Belgium; 3) Centre de Génétique Humaine, Institut de Pathologie et Génétique, Charleroi, Belgium.

OBJECTIVES: To implement non-invasive prenatal testing (NIPT) for fetal aneuploidies with semiconductor sequencing in an academic cytogenomic laboratory and to evaluate the first 15 months experience on clinical samples. METHODS: We validated a NIPT protocol for cell-free fetal DNA sequencing from maternal plasma for the detection of trisomy 13, 18 and 21 on a semiconductor sequencing instrument. Fetal DNA fraction calculation for all samples and several quality parameters were implemented in the workflow. 1081 clinical NIPT samples were analyzed following the described protocol. RESULTS: NIPT was successfully implemented and validated on 201 normal and 74 aneuploid samples. For trisomy 21, 18 and 13 a specificity of 100%, and a sensitivity of respectively 98%, 94% and 100% was reached and accreditation for the NIPT workflow was obtained. From 1081 clinical samples, 17 samples showed an abnormal result: fourteen trisomy 21 samples, one trisomy 18 and one trisomy 16 were detected. Also a maternal CNV on chromosome 13 was observed, which could potentially lead to a false positive trisomy 13 result, but was reported as normal. One sex discordant result was reported, probably possibly attributable to a vanishing twin. No false-negative results have been reported to date. Moreover, our combined fetal fraction calculation enabled a more reliable risk estimate for trisomy 13, 18 and 21. CONCLUSIONS: NIPT for trisomy 21, 18 and 13 detection has a very high specificity and sensitivity. Due to several biological phenomena diagnostic invasive confirmation of abnormal results remains required.

3260T

Assessing the mutational spectrum of 7-dehydrocholesterol reductase and the toxicological effects of pharmacological inhibition during the prenatal period. *M.R. Boland[1,2,3,4], N.P. Tatonetti[1,2,3,4].* 1) Biomedical Informatics, Columbia University, New York, NY; 2) Systems Biology, Columbia University, New York, NY; 3) Medicine, Columbia University, New York, NY; 4) Observational Health Data Sciences and Informatics (OHDSI), Columbia University, New York, NY.

Seasonal factors occurring during the prenatal or perinatal period can affect long-term disease outcomes. Previously, we used clinical data from 1.7 million patients to identify 55 diseases correlated with birth season. However, this work did not reveal the environmental drivers underlying these associations. Vitamin D is a seasonally varying compound synthesized in the skin using 7-dehydrocholesterol and ultraviolet B radiation. A competing reaction occurs to convert 7-dehydrocholesterol to cholesterol via 7-dehydrocholesterol reductase (DHCR7). Deleterious mutations in *DHCR7* decrease the ability to produce cholesterol and enhance production of vitamin D. Thus making *DHCR7* mutations more evolutionarily favored in regions of the world with historically low sunlight access. Importantly, a rare Mendelian disease - Smith-Lemli-Opitz Syndrome (SLOS [MIM 270400]) is characterized by compound heterozygous mutations in *DHCR7*. SLOS results in severe fetal deformities and malformations. We discuss the toxicological information gleaned from a deep exploration of *DHCR7*. First, we present a compilation of SLOS-inducing *DHCR7* mutations and the geographic distribution (S. America, Europe, Australia, Asia) of those mutations among diseased populations. We describe several hypotheses for *DHCR7* mutations that would maximize vitamin D production via increased evolutionary pressure. Next, we looked at the mutational spectrum of *DHCR7* in an ethnically diverse, presumed healthy population from ExAC (http://exac.broadinstitute.org/). We observed that several mutations thought to be disease causing occur in healthy populations as well, sometimes with high frequencies, indicating an incomplete understanding of SLOS and highlighting new research opportunities. We also highlight several *DHCR7* variants found in ExAC that could represent SLOS carriers in under-represented ethnic groups. Our knowledge of the importance of vitamin D during the prenatal period coupled with our enhanced global understanding of the *DHCR7* mutational spectrum allowed us to hypothesize that exposure to DHCR7 inhibitors would result in deleterious effects similar to SLOS. We tested our hypothesis by investigating the fetal outcomes following prenatal exposure to DHCR7 modulators. First-trimester exposure to DHCR7 inhibitors resulted in outcomes similar to those of known teratogens (50% vs. 48% born-healthy). DHCR7 activity should be considered during drug development and prenatal toxicity assessment.

3261F

Dynamics of mitochondrial DNA heteroplasmy in human multi-generation pedigrees. *M. Su[1], N. Stoler[2], B. Rebolledo-Jaramillo[2], B. Dickins[3], D. Blankenberg[2], M. Madhava[1], F. Chiaromonte[5], R. Nielsen[4], I. Paul[6], A. Nekrutenko[2], K. Makova[1].* 1) Department of Biology, Pennsylvania State University, State College, PA; 2) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 3) School of Science and Technology, Nottingham Trent University, Nottingham NG1 4BU, United Kingdom; 4) Department of Integrative Biology, University of California, Berkeley, CA; 5) Department of Statistics, Pennsylvania State University, University Park, PA; 6) Department of Pediatrics, College of Medicine, Pennsylvania State University, Hershey, PA.

Mitochondrial DNA (mtDNA) accumulates de novo mutations causing heteroplasmy, the presence of more than one mtDNA variant within an individual. Studies of heteroplasmy in healthy individuals have shed light on understanding the dynamics of changes in heteroplasmy frequency from generation to generation. Such changes are governed by mtDNA bottleneck, a reduction in the number of mtDNAs during oogenesis. The variation of accumulated heteroplasmic mtDNA was found to correlate with the age of a mother at the time of fertilization in our previous study. To elucidate the dynamics of maternal transmission of human mtDNA and to further explore the effect of maternal age on mitochondrial mutations, we analyzed patterns of mtDNA heteroplasmic mutations in 85 two-generation human pedigrees with multiple siblings. We performed whole mtDNA genome sequencing on blood and buccal epithelial genomic DNA from two or more offspring from each pedigree and sequenced 613 mtDNA genomes in total. The focus of this study is on heteroplasmy changes between tissues, generations, pedigrees, and age groups. The association of mitochondrial heteroplasmy with mitochondrial haplogroups, the maternal age effect on the accumulation of mtDNA mutations, and the evolutionary mutation spectrum of human mtDNA in germline and somatic tissues will be discussed.

3262W

Interphase Chromosome Profiling (ICP) as a rapid, sensitive and cost-effective diagnostic tool for Amniotic Fluid (AF) and Products of Conception (POC) testing. *S.K. Bhattacharya[1], V.R. Babu[2], V. Lal[1].* 1) Cytogenetics, Dr. Lal Path Labs. Pvt. Ltd., Newdelhi, Delhi, India; 2) InteGen LLC, Orlando, FL 32819, USA.

Karyotype determination has an important role in the genetic work-up of POC specimens, since approximately one half of miscarriages are due to chromosomal imbalances. The three primary methods used to obtain karyotype are 1) Classical cytogenetics 2) Targeted FISH and 3) aCGH . Each of these methods has its advantages and disadvantages. While the classical approach covers the whole genome, the culture success rate is very low and as many as 30% of the cultures fail to produce results. Additionally, even in a successful culture, the TAT is significantly long. The targeted FISH, as the name indicates, only covers 5-7 chromosomes and therefore provides incomplete information. It also cannot detect any structural abnormalities. Microarray is highly sensitive in detecting most of the abnormalities, but the limitations associated with this technology are 1) inability to detect balanced translocations including Robertsonian translocations 2) inconsistent ploidy detection and 3) problem with low level mosaicism. Prenatal diagnosis by karyotype determination is done mostly to provide assurance, since majority of the pregnancies would have a normal karyotype. Therefore, fast and accurate information is highly critical for management of the pregnancy. However, the same limitations mentioned for POC also apply to prenatal diagnosis. To overcome these challenges, we recently validated and adapted a novel cytogenetic technology 'Interphase Chromosome Profiling (ICP) (InteGen LLC, USA) to assess the molecular karyotype of 200 miscarriage material and 80 amniotic fluid samples using interphase nuclei. Based on the clinical situation, ICP probes can be used to get high (550 band) or standard resolution. For POC and AF samples, using probes targeting sub-telomeres and centromeres, all numerical, most balanced and unbalanced structural aberrations, and all Robertsonian translocations can be detected. Using ICP, we obtained results from all (100%) samples and the TAT was significantly reduced to less than 7 days. However, with a proper workflow, results can be delivered in less than 48 hours.

3263T

Accurate detection of segmental aneuploidy in preimplantation genetic screening using targeted next-generation DNA sequencing. *M. Umbarger, K. Germain, A. Gore, B. Breton, E. Tsung, L. Walters-Sen, T. Mullen, N. Faulkner.* Good Start Genetics, Cambridge, MA.

Preimplantation genetic screening (PGS) aims to identify embryos that are chromosomally abnormal and thereby have limited reproductive potential. We have previously described a targeted next-generation DNA sequencing (NGS)-based method for PGS that has been validated for accurate detection of full chromosome aneuploidy, the most common form of chromosomal abnormality in preimplantation embryos. This method leverages FAST-SeqS technology to amplify and sequence tens of thousands of loci across the genome and therefore has the potential to detect sub-chromosomal or segmental aneuploidy. Here, we sought to characterize the performance of our targeted NGS PGS strategy with respect to segmental aneuploidy. First, we tested samples derived from 138 cell lines with previously reported segmental duplication/deletions ranging in size from 0.1 to ~80 Mb as well as 170 known normal samples and compared the resulting calls to the reported karyotypes, so as to assess the segmental calling sensitivity and specificity. We found that 112 of the 115 segmental copy number variants (CNVs) ≥10 Mb were called and that a total of 6 samples were falsely classified as aneuploid, yielding sensitivity and specificity estimates of 0.974 (0.926,0.991) and 0.965 (0.925,0.984), respectively. Next, to evaluate performance with respect to CNVs falling in parts of the genome for which samples were not available, we implemented a simulation strategy that enables genome-wide assessment of segmental calling sensitivity. Using data derived from trophectoderm biopsies from embryos that were classified as normal on our targeted NGS and microarray comparative genomic hybridization PGS platforms, we simulated segmental duplications of varying sizes across the genome and evaluated the frequency at which these duplications were called. These simulations revealed while there are positions in the genome where segmental calling sensitivity is reduced, these regions often coincide with heterochromatic and/or repetitive regions that are likely problematic with most PGS technologies. Collectively, the data suggest that our targeted NGS-based PGS approach is capable of detecting sub-chromosomal changes of ≥10 Mb with a sensitivity of >97% and a specificity that is comparable to that noted in our previous studies. In addition, our simulation methodology enables a level of comprehensiveness of sensitivity assessment that is difficult to achieve using samples.

3264F

Study the high incidence and mechanism of chromosome 22 aneuploidy in the embryos of male Robertsonian translocation carriers. *W. Wei, D. Chen Yue, S. Jian Dong, G. Chao, X. Jiazhi, L. Jiayin.* Reproductive Medicine, Jiangsu Provience Hosipital, Nanjing,Jiangsu,China, Jiangsu, China.

Background Approximately 2% to 3% of infertile patients are Robertsonian translocation carriers. Robertsonian translocation carriers are at high risk for infertility, spontaneous abortions, or offspring with unbalanced chromosome. **Objective** We try to find out the mechanism of the embryos from biparental robertsonian and reciprocal translocation carriers which emerge the aneuploidy. We will combine next generation sequencing with biological methods to discover the mechanism of genetics and better understand the influence of biparental translocation carriers on embryos. **Method** Analyzed the data from the PGD patients in reproductive center There were 30 maternal robertsonian translcation carriers, 49 paternal robertsonian translcation carriers,64 maternalbalanced translocation carriers, 104 paternal balanced translocation carriers, in total 247 patients. We choose two stages of embryos include blastomere and trophectoderm by sequencing. We use the two kinds of next generation sequencing machines to get the datas, Hiseq2000 and Ion torrent PGM. **Results** We analyzed the datas from PGD patients in reproductive center, the results show that the incidence of embryos from paternal robertsonian translocation carriers with chromosome 22 aneuploidy (70%) is higher than other three groups. The NGS results of chromosome 22 aneuploidy embryos from paternal translocation show that the mutation of gene CECR2, MED15 and POM121L4P may related to the high incidence of chromosome 22 aneuploidy of play a key role in it. **Conclusion** It is common that the embryos from robertsonian and reciprocal translocation carriers are aneuploidy. Our research firstly reports that the rate of chromosome 22 aneuploid embryos from paternal robertsonian translocation carriers is 70%, which is higher than other three groups. We found that CECR2, MED15 and POM121L4P are related to chromosome 22 aneuploidy. All of these supply a proof to deep understand the mechanism of aneuploid embryos.   .

3265W

Novel autosomal genes linked with male infertility. *D.V.S. Sudhakar[1], R. Phanindranath[1], N.J Gupta[2], M. Deenadayal[3], R. Singh[4], J. Reshma Devi[1], S. Yogendra[1], K. Thangaraj[1].* 1) Centre for Cellular and Molecular Biology, Hyderabad, India; 2) Institute of infertility and reproductive medicine, Kolkata, India; 3) Infertility institute and research center,Hyderabad, India; 4) Central drug research institute, Lucknow, India.

Introduction: Approximately 15% of the couples are infertile worldwide. Impaired fertility of male partner is causative in approximately 50% of all couples unable to conceive spontaneously. Earlier studies from our lab have shown that about 8.5% infertility among Indian men is due to the Y chromosome microdeletions. Further, analysis of several autosomal genes and mitochondrial genome accounted for additional 20.5% of the genetic factors responsible for infertility among Indian men. However, etiology of large proportion (74%) of infertile men still remained unknown. Our aim was to understand the role of novel autosomal genes in male infertility. **Materials and methods:** we sequenced exome of 44 idiopathic infertile men using Illumina Hiseq -2000 platform at 100X coverage. Raw data (.fastq files) were trimmed, mapped, aligned and further variants were called using GATK unified genotyper. Genotyping of the identified variants in more than 1400 samples (cases and controls) was carried out using iPLEX SNP Genotyping analysis on a Sequenom MassARRAY System. *CETN1* gene is one among the candidates identified from exome sequencing. We have sequenced the complete *CETN1* gene using peripheral blood DNA from 875 infertile men and 552 ethnically matched fertile controls. Functional studies were carried out using biophysical and cell biology techniques to understand the affect of *CETN1* variants on male infertility. **Results:** After extensive analysis of exome data we identified 32 novel and rare variants from 30 genes, that were mostly, stop gain and missense. In addition, CETN1 gene was selected from the exome data and the entire gene was sequenced; 5'UTR variant (rs 367716858) and missense (rs61734344) variant from CETN1 gene were found to be strongly associated with male infertility *(P_{Corr}<0.005)*, *(P_{Corr}=0.001)* respectively. Biophysical studies have shown that missense mutation (rs61734344) affects the calcium binding affinity, thermodynamic properties and surface hydrophobicity of the mutant (p. Met72Thr) CETN1 protein. Functional analysis of 5' UTR variant g580393C>T (rs367716858) has shown that mutation leads to increased expression of CETN1 protein. **Conclusion:** Our study has identified novel autosomal genes that are linked with male infertility.

3266T

Single cell testing versus amniocentesis and NIPT: Ultra-high resolution preimplantation genetic screening for microduplication-microdeletion syndromes. *G. Altarescu[1], T. Dror[1], E. Elinor Hakam-Spector[1], D. Kort[1], E. Farhi[1], M. Ben Shlomo[1], Z. Zeligson[1], R. Beeri[1], P. Renbaum[1], T. Eldar-Geva[2], E. Levy-Lahad[1], D.A. Zeevi[1].* 1) Preimplantation Genetic Unit, Medical Genetics , Shaare Zedek Medical Center, Jerusalem, Israel; 2) IVF Unit, Shaare Zedek Medical Center, Jerusalem, Israel.

   Introduction: Invasive prenatal testing and recent advances in noninvasive prenatal testing (NIPT) have afforded pregnant couples the ability to test for some pathogenic microdeletions during pregnancy. Nonetheless, due to the small size of these deletions (ranging from 1Mb-3Mb long), preimplantation screening for high-risk de-novo microdeletions remains too technically challenging for large-scale application. Aim: To assess whether moderately low coverage next generation sequencing can be utilized to screen clinically relevant microdeletion/microduplication syndromes in preimplantation embryos.   Materials and Methods: Structural chromosomal aberrations, resulting from parental balanced translocations, were tested in 96 different blastomere/blastocyst biopsies by array comparative genomic hybridization (aCGH) and moderately low coverage whole genome shotgun sequencing. For sequencing data analysis, custom in-house software was established for the determination of ultra-high resolution chromosomal copy number across the entire human genome. This same software was applied for the detection of pathogenic microdeletions in 12 polar body samples from the Hou et al (Cell.2013.155(7):1492-506) deep sequencing data set.   Results: Deep sequencing identified all 84 unbalanced and 12 balanced/euploid embryo samples that were detected by aCGH (100% of the cases). Strikingly, in the Hou et al. data even CNVs as small as 0.3Mb and 0.5Mb were detected with high confidence.   Conclusions: We have expanded the scope of preimplantation genetic screening (PGS) to include genetic disorders that are currently screened only by amniocentesis or NIPT. Indeed, our deep sequencing workflow offers the requisite high resolution CNV detection which is crucial for preimplantation diagnosis of common microdeletion disorders such as DiGeorge, and Prader-Willi/Angelman syndromes. Moreover, unlike aCGH our method is rapid, robust, and cost-effective.

3267F

Prospective analysis of maternal duplications improves positive predictive value in prenatal cell-free DNA screening. *M.D. Maxwell[1], C.M. Strom[2], B. Anderson[2], R. Owen[2], K. Zhang[3], F.L. Lacbawan[2].* 1) Medical Affairs, Quest Diagnostics, San Juan Capistrano, CA; 2) Research and Development, Quest Diagnostics, San Juan Capistrano, CA; 3) Research Support, Quest Diagnostics, San Juan Capistrano, CA.

   Introduction: For trisomy 21, the positive predictive value (PPV) is >90% for cell-free DNA (cfDNA) screening. However, the PPVs for trisomies 18 and 13 are substantially lower. Lower PPVs for trisomies 18 and 13 have been attributed to their low prevalence, vanishing twins, and confined placental mosaicism (CPM). However, maternal duplications can also contribute to lower PPV. Here, we determined if identification of maternal duplications can improve PPV for trisomies 21, 18, and 13. **Methods**: The cohort included 38,668 patients who elected prenatal cfDNA screening for advanced maternal age, abnormal ultrasound finding(s), abnormal maternal serum screening, or personal or family history of chromosomal abnormality. When elevated Z-scores for chromosomes 21, 18, or 13 were observed, a karyogram was generated to prospectively analyze for maternal duplications. When maternal duplications were identified, these cases were reported as negative for the corresponding fetal trisomy with the maternal duplication reported as an incidental finding. PPVs were calculated with and without prospective analysis and compared. **Results**: Prospective analysis of suspected maternal duplications improved PPV for trisomy 21 by 3.9%, trisomy 18 by 18.4%, and trisomy 13 by 12.1%, resulting in PPVs of 97%, 90%, and 72% respectively. For chromosome 21, 15 of 389 (3.9%) cases with elevated Z-scores had maternal duplications. For chromosome 18, 26 of 141 (18.4%) cases with elevated Z-scores had maternal duplications. For chromosome 13, 21 of 107 (12.1%) cases with elevated Z-scores had maternal duplications. Subsequent chromosomal microarray analysis on maternal blood confirmed the suspected maternal duplication in all 17 cases for which follow-up testing was ordered: 2/2 for chromosome 21, 4/4 for chromosome 18, and 11/11 for chromosome 13. **Conclusions**: Identification of maternal duplications in our large cohort has allowed us to improve our PPV for trisomies 21, 18, and 13. Prospective analysis of suspected maternal duplications improved the overall PPV by 12.9%, with the more dramatic improvement in PPV for trisomies 18 and 13.

3268W

Clinical utilization of NGS in Preimplantation Genetic Diagnosis (PGD)/ Screening (PGS) for chromosomal rearrangements. *S. Madjunkova[1], R. Antes[1], V. Kuznyetsov[1], C. Librach[2,3,4,5].* 1) Department for Preimplantation Genetics, Create Fertility Centre, Toronto, Ontario, Canada; 2) Create Fertility Centre, Toronto, Ontario, Canada; 3) Department of Obstetrics and Gynecology, University of Toronto, Toronto, Canada; 4) Department of Physiology, University of Toronto, Toronto, Canada; 5) Department of Gynecology, Women's College Hospital, Toronto, Canada.

   Introduction: Data from preimplantation genetic screening (PGS) of human embryos has confirmed that the incidence of structural and numerical chromosomal rearrangements increases significantly with maternal age. Selection of euploid embryos using aCGH has improved implantation and pregnancy outcome rates in assisted reproduction (ART) programs and has offered a non-invasive therapeutic option for carriers of balanced rearrangements. We have previously validated the next-generation sequencing (NGS) VeriSeq™ PGS kit and showed its accuracy, reliability and high level of consistency with aCGH. The aim of this study was to prospectively evaluate the performance of NGS PGS/PGD for chromosomal rearrangements in a clinical environment. **Material & Methods:** This study had institutional REB approval. Trophoectoderm biopsy (2-5 cells) of D5/6 blastocysts was performed for PGS/PGD on 315 embryos (121 IVF cycles) from February to May 2016. PGD for chromosomal rearrangements was performed on 10 embryos from 4 couples (2 carriers of reciprocal translocations, 1 with a Robertsonian translocation and 1 with a pericentric inversion). The same whole genome amplification product from each embryo was assessed with both the VeriSeq™ PGS kit on the MiSeq system (Illumina) and the 24Sure™/plus™ aCGH kit (BlueGnome). Results from the NGS were analyzed blindly and then compared to the aCGH results.   **Results:** Parallel analyses of the aCGH and NGS results revealed 100% concordance between the two platforms when the cut-off for reporting was ≥10Mb (del/dup) and mosaicism ≥50%. PGS analysis detected 69 single chromosome aneuploidies (22.3%), 71 complex (≥2) chromosome aneuploidies (22.9%) and 34 segmental aberrations (11%). Unlike aCGH, the NGS platform was able to identify aberrations in the presence of ≤50% mosaicism in 17% of the embryos. No embryos identified as euploid by aCGH were found to be abnormal by NGS. However, NGS gave less inconclusive results (1.3%) than aCGH (2.2%). PGD was successful in all analyzed embryos resulting in 3 euploid/balanced embryos. The hands-on time and cost per sample were lower with NGS when 24 samples were analyzed in a same run.   **CONCLUSION:** NGS is a reliable and robust high-throughput method for human embryo PGS/ PGD. In a clinical setting with high volume of analyses, NGS outperformed aCGH by reducing the hands on time and costs, and demonstrated a greater precision and accuracy. Considering the advantages of the NGS methodology we are now employing it routinely in our current clinical workflow.

3269T

A change in fetal risk after prenatal diagnosis: A case of HbS/City of Hope. *B. Clark[1], B. Pullon[2], T. Senior[1], A. Sajoo[3].* 1) Kings College Hospital, London, United Kingdom; 2) Waikato Hospital, Hamilton, New Zealand; 3) Guy's and St Thomas' NHS Foundation Trust, London, UK.

   A 33 year old woman of Nigerian background attended the genetics clinic to discuss the diagnosis of sickle cell anemia (SS) in her toddler, who was said to have had a mild course. The recessive nature of the condition, coping with her child's diagnosis and reproductive options including preimplantation genetic diagnosis were discussed. In a subsequent pregnancy, the patient requested prenatal diagnosis and was referred by her General Practitioner to the Fetal Medicine Unit. A maternal but not paternal sample was sent with the CVS sample to the laboratory. Results: DNA analysis showed that the fetus was heterozygous for the c.20A>T; p.Glu7Val sickle mutation and the c.208G>A; p.Gly70Ser City of Hope mutation in the HBB gene. A Consultant Haematologist saw the parents and affected child and obtained a paternal sample was obtained. HPLC analysis identified the father to be a sickle carrier, with normal hemoatology. Beta globin sequencing confirmed he was a compound heterozygote with HbS/City of Hope. Discussion: City of Hope is electrophoretically silent (Rahbar, 1984). Carriers are expected to be asymptomatic with normal hematology parameters. However, a case report (Paradisi, 2010) of a deceased child of African background with HbS/City of Hope genotype, severe anemia, bone marrow hypoplasia and an underlying congenital immunodeficiency was identified. Hb City of Hope in combination with beta-zero thalassemia has also been reported as clinically significant in a case report by Vinciguerra et al. (2014). The PND result, father's genotype and case reports were discussed with the patient by the Hematologist. The patient opted to end her pregnancy. Summary: This patient's reproductive risks for sickle cell-related disorders were altered after PND and fetal genotyping, which would not have been indicated if a paternal sample had sent initially with the CVS sample. Due to the potential risk of interaction between HbS and Hb City of Hope, the patient ended her pregnancy despite having a phenotypically normal partner heterozygous for the same genotype. The patient has subsequently had a child who is a carrier of the City of Hope variant.

3270F

Identification and characterization of rare hemoglobin variants: Experience at good start genetics. *J. Wilkinson, M. Zhu, M. Sardella, K. Robinson, N. Faulkner. Good Start Genetics, Cambridge, MA.*

Hemoglobin variants in the *HBB* and *HBA1/HBA2* genes are found with varying frequencies across ethnic groups. Individuals of African, Southeast Asian, and Mediterranean descent are at increased risk to be carriers for hemoglobinopathies and ACOG recommends that they be offered carrier screening. The appropriate screening tests include a complete blood count and hemoglobin electrophoresis. Several common *HBB* variants with known clinical consequences [Hb S, Hb C, Hb E, and Hb D-Los Angeles (Hb D-Punjab)] are easily identified by their electrophoretic profiles and confirmed with molecular sequencing. However, rarer variants in the *HBB* and *HBA1/HBA2* genes can be challenging to identify and interpret the potential clinical significance. Over the last 18 months at Good Start Genetics, 798 blood samples with variant peaks on hemoglobin electropherograms suggestive of an *HBB* or *HBA1/2* mutation were reflexed to targeted gene sequencing. As expected, the common *HBB* variants were most frequently identified (n=742; 93.2%). We also identified 31 rare or novel hemoglobin variants in the *HBB, HBA1* and *HBA2* genes, accounting for 6.4% (n=53) of all hemoglobin variants seen. Only 3 samples (0.4%) had variants that were still unidentified after using this testing approach. Many of the rare hemoglobin variants have little or no published information regarding the clinical consequences of inheriting the variant, alone or in combination with known pathogenic mutations. Publications are often single case reports with limited information about the clinical and hematological findings. We discuss rare variant interpretation in a well characterized disease. In the setting of a molecular testing laboratory, clinical information about the patient is often not available, and in the context of carrier screening, clinical findings in heterozygotes would not typically be expected. However, given the paucity of information known about these rare hemoglobin variants, more information is needed to accurately characterize these variants and how to counsel patients about their reproductive risks and diagnostic testing options.

3271W

Evaluating early pregnancy loss by whole exome sequencing. *H.M. Byers[1], A. LaCroix[2], T. Naluai-Cecchini[2], J.X. Chong[2], D.R. O'Day[2], M. Bamshad[2,3,4], I.A Glass[2,4], H.C. Mefford[2,4], University of Washington Center for Mendelian Genomics. 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, USA; 2) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA, USA; 3) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 4) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA, USA.*

Background: Pregnancy loss is a heterogeneous condition influenced by both genetic and environmental risk factors. One-percent of couples experience recurrent pregnancy loss (RPL), classically defined as three consecutive losses or, more recently, as ≥2 losses. For 50% of couples no etiology can be identified, denying couples an explanation and making it impossible to provide a precise risk of recurrence. Hypothesis: We hypothesized that that variants in genes critical to fetal development and/or intolerant of variation contribute to early pregnancy loss. Methods: The study is being conducted at a tertiary, academic medical center. Patients are recruited from this center, two private fertility clinics, a healthcare maintenance organization, or self-referral. Inclusion criteria are: ≥2 first trimester losses or ≥1 second or 3rd trimester loss and participation of the complete couple. Ectopic and molar pregnancies are not included. Exclusion criteria included: known cause for pregnancy loss; Mullerian anomaly; major maternal medical condition; losses at maternal age ≥40 or maternal BMI >40. Pedigrees suggestive of an autosomal dominant or recessive mode of inheritance (MOI) were prioritized for whole exome sequencing. Results: To date, 49 families, including 9 trios (male-female-1 loss) and 4 quads (couple + 2 losses), consented to participate were prioritized for whole exome sequencing. Analysis was focused on genes known to be highly conserved, underlie Mendelian conditions, or known to be critical for development as assessed in animal models. Conclusion: There are many challenges in studying genetic causes of pregnancy loss. Technological advances and declining costs of massive parallel sequencing have made it a more feasible area of study. There is enormous patient interest in this topic; identifying genetic causes and risk factors may ultimately benefit the millions of couples struggling with recurrent pregnancy loss.

3272T

A systematic evaluation of RNA quality of human conceptual tissue and its relationship to the post-mortem interval. *D. O'Day, M. Deng, K. Muggli, J. Corona, T. Naluai-Cecchini, T. Cox, I.A. Glass.* Department of Pediatrics, University of Washington, Seattle, WA.

To study tissue specific gene expression in early human prenatal development, high quality tissue samples are required. Quality assessment of starting material is vital for generating valid, accurate and reproducible data for a variety of biomedical research applications. Factors that may impact tissue quality include the post-mortem interval (PMI), tissue handling, and tissue storage prior to performing downstream experiments. Here, we systematically compare two tissue storage methods prior to Total RNA isolation from several human fetal tissues across a range of PMIs. RNA quality was evaluated using an Agilent Bioanalyzer 2100 to obtain the RNA integrity number (RIN). Tissue samples were divided in two and stored separately at -80°C dry and at 4°C in the RNA stabilization and storage solution, RNAlater, before processing the samples for RNA extraction. We determined that tissues stored in RNAlater, according to manufacturer guides, generated RIN scores consistently higher (mean RIN 9.5 ± 0.3 standard deviation; range 9.0 to 9.9) than those of tissues stored at ultra-low temperature (mean RIN 7.1 ± 1.5 standard deviation; range 4.9 to 9.7) prior to RNA extraction. Interestingly, only minimal RNA degradation is seen in cases with a PMI of up to 28 hours at 4°C. This data shows that proper handling and storage of samples can result in high quality RNA from several human fetal tissues including brain, heart, lung, kidney, muscle, and intestine within a PMI of 24 hours.

3273F

Preconception carrier testing using genome sequencing: What categories of results do patients want to receive? *P. Himes[1], K. Bergen[2], M.J. Gilmore[1], J.A. Reiss[2], T.L. Kauffman[2], C.K. McMullen[2], J.V. Davis[2], M.C. Leo[2], J.L. Schneider[2], G.P. Jarvik[3,4], L.M. Amendola[3], M.O. Dorschner[5], D.A. Nickerson[4], C.S. Richards[6], S. Punj[6], D.K. Simpson[1], A. Rope[1], K.M. Porter[7], K.A.B. Goddard[2], B. Wilfond[7].* 1) Department of Genetics, Kaiser Permanente Northwest, Portland, OR; 2) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 3) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Department of Pathology, University of Washington, Seattle, WA; 6) Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR; 7) Seattle Children's Research Institute, Treuman Katz Center for Pediatric Bioethics, Seattle, WA.

Background: As part of the Clinical Sequencing Exploratory Research (CSER) consortium, we investigated the use of genome sequencing (GS) for clinical preconception carrier screening. Compared with currently available clinical tests, single-gene panels, and multi-gene panels, GS can identify variants associated with hundreds to thousands of genetic conditions. However, the preferences of preconception couples regarding the types of carrier results returned through GS are not well studied. Methods: Women who had clinical carrier testing and were planning a pregnancy as well as the male partners of identified carriers were recruited, provided a taxonomy of genetic conditions, and asked to select return of results from broad categories (lifespan limiting, serious, mild, unpredictable, adult onset) rather than specific conditions. Participants could also choose whether to receive medically actionable incidental findings. We examined participants' selections and any differences between those choosing all versus fewer categories based on demographic characteristics and family history (Fisher's exact test, α=.05). Results: Most study participants (92% total; 91% of 279 females, 94% of 46 males) selected all categories. There were no statistically significant differences between non-Hispanic whites and nonwhites (92% vs. 85%; p=.07), participants with knowledge of a family with a genetic condition and those who did not (94% vs. 88%; p=.07), low or high income (91% vs. 90%; p=1.00), having a family history of a genetic condition or not (96% vs. 91%; p=.0.34) or having children or not (93% vs. 89%; p=0.17) on whether they chose all categories. Among the participants who decided against receiving return of results from all categories, 69% declined unpredictable findings, 52% adult-onset findings, 34% mild findings, 14% serious findings, and 10% incidental findings. Conclusions: In a setting with limited barriers to receiving additional information from genome sequencing, most participants elected to receive all possible information. The rate at which individuals selected all genetic results versus a subset of results did not differ by demographic factors. Additional work is needed to explore how participants understand, value, and use the information from each category.

3274W

Investigating *HOXA4 and HOXA11* **expression-affected implantation under the influence of N-acetyl cysteine.** *P. Afsharian[1], V. Mokhtari[1,2,3], M. Shahhoseini[1], S.M. Kalantar[2], A. Moini[3,4].* 1) Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 2) Department of molecular cytogenetics, Research and clinical center for infertility, Shahid Sadoughi University of Medical Sciences, Yazd, Iran; 3) Department of Endocrinology and Female Infertility, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 4) Department of Obstetrics and Gynecology, Faculty of Medicine, Tehran University of Medical Science, Tehran, Iran.

   Statement of purpose: N-acetyl cysteine (NAC) is a well-tolerated mucolytic drug that moderates clinging mucous secretions and enhances glutathione S-transferase activity. Various studies concluded that NAC can be effective on ovulation and pregnancy rate improvement in infertile patients when was co-administered with another drugs such as folic acid. Also, successful implantation rates under the effect of high expression of some genes especially in *HOXA* cluster have been reported in the literature. The main focus in this study is on the efficacy of NAC on implantation by effect on *HOX-A4* and *-A11* expression. As a new aspect of this study, we evaluated the expression of these genes in NAC supplementation during in vitro fertilization (IVF) cycle. **Methods used**: Thirty women aged 22 to 40 with a history of at least two recurrent implantation failure (RIF) who were undergoing IVF cycle were included in the study. The study was of the type a single center, double blinded, placebo controlled, randomized trial. Subjects received either NAC or placebo with both effervescent tablets having similar color, size and appearance. Expression of *HOX-A4* and *-A11* were assessed on the day of WOI (using Real Time PCR) biopsies from endometrium. The patients were randomly categorized in two groups (A/B) to receive NAC 1200 mg/day or placebo, for at-least 9 weeks before starting ovarian stimulation. Pipelled-based biopsy from endometrium was done on specific day (19-21) of their cycle. Then patients were undergone ovarian stimulation (using NAC) ended to IVF treatment. Total RNA-extraction and cDNA synthesis were performed from endometrium samples and evaluatied by RealTime PCR. **Summary of results**: Our obtained results have depicted that expression levels of *HOXA11* gene was significantly (*P* values= 0.046) different in comparing both A and B groups, however we have seen no expression of *HOXA4* in any groups on the day of WOI. As a remarkable finding, implantation rate was tailored in group B in comparison with that in group A. **Conclusion:** There is a significant difference in expression of *HOXA11* in comprising two NAC supplementation and placebo groups during IVF protocols for RIF patients unlike *HOXA4*.

3275T

Nelf **knockdown alters gene expression in GnRH neuron migration.** *E. Ko, L.P. Chorich, M.E. Sullivan, L.C. Layman.* Department of Neuroscience and Regenerative Medicine, Augusta University , Augusta, GA.

   Hypothalamic-pituitary-gonadal (H-P-G) axis function, which is regulated by gonadotropin releasing hormone (GnRH), is critical in pubertal development. Thus proper migration of GnRH neurons from the olfactory placode to the hypothalamus is important. Mutations in Nasal embryonic LHRH factor (NELF), which is also known as NMDA receptor synaptonuclear signaling and neuronal migration factor (NSMF), have been reported in normosomic hypogonadotrpic hypogonadism (nHH) and Kallmann syndrome (KS) patients. *Nelf* knockdown in NLT cells and *Nelf* knock out mice have shown disruption in GnRH neuron migration. However, the mechanism by which NELF regulates GnRH neuron migration is unclear. We reported that NELF is predominantly located in the nucleus and contains two atypical zinc finger domains. Jacob protein, the rat ortholog, is 98% identical to NELF, interacts with CREB after translocating to the nucleus by stimulating the NMDA receptor, and then affects other genes. We hypothesize that NELF could be involved in GnRH migration indirectly through regulation of other proteins (protein-protein interaction) and not by direct interaction (DNA–protein binding).   To determine what genes are potentially regulated by *Nelf*, cDNA arrays were performed following *Nelf* knockdown (KD) in mouse migratory neuronal NLT cells using *Nelf* shRNA Lentiviral particles. We selected 181 genes with at least 2 fold change either up or down-regulated and with P<0.028 in *Nelf* knockdown compared to a scrambled control. We found two known genes, *Cadm1* and *Grina,* which are involved in GnRH neuron migration, to be reduced in cDNA arrays by 52-54% with *Nelf* knockdown. To validate these genes, RT–qPCR was performed. In RT-qPCR, both of these genes were down-regulated, ranging from 58-94%. In addition, we validated 4 of the highest down-regulated genes, *Vnn1, Tmem176b, Rtp4* and *Efemp1,* ranging from 85-100 %, and 3 of the highest up-regulated genes, *Vcan, Hhip, Neot 2,* ranging from 99 -100%, by RT-qPCR. Confirmation of protein levels is ongoing. Findings from this study will further enhance our understanding of the function of *Nelf* in GnRH neuronal migration.

3276F

Finding genes for infertility: A balanced translocation in an oligospermic male uncovers *SYCP2* **overexpression.** *S.L.P. Schilit[1,2], T. Kammin[3], C. Hanscom[4], C. Redin[2,4,5], C.B. Gurumurthy[7], A.J. MacQueen[8], J.F. Gusella[2,4,5,6], M.E. Talkowski[2,4,5,6,9,10], C.C. Morton[2,3,6,11,12].* 1) Program in Genetics and Genomics, Biological and Biomedical Sciences Program, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA; 2) Harvard Medical School, Boston, MA, USA; 3) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA, USA; 4) Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 6) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 7) Developmental Neuroscience, Munroe Meyer Institute, Mouse Genome Engineering Core Facility, University of Nebraska Medical Center, Omaha, NE, USA; 8) Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT, USA; 9) Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 10) Department of Pathology, Massachusetts General Hospital, Boston, MA, USA; 11) Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; 12) Manchester Academic Health Centre, University of Manchester, Manchester, UK.

Infertility affects 10-15% of couples, making it one of the most common disorders for reproductive aged individuals. Over 20% of cases are idiopathic in nature, and it is commonly thought that many have an underlying genetic etiology. Identifying genes involved in unexplained infertility not only informs an understanding of the mechanisms regulating fertility, but also provides information to support diagnosis, genetic counseling and eventual therapeutic intervention. One approach to identifying genes involved in infertility is to examine the phenotype-genotype correlation in subjects with a clinical phenotype (*e.g.*, unexplained infertility) accompanied by a *de novo* balanced chromosomal anomaly (BCA), as is the foundation of gene discovery in the Developmental Genome Anatomy Project (DGAP, dgap.harvard.edu). One subject in this study, designated DGAP230, has oligospermia and 46,XY,t(20;22)(q13.3;q11.2). While it is generally thought that BCAs may reduce fertility due to production of unbalanced gametes, another possibility is that the chromosomal rearrangement disrupts or dysregulates genes important in fertility. By large-insert whole genome sequencing, we refined the location of chromosomal breakpoints in DGAP230. Through assessment of topological associated domains (TADs) at the sites of rearrangement, we have discovered overexpression of *SYCP2*, encoding synaptonemal complex 2 (MIM 604105), a structural protein involved in synapsis of homologous chromosomes in meiosis I. We hypothesize that *SYCP2* overexpression impairs proper spermatogenesis leading to oligospermia by meiotic nondisjunction. This effect could be mediated by either increased adherence between synapsed homologous chromosomes, or conversely, by poorly synapsed chromosomes that result from improper synaptonemal complex stoichiometry. To test causality and molecular mechanism we are modeling *SYCP2* overexpression in mice and *Saccharomyces cerevisiae*, respectively. In addition to pinpointing a putative genetic etiology for DGAP230's oligospermia, our study more generally demonstrates an alternative explanation for BCA-mediated infertility.

3277W

First day urine of discordant monozygotic twin reflected fetal metabolic programming in metabolomics study. *M.K. Thong[1], R.D. Muthukanoo[1], M.F. Loke[2], Y.M. Choo[1], A.A. Kamar[1], M.T. Ishak[1], J. Vadivelu[2].* 1) Department of Pediatrics, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; 2) Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.

Introduction: Metabolomics, is believed to be a promising new tool for clinical diagnosis. Refined clinical management of ill premature newborns may be achieved with knowledge available on neonatal development processes and their metabolomics profile. Objectives: The aim of this study is to investigate whether dynamic urinary metabolomics in a set of premature monozygotic twins reflects their metabolic status and its usefulness in predicting subsequent clinical conditions. Methodology: This is a longitudinal follow up study of a set of Malay female twin who developed multiple life-threatening conditions over the 6 week period of sample collection. A total of 57 urine samples were collected on 'day one' (i.e. within 24 hours of birth), "week two" and "week six" of life from neonates born premature (2 cases and 17 controls). Urinary metabolomics profiling was performed using liquid chromatography-mass spectrometry (LC/MS). Mass Profiler Professional (MPP) and statistic method including ANOVA and principal component analysis (PCA) in conjunction with pathway analysis were used for discovering discriminating metabolites. Results: PCA analysis revealed that urinary metabolomics profiles were significantly different among the twins on day 1, as well as compared to the control neonates, but were comparable on week 2 and week 6. Discussion: Metabolites in urine excreted at day one of life may predict differences in adaptation and alteration caused by the pathological conditions even though it is of discordant monozygotic twins. Pathways associated with vitamin B12, folate, selenium and tryptophan metabolism were significantly different between the twins. In addition, dopamine, melatonin and estrogen metabolites were significantly different among twins. Conclusion: Urinary metabolomics analysis may have a role in early identification of premature newborns with high risk of developing clinical complications and enable timely interventions and management.

3278T

Effects of paternal age on child's telomere length and number of *de novo* **mutations.** *S.W.W. Wong[1], B.D. Solomon[1,2,3], J.E. Niederhuber[1,4].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) Department of Pediatrics, Virginia Commonwealth University School of Medicine, 1201 E Marshall St, Richmond, VA; 3) Department of Pediatrics, Inova Children's Hospital, Inova Health System, Falls Church, VA; 4) Johns Hopkins University School of Medicine, 733 North Broadway Street, Baltimore, MD.

Telomeres play a vital role in maintaining genomic integrity and chromosomal stability in humans. Males tend to have shorter telomeric lengths (TLs) than females of the same age, which leads to the belief that females have less telomere attrition than males. Previous studies have shown that the child's TL is correlated with the father's age at conception, but have also provided contradictory results on whether the child's TL is correlated with either parents' TL. Most of these studies suffer from small sample sizes (number of families <100), as well as studying as an adult, where other intrinsic and extrinsic events during life introduce variance. In this study we estimate the TLs using the software telseq in 1315 whole genome sequenced (~40X) newborns and their parents. We also detected single nucleotide de novo mutations (DNMs) in the probands and were able to determine the parent of origin of over 40% of the DNMs using familial transmission and read haplotype assembly. This cohort is part of the "First 1,000 Days and Beyond Study" at the Inova Translational Medicine Institute, which represents a group unselected for health or disease. Using multiple linear regression, we find that the TL of the newborn child is positively correlated with both parents' TLs ($P<10^{-16}$) and father's age at conception ($P=2\times10^{-7}$), and is negatively correlated with gestational age ($P=6\times10^{-15}$) and male gender ($P=2\times10^{-3}$). The mother's age at conception, newborn's ancestry, and parents' BMI do not appear to be correlated with newborn's TL after controlling for the above factors. We performed separate multiple linear regressions for female and male newborns and found no statistical difference between the two genders in telomere attribution during gestation. Finally, we discovered that the number of DNMs in the newborn from the father is negatively correlated with father's TL ($P=9\times10^{-3}$), after controlling for father's age ($P<10^{-16}$) whereas the number of DNMs from the mother does not correlate with mother's TL, after controlling for mother's age. These observations present an interesting observation where an increase in paternal age may correlate with certain health risks in the child due to more do novo mutations, but at the same time may have a beneficial effect due to increased TL.

3279F

Performance evaluation and clinical implementation of a new paired-end MPSS approach for cfDNA based prenatal screening of common chromosome aneuploidies. *V. Cirigliano[1,2], E. Ordoñez[1,2], L. Rueda[1,2], I. Castilla[1], M. Grau[1], C. Fuster[2], M.P. Cañadas[1].* 1) Molecular Genetics, Labco Diagnostics/Synlab, Barcelona, Spain; 2) Biologia Celular, Universitat Autonoma de Barcelona, Bellaterra, Barcelona, Spain.

Objective: To assess the performance of a novel paired-end MPSS based cfDNA screening method on a large cohort of plasma samples from pregnancies of known outcome. Following the preliminary validation study, the efficiency of this approach was also evaluated on a large set of consecutive clinical samples from average risk pregnancies **Methods:** Retrospective blind study of 1730 1st trimester plasmas including 66 confirmed T21, 34 T18,13 T13 and 8 XY aneuploidies. The new NeoBona test was performed on cfDNA from 1 ml of plasma using Illumina paired-end MPSS and data analysis. An innovative algorithm was applied to generate a new Tscore reflecting the likelihood of aneuploidy based on fetal fraction, counting statistics, cfDNA sizes distribution and sequencing depth. Cut-offs at Tscore values were set to discriminate euploid/aneuploid cases and results compared with fetal karyotypes or neonatal outcomes. The NeoBona was than used to screen 6000 consecutive clinical samples from pregnancies above 10w of gestation (102 twins) regardless their risk category. **Results:** Results were obtained in 98.8% of cases from the validation cohort. All 21, 13 and XY aneuploidies were detected (100% sensitivity and specificity) as also 33/34 T18. Normal/aneuploid cases were correctly scored even at fetal fractions <1%. Results were obtained for 98% clinical samples, 96 trisomies 21, 17 trisomies 18 and 13 trisomies 13 were identified; 2 false positive results for T21 and one for T13 were observed (FPR 0.03% and 0.02% respectively). XY aneuploidy screening was also requested in 57% of clinical samples and 15 cases identified. Invasive procedures were performed in 9/15 cases and 1 false positive 45,X result was observed (FPR 0.05%). No follow up for normal results was available yet, the majority of which are still ongoing. Test failures occurred in 125 samples (2 twins), blood redraw provided valid results in 81% of cases. Thus, the NeoBona test showed overall success rate of 99.6% at redraw rate of 2%. **Conclusions**: Paired-end MPSS allowed simultaneous assessment of fetal fraction, cfDNA size distribution and counting which, combined with sequencing depth into a new analysis algorithm, allowed the NeoBona test to be successful even at fetal fractions <1% while reducing FPR. Setting cut-offs at the new multifactorial Tscore eliminates the lower limit at FF for reportable cases thus, potentially extending the benefits of cfDNA screening to a larger proportion of pregnancies. .

3280W

Computing confidence intervals on positive predictive value for non-invasive prenatal screening. *I.S. Haque, C. Haverty, J.D. Goldberg, E.A. Evans.* Counsyl, South San Francisco, CA.

While non-invasive prenatal screening (NIPS) for fetal aneuploidy has high sensitivity and specificity, prevalence varies significantly by maternal and gestational age. Variable prevalence affects the probability that a positive test indicates an affected fetus (positive predictive value, PPV). While ACOG and SMFM direct laboratories to report PPV individualized to the particular patient, no previous work has addressed how uncertain PPV calculations are or to what precision PPV can be estimated.We introduce a statistical framework to estimate the confidence interval (CI) around NIPS PPV. We estimate the uncertainty in NIPS analytical performance by fitting beta distributions to the 95% CIs on test sensitivity and specificity from large meta-analyses [1]. We sample the posterior distribution of aneuploidy prevalence by using original population data [2,3,4,5] to sample from beta distributions or bootstrap Kaplan-Meier curves as appropriate. We estimate the portion of the CI arising from each source by holding individual parameters fixed.With this method, we evaluate the 95% CI of NIPS PPV for trisomies 13, 18, and 21 (T13/18/21). Sampling all sources of uncertainty shows that CI breadth mostly varies by maternal age (MA): for example, the CI width for T21 is +/- 11% at age 20 down to +/- 4% at age 40. Low mean PPV for T13 and T18 leads to asymmetric CIs: the T13 PPV CI at MA=20yr and 12 weeks' gestation has a lower bound 13% below the mean of 20.8% but an upper bound 23% above the mean. All CIs are dominated by uncertainty in test specificity: at MA=20yr, the T21 CI is +/- 10.5% when sampling only over test specificity, +/-4% using only prevalence, and +/- 0.1% using only test sensitivity.Our results show that larger meta-analyses of NIPS specificity could narrow the PPV confidence interval two-fold for T21 or four-fold for T13 and T18 (motivating outcome data sharing among laboratories), but that additional gestational prevalence data will then be needed to further improve the precision of PPV estimates. The utility of providing PPV is well-justified even with the estimated CI. This work further demonstrates the need to use large meta-analyses to establish sensitivity and specificity rather than smaller individual studies.[1] Taylor-Phillips S et al. BMJ Open 2016 [2] Hecht CA and Hook HB. Prenat Diagn 1994 [3] Snijders RJM et al. Ultrasound Obstet Gynecol 1999 [4] Morris JK and Savva GM. Am J Med Genet A 2008 [5] Savva GM et al. Prenat Diagn 2010.

3281T

Development of a national NIPT in Brazil. *C. Malcher[1], G.L. Yamamoto[1], P. Burnham[2], S.A.M. Ezquina[1], D.S. Marco Antonio[1], N.C.V. Lourenço[1], D. Bertola[1], I. De Vlaminck[2], M.R. Passos-Bueno[1].* 1) Human Genome Research Center, Institute of Biosciences, Universidade de São Paulo, São Paulo, Brazil; 2) Department of Biomedical Engineering, Cornell University, Ithaca, NY.

Background: Noninvasive prenatal tests (NIPT) performed in Brazil nowadays rely on outsourcing technology or the test itself. Our aim is to develop a national NIPT test at the Human Genome and Stem Cell Research Center using next generation sequencing **Method**: Libraries were prepared, indexed, multiplexed, captured for a gene panel and sequenced in MiSeq. We aligned the fastq files, removed duplicated reads, realigned based on known local indels and also removed reads with more than two mismatches. For trisomy detection (T21) we used 240 probes across chromosome 21, while for fetal sex determination we used 1 probe on the *SRY* gene on chromosome Y. We performed read count for these regions, followed by a Z-score test. For fetal fraction estimation we developed an in-house model to infer fetal fraction using SNPs (developed in R) detected from 6739 probes across 388 autosomal genes. **Results**: We used 68 samples (33 pregnant samples and 35 mock samples mimicking fetal fraction increase during pregnancy). Sequencing yielded a mean of 15.2 millions raw fastq reads (8.7-33.5 million) and mean coverage of bam files was 191.65 X (39.99 X-294.3X). Mean fitted fetal fraction for pregnant samples was 0.12 (0.02-0.3). Correlation analysis showed strong positive correlation between fetal fraction and gestational age (Pearson correlation r^2 =0.568, p-value = 5.5E-4). We did not find a significant association between fetal fraction and maternal weight (Pearson correlation r^2= -0.219; p-value=0.236). For fetal sex determination, we correctly classified all the samples, with a sensitivity of 100% (95% CI:87%-100%) and specificity of 100% (95% CI:87%-100%). Using a fetal fraction threshold of 0.06 for trisomy detection, we classified: 9/10 T21 pregnant samples with a Z-score above 3, 1/10 T21 pregnant samples with a Z-score below 3 (false negative) and 1/58 of the samples with a not-affected fetus presented a Z-score above 3 (false positive; Z-score=5.6, fetal fraction=0.18). Therefore, we have a sensitivity of 90% (95% CI:55.5%-99.75%) and a specificity of 98.28% (95% CI:90.76%-99.96%) for T21 detection. **Conclusion**: We are the first group in Brazil to develop a national noninvasive prenatal test of genetic disorders. We used high-coverage targeted next-generation sequencing in order to detect fetal genetic disorders, determine fetal sex (gender-independent) and estimate fetal fraction using only sequencing data, without any additional laboratory steps. Support: Fapesp/CNPq.

3282F

Impact of genetic disease, mode of transmission and mutation type on the outcome of preimplantation genetic diagnosis. *A.J. Hehr[1], B. Seifert[2], B. Paulmann[2], L. Eichhammer[1], C. Hofmann[1], D. Seifert[2], C. Gaßner[2], U. Hehr[1].* 1) Center for and Department of Human Genetics, Regensburg, Bayern, Germany; 2) KITZ, Regensburg, Bayern, Germany.

   Introduction: In the unique setting of the German Embryo Protection law preimplantation genetic diagnosis (PGD) for monogenic disorders in Germany could only be performed by polar body diagnosis (PBD) until 2015. With an average clinical pregnancy rate of 28% per PBD cycle the results are comparable to the PGD results obtained after blastomere biopsy on day 3. Preliminary data of our unique PBD cohort indicate an important impact of the underlying monogenic condition on PGD outcome. While a reproductive disadvantage during ART for female *FMR1* premutation carriers is well known, our data suggest reduced PGD success rates for further Repeat expansion disorders (RED) including Myotonic Dystrophy type 1 (MD1).   **Methods:** retrospective assessment of 208 PBD cycles for 100 families and 37 different monogenetic disorder including 30 female carriers for RED - Fragile X Syndrome (16), MD1 (11), Huntington disease (2) and SCA1 (1). **Results:** Transfer of 115 embryos in 67 PBD RED cycles resulted in 13 clinical pregnancies (19,1% per embryo transfer=ET), 6 missed abortions (46,2%) and 7 live births/ongoing pregnancies. In contrast, In 119 PBD cycles with ET for all other monogenic disorders (non-RED; 33) we obtained a clinical pregnancy rate of 32,1% per ET with 37 births/ongoing pregnancies and 15,9% missed abortions. In 26 PBD treatment cycles for 11 female MD1 carriers similar numbers of retrieved 12,9 vs. 14,1 non RED), mature (9,7 vs. 11,4) or fertilized oocytes (6,5 vs. 7,0) per PBD cycle were obtained at a median female age of 32,88 years (34,0 years non RED cohort). We did, however, observe a reduced implantation rate of 13,3% vs. 19,8% per transferred embryo and birth rate of 7,6% vs. 26,2% per PBD RED cycle compared to our non-RED cohort, correlated with decreased Anti-Mueller-Hormone levels. **Conclusion:** Assessment of our unique PBD cohort is indicating an important impact of the underlying genetic disorder and type of mutation on PGD outcome. We suggest, that female carriers of RED should be counseled about their reduced PGD success rates in order to allow informed reproductive decisions. In an ongoing study we currently address potential predictive biomarkers for favorable PGD outcome in order to improve individual pre-PGD counseling of female carriers interested in RED-PGD. Our preliminary RED data point towards impaired oocyte/embryo quality or endometrium receptivity and may guide further improvement of ART strategies for this PGD cohort.

3283W

A review of a 20-year experience with prenatal diagnosis records, 9,297 cases from Turkey. *H. Bolat[1], B. Durmaz[1], Z. Cengisiz[1], E. Karaca[1], A. Durmaz[1], A. Aykut[1], E. Pariltay[1], T. Sozen Turk[1], A. Rashnonejad[3], E. Arslan Ates[1], T. Atik[2], B. Erturk[1], H. Taslidere[1], N. Aliyeva[1], C. Gunduz[4], H. Onay[1], F. Ozkinay[1,2], H. Akin[1], O. Cogulu[1,2].* 1) Department of Medical Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 2) Department of Pediatric Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 3) Department of Biotechnology, Ege University Faculty of Medicine, Izmir, Turkey; 4) Department of Medical Biology, Ege University Faculty of Medicine, Izmir, Turkey.

   Introduction: Although novel procedures such as fetal DNA isolation from maternal blood are avaliable, invasive prenatal tests are still extensively being used for prenatal diagnosis in most countries. The aim of this study is to evaluate the demographic data, indications and cytogenetic results of prenatal samples. **Methods:** We conducted this study using records from 9297 patients that were referred to Ege University Hospital. The records of these referrals were retrospectively evaluated between 1995 and 2015. Chi-square and logistic regression analysis tests were performed for the statistical analysis of different events. **Results:** Analyses were carried out for the results of 8363 amniocentesis, 626 chorionic villus sampling and 308 cordocentesis samples. 50.80% of patients were ≥35 years of age. The number of referrals significantly decreased after 2009. Choromosome anomalies were determined in 538 (%5.8) cases (60.41% numerical, 39.59% structural). Autosomal trisomies were detected in 2.38%. Balanced structural anomalies were found in 0.9% whereas unbalanced structural anomalies were described in 0.1% of all samples. The demand for the prenatal testing of single gene disorders has increased throughout the years. Parental anxiety made up the 0.9% of the indications. The indications for prenatal diagnosis were advanced maternal age (48.2%), abnormal maternal screening tests (25.7%), abnormal ultrasound (USG) findings (11%). The indication of abnormal USG findings is found to be the most strongly associated with chromosomal abnormalities. Risk of having trisomy 21 significantly increased with advanced maternal age however trisomy 18 and 13 risk also increased but the difference was not statistically significant. The risk of having a trisomy 21 pregnancy was compared in different age groups. It was doubled in the 36-40 age group, 5 times and 10 times increased in the 36-40 and 41-45 age groups respectively compared to 21-25 age group. We found a cut-off value of 1/250 for trisomy 21, but we could not find a significant linear correlation between maternal serum screening test results higher than 1/250 with trisomy 21. **Conclusion:** Early diagnosis of genetic diseases has become very important in the health care system. We believe that, the review of these indications and their results might be helpful not only in gentic counseling before prenatal diagnosis but also in developing long term and effective strategies of genetic services.

3284T

Targeted gene set enrichment of placental transcriptome in preeclamp-sia. *J. Schuster[1], A. Uzun[1,2], J. Padbury[1,2,3].* 1) Pediatrics, Women and Infant's Hospital, Providence, RI; 2) Brown Alpert Medical School, Providence, RI; 3) Center for Computational Molecular Biology, Brown University, Providence, RI.

   Preeclampsia is a systemic and potentially life-threatening hypertensive disorder of pregnancy. Semantic data mining and natural language processing were used to curate published literature to identify genes associated to pre-eclampsia. We built a Database for Preeclampsia (dbPEC) consisting of these genes along with their associated clinical features, concurrent conditions, and published supporting literature. We previously aggregated the mined genes into gene sets associated with severity and timing of the phenotype, as well as tissue source and concurrent conditions (i.e., fetal growth restriction (FGR), gestational hypertension, or hemolysis, elevated liver enzymes, and low platelet count). Here we sought to identify enrichment of these gene sets in well phenotyped, published placental genome-wide expression data sets. The dbPEC gene sets associate to the corresponding phenotypic specifications in these expression data sets. One expression array compared severe early on-set preeclamptic patients (n=5) to normotensive patients (n=4) and we found significant enrichment for the gene set of fetal, early onset genes (p<0.0001). For a second expression array comparing severe preeclamptic patients (n=8) to normotensive patients (n=8) both the fetal, early onset gene set and the fetal, severe preeclamptic gene sets were nominally enriched (p<0.05). We also interrogated the networks of genes contributing the enrichment score for the most enriched gene set (*FLT1, CRH, PAPPA2, ABCA1, TGFB1, KDR, and NOV*) using GeneMania and Ingenuity Pathway Analysis. We assessed the biological functions, molecular functions and cellular components of these genes using the Gene Ontology Database and found the following signifi-cant GO terms: positive regulation of phosphate metabolic process, vascular endothelial growth factor-activated receptor activity, growth factor binding, and extracellular region. The approach is more successful than univariate analysis of genome-wide expression data because it leverages the genetic architecture of preeclampsia, by examining the combinatorial effect of genes using knowl-edge-based inferences.

3285F

A modified genotyping method to reliably identify single fetal cells in the maternal circulation. *X. Zhuo, Q. Wang, A. Breman, A. Beaudet.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

   Background: As part of an effort to develop fetal cell based noninvasive prenatal testing (CB-NIPT), we have sought a rapid, robust, and low cost method to genotype individual presumptive fetal cells especially to distinguish these from maternal cells in female pregnancies. We have chosen to modify a method described by others (Debeljak et al. PMID: 25132481) using haplotype counting by next-generation sequencing (NGS) for analysis of single cells. **Methods:** For cells from 12-16 weeks gestation pregnancies, both cytoker-atin (CK)-positive putative fetal cells and CK-negative maternal white blood cells were picked from slides from maternal blood using methods submitted for publication. The methods for recovering fetal cells involve fixation and permeabilization of all cells followed by staining for CK which is positive in fetal trophoblasts and for CD45 which is negative in fetal cells but positive in most maternal cells. Highly polymorphic genomic regions with 10-20 very informative SNPs within a 200-300 bp interval were amplified as described by Debeljak et al. A region of HLA-A was used primarily but other amplicons were also utilized including Y-specific regions. NGS was performed on the Ion Torrent platform. **Results:** This method allowed reliable differentiation of fetal and maternal cells. The HLA-A amplicon included >11 highly informative SNPs and was informative for distinguishing maternal and fetal cells in >90% of samples. Additional amplicons can be used to resolve uninformative cases, and a Y-specific amplicon is useful for male fetuses. Read depth per sample is usually >1000 reads of NGS data. Typically in fully informative cases, two haplotypes are found within the maternal reads, and fetal cells show reads with one but not the second maternal haplotype while also showing a novel paternal haplotype not present in the mother. A paternal DNA sample is not re-quired for use of this method to distinguish maternal and fetal cells in informa-tive cases. **Conclusion:** We found this method to be superior to SNP arrays, multiplex genotyping methods, and STR analysis considering cost, throughput, and reliable distinction of fetal from maternal cells. This method derives its power by providing a highly informative haplotype within individuals reads for single molecules. This method will facilitate genotyping of every single fetal cell within an overall platform for performing CB-NIPT.

3286W

Placental *MTHFR* 677C>T genotypes in pregnancy pathologies. *G.F. Del Gobbo[1,2], E.M. Price[1,2], C.W. Hanna[3,4], W.P. Robinson[1,2].* 1) University of British Columbia, Vancouver, Canada; 2) Child & Family Research Institute, Vancouver, Canada; 3) Babraham Institute, Cambridge, United Kingdom; 4) Cambridge University, Cambridge, United Kingdom.

Background: One-carbon metabolism is the biochemical pathway that activates and transfers one-carbon units for methylation reactions and nucleotide synthesis. A key step in this pathway is catalyzed by 5,10-methylenetetrahydrofolate reductase (MTHFR), which irreversibly commits one-carbon units to methylation reactions. A non-synonymous SNP in *MTHFR*, c.677C>T, produces a thermolabile enzyme with reduced function, which can be further exacerbated under folate-limited conditions. During pregnancy, the placenta extracts and concentrates folate from maternal circulation. Both maternal and placental/fetal *MTHFR* variants have been associated with pregnancy complications including miscarriage, neural tube defects (NTDs), and pre-eclampsia (PE). **Objectives:** To determine whether the placental *MTHFR* 677 T/T genotype is associated with NTDs, placental insufficiency (PI) (including PE and/or intrauterine growth restriction (IUGR)) and/or reduced birth weight/placental weight (BW/PW) ratio. **Methods:** Placental samples were acquired from pregnancies delivered at BC Women's Hospital in Vancouver, BC. Placentas from 195 healthy control, 55 NTD, and 65 PI pregnancies were genotyped by pyrosequencing at *MTHFR* 677. Fisher's exact tests were used to assess increased T/T genotype frequencies in pathology groups. ANOVA was used to compare BW/PW ratio between *MTHFR* 677 C/C + C/T genotype and T/T genotype placentas in a subset of N=173 healthy placentas. **Results:** Study population genotype frequencies were C/C: 0.498, T/C: 0.422, and T/T: 0.076, consistent with expectations for Hardy-Weinberg equilibrium and for a majority Caucasian and Asian population. NTD T/T frequency (0.091) was not significantly different from controls (T/T: 0.051), ($p = 0.214$), while the PI group had a marginally significant increase in the T/T genotype (T/T: 0.138), ($p = 0.024$) compared to controls. When the PI cases were subdivided by pathology, i.e. pre-eclampsia (both early- and late-onset) and normotensive IUGR, we found that no individual pathology drove the increased T/T frequency in the PI group. BW/PW ratio, a measure of placental efficiency, was not associated with placental *MTHFR* 677 T/T genotype ($p = 0.647$). **Conclusion:** Fetal *MTHFR* c.677C>T genotype does not contribute significantly to risk of pregnancy complications or altered BW/PW ratio in our folate-supplemented population. However, follow-up studies with larger sample sizes may be warranted for specific PI sub-pathologies.

3287T

Genome-wide analyses for age at menopause identify gene-environment interactions and causal relationships with other complex traits. *K.S. Ruth, The ReproGen Consortium (www.reprogen.org).* Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom.

As more women delay childbearing, reproductive ageing has increasing relevance. Menopause timing has also been associated with adverse health outcomes such as breast cancer, heart disease, osteoporosis and type 2 diabetes. Previous GWAS meta-analysis by the ReproGen consortium focused on HapMap 2 imputed variants and in 70,000 women identified 56 mostly common genetic variants that together explain ~6% of population variance in age at natural menopause (ANM). We carried out a meta-analysis of 108,000 women imputed to the 1000 Genomes reference panel and thus including analysis of rarer variants. We used UK Biobank (35,000 menopausal women) to test a genetic-risk score for ANM for interactions with risk factors for ANM, and to perform Mendelian Randomization (MR) analyses. We identified 123 independent signals at 101 loci (MAF 1–49%; per allele effect -0.1 to -1.4yrs; smallest $P=1\times10^{-168}$) including 60 new signals, multiple independent signals at 14 loci and 4 signals on chr X. Non-synonymous variants were found in LD with 25 signals ($r^2>0.8$). As in previous analyses, DNA repair genes were enriched; new signals included variants in *FANCA* (per allele effect -0.1yrs; $P=2\times10^{-9}$) and *FANCM* (per allele effect -0.2yrs; $P=1\times10^{-10}$), two of which were non-synonymous. Other new signals included a non-synonymous variant in *CALCR* (per allele effect -0.1yrs; $P=8\times10^{-9}$) and an intronic variant in *FTO* (per allele effect -0.3 yrs; $P=3\times10^{-8}$), which was an eQTL for *FTO* ($P=2\times10^{-7}$) and was independent of the known BMI signal at this gene. Genetic risk for menopause appeared to be modified by maternal smoking during pregnancy, with both the combined SNP score and four individual variants exhibiting signs of significant G×E effects ($P<0.01$). MR analyses supported a causal role for later menopause on increased breast cancer risk, increased bone mineral density, decreased risk of osteoporosis and increased risk of uterine fibroids. By increasing the sample size and using a denser imputation we have more than doubled the number of known genetic signals for ANM and explained 8.5% of the population variance. Our results suggest a novel role for *FTO* in reproductive lifespan. MR analyses support the epidemiological relationship of ANM with osteoporosis. Finally, gene–maternal smoking interactions highlight the importance of the uterine environment during oocyte formation, and suggest that *in utero* exposures can affect reproductive lifespan.

3288F

A comprehensive strategy for exome-based preconception carrier screening. *S.C.E.H. Sallevelt, B. de Koning, R. Szklarczyk, A.D.C. Paulussen, C.E.M. de Die-Smulders, H.J.M. Smeets.* Department of Clinical Genetics, Maastricht University Medical Center+ (MUMC+), Maastricht, The Netherlands.

Whole exome sequencing (WES) provides the possibility of genome-wide preconception carrier screening (PCS). Here, we propose a filter strategy to rapidly identify the majority of relevant pathogenic mutations, which was developed on WES data from 8 consanguineous and 5 fictive non-consanguineous couples, and subsequently applied to 20 other fictive non-consanguineous couples. Presumably pathogenic variants based on frequency and 1) database annotations (HGMD/ClinVar) or 2) generic characteristics and mutation type (nonsense, frameshift and splice-site mutations), were selected in genes shared by the couple and in the female's X-chromosome. Unclassified variants, predominantly missense mutations and in-frame deletions/insertions not present in HGMD or annotated pathogenic in ClinVar, were not included. This yielded on average 29 [19-51] variants in genes shared by the consanguineous couples (n=8) and 15 [6-30] by the non-consanguineous (n=25). For X-linked variants the numbers per female were 3 [1-5] and 1 [0-3], respectively. Remaining variants were verified manually. The majority could be quickly discarded, effectively leaving true pathogenic variants. Dependent on the clinical consequences, reproductive choices can be offered to the couple. Addition of gene panels for filtering was not favorable, as it resulted in missing pathogenic variants. We conclude that WES is eligible for PCS, both for consanguineous and non-consanguineous couples, with the remaining number of variants being manageable in a clinical setting. It is important to develop and continuously curate databases with pathogenic mutations in order to further increase sensitivity of WES-based PCS.

3289W

Toll-like receptor 1 SNP markers (*TLR1* N248S, *TLR1* H305L) and preterm birth in a Wisconsin cohort. *T. Ranade[1], W. Luo[2], J. Eickhoff[2,3], B.R. Pattnaik[2,4], M. Baker[2,5,6], S.A. Tokarz[2], D.-A.M. Pillers[2,5,7].* 1) University of Wisconsin, Madison, WI; 2) Department of Pediatrics, University of Wisconsin, Madison, WI; 3) Department of Biostatistics, University of Wisconsin, Madison, WI; 4) Department of Ophthalmology and Vision Sciences, University of Wisconsin, Madison, WI; 5) Department of Medical Genetics, University of Wisconsin, Madison, WI; 6) Wisconsin State Laboratory of Hygiene (WSLH) Newborn Screening Laboratory, Madison, WI; 7) J.F. Crow Institute for the Study of Evolution, University of Wisconsin, Madison, WI.

Background: A subset of premature births occur in pregnancies complicated by infection. The first line of defense against infection is the innate immune system at the Toll-like receptor (TLR). TLRs are pattern-recognition receptors that recognize pathogen-associated molecular patterns (PAMPs). Activation of TLRs initiates an inflammation cascade of signaling moieties and cytokines. Genetic variation in these pathways is associated with diseases involving chronic and deregulated inflammation. Inflammation is associated with preterm birth, and TLR alterations are an area of investigation in prematurity research. We have previously reported results for SNPs in three TLR receptors (TLR1, 2, and 4). We found that TLR1 SNP rs4986791 (N248S) is associated with preterm birth in Wisconsin Black infants. Purpose: We sought to validate our association of TLR1 with preterm birth by studying a second TLR1 SNP in a discrete position within the gene. We hypothesized that TLR1 SNP rs3923647 (H305L) also marks an increased risk for preterm birth, and that the combined effects of TLR1 SNP (H305L) and (N248S) may enhance the association. Methods: We used DNA samples from residual newborn screening specimens from Wisconsin infants of gestational ages (GA), 23-42 wks. IRB approval was secured. 3040 samples were used for TLR1 SNP (N248S) analysis, and 1616 of these samples were adequate for subsequent TLR1 SNP (H305L) analysis. SNP genotype was determined using TaqMan® SNP Genotyping Assays (Lifetechnologies, CA). Results: The H305L minor allele T was associated with preterm GA< 34 wks with Odds Ratio (OR) 1.62 (P < 0.0003) and with GA< 28 wks with OR 1.56 (P < 0.0024). The combined effects minor alleles of TLR1 SNP 305 (T) and SNP 248 (C) also showed an association with prematurity (GA < 34 and <28 wks) (P < 0.00001), as well as low birth weight (< 1250g) (P < 0.0000015). The combined effects were most prominent in the Wisconsin Black cohort (GA<34 wks; OR for Whites is 1.81 vs. 3.27 for Blacks). SNP 305 heterozygotes were under-represented in our sample. This may reflect geographic population effects described in Wisconsin and is under ongoing investigation. Discussion: We found a significant association between TLR1 SNP (H305L) and preterm birth in a Wisconsin newborn cohort. The combination of minor alleles for both TLR1 SNP (H305L) and (N248S) was also associated with preterm birth. We recommend further study of TLR1 and its potential relationship to premature birth.

3290T

Genome-wide maternal uniparental disomy mosaicism identified in a products of conception sample tested via 24-chromosome single nucleotide polymorphism microarray with bioinformatics. *K. Merrion[1], M. Maisenbacher[1], D. Clark[1], M. Young[1], B. Levy[1,2].* 1) Natera, Inc., San Carlos, CA; 2) Columbia University Medical Center & the New York Presbyterian Hospital, New York, NY.

PURPOSE: Report a unique case of genome-wide maternal uniparental disomy (UPD) mosaicism in a products of conception (POC) sample. METHODS: POC testing was performed on a reported identical-twin loss at 16 weeks gestation. A fresh POC sample was sent to a reference lab with a maternal blood sample for analysis. Genotyping was performed using Illumina CytoSNP-12b microarrays with bioinformatics to confirm parental origin of abnormalities and rule out maternal cell contamination (MCC). RESULTS: A total of 5 dissections were performed: 2 were reported as MCC; 1 from the gestational sac showed an abnormal male result with triploidy of maternal origin; 2 taken from fetal skin showed a normal female result with all chromosomes being maternal in origin, consistent with genome-wide maternal heterodisomic UPD (rather than MCC, as the specimen was of clear fetal origin). Combined, the data reveal a complex case of mosaicism involving triploidy of maternal origin and a separate cell line with genome-wide maternal UPD. CONCLUSIONS: Although genome-wide paternal UPD, associated with complete molar pregnancy, is well documented in miscarriage, genome-wide maternal UPD has rarely been reported. In this case, an additional cell line with maternal triploidy, which has a known mechanism of either fertilization of a diploid egg, retention of a polar body, or fertilization of an ovulated primary oocyte,[1] was present. To our knowledge, there have been no prior reports of maternal triploidy/genome-wide maternal UPD mosaicism, and, the mechanism for this finding is unknown. There have been published reports of diploid/triploid mosaicism in a limited number of patients with intellectual disability and dysmorphic features,[2] and for cases in which parental origin of the extra set of chromosomes was determined, the majority were maternal in origin.[2] Several authors proposed a likely mechanism of digynic diploid/triploid mosaicism being inclusion of a second polar body into a diploid embryo early in development.[2,3] In this case, given the diploid cell line showed genome-wide maternal UPD, our proposed mechanism is a triploid conception followed by a postzygotic maldivision. References: 1) Gardner, R. J. McKinlay and Sutherland, Grant R. Chromosome Abnormalities and Genetic Counseling; 3 edition (August 28, 2003): 235. 2) Van de Laar, et al. Clin Genet 2002: 62: 376-382. 3) Muller, et al. J Med Genet 1993; 30: 597-600.

3291F

Embryo to embryo hybridisation improves the quality and reduces the cost of preimplantation genetic diagnosis for chromosomal rearrangements using microarrays. *J.W. Ahn[1], A. Corrigan[2], A.R. Jones[1], S. Bint[2], A.F. Davies[2], H. Hewitson[2], P. Renwick[3], C. Mackie Ogilvie[3].* 1) Genome Informatics, Guy's and St Thomas' NHS Foundation Trust, London, UK; 2) Genetics Laboratories, Viapath, London, UK; 3) Genetics Laboratories, Guy's and St Thomas' NHS Foundation Trust, London, UK.

Preimplantation Genetic Diagnosis (PGD) is offered to couples that carry chromosome rearrangements and are therefore at risk of conceiving an abnormal child. Five to seven days following intracytoplasmic sperm injection, 5-10 cells from the embryonic trophectoderm are collected for testing. Whole genome amplification (WGA) is required due to the small amount of starting material; multiple displacement amplification of DNA is performed before oligonucleotide microarray comparative genomic hybridization. Standard protocols use commercial DNA derived from cell lines as control samples; this DNA is serially diluted to ~10 genomes before WGA. Amplification bias results in noise being introduced into the microarray data which can make detection of chromosome abnormalities challenging, especially with smaller segment imbalances. We hypothesized that this noise can be reduced by using control DNA extracted from a similar tissue type to the test sample. To evaluate this, we used as a control amplified DNA from another embryo biopsied at the same time as the test embryo. This reduced the mean derivative log ratio standard deviation (a measure of noise in array data) from 0.58 (n=46) to 0.31 (n=25), which was significant (p<0.01). Cases with known imbalances ranging 2.5-8.9 Mb were selected as test embryos; all imbalances were detected. We also developed software (EvE) to virtually hybridise embryos against each other. This allowed us to check that the two embryos that were physically hybridized together were not masking imbalances in each other. As a result of this work, we have improved the resolution of this test, and can now confidently detect 5Mb imbalances, down from our previous detection limit of 10Mb. Furthermore, the consumable costs associated with the test are reduced. This workflow therefore represents an advance in delivering PGD for chromosome imbalance.

3292W

Enhancing X-linked disease risk assessment with a non-invasive prenatal molecular assay. *M. Sekedat, L. Dohany, Y. Hong, J. Kim, H. Mueller, T. Reeder, J. Stoerker.* Progenity, Inc., Ann Arbor, MI.

   Cost-efficient sequencing technologies have enabled significant expansion of carrier screening panels for pregnant patients. Laboratories routinely perform expanded carrier screening for hundreds of genetic disorders, many of which are X-linked. For most mutations, parental carrier status only indicates the probability of giving birth to an affected child. Additional invasive testing is required for determining if the fetus is affected. However, there is an opportunity to refine the risk assessment for X-linked recessive disorders since children with two X chromosomes are very rarely affected. To address this, we developed a molecular assay that would accurately determine the presence or absence of Y chromosome material in maternal plasma as early as 10 weeks gestation. This high-throughput assay combines automated DNA extraction from maternal plasma, library preparation targeting sites on all chromosomes with emphasis on selection of the ChrY sites, and shallow-depth Illumina sequencing. A training set of 46 samples from pregnant patients carrying an XX fetus was used to generate a modified z-score for ChrY – referred to as the Y-score – for all samples. A second training set of 46 samples from pregnant patients carrying an XY fetus was used to determine the minimum Y-score for samples positive for ChrY material. The trained Y-score algorithm was tested on 433 samples having previously undergone cell-free DNA (cfDNA) testing (48.7% XX and 51.3% XY). The results were 100% concordant, indicating test accuracy equal to cfDNA testing for determining fetal ChrY status. **Conclusions:** The presence or absence of ChrY material can be used to enhance the risk assessment and guide management choices for pregnant patients who are carriers of X-linked disorders. This assay could be performed in addition to expanded carrier screening, offering valuable information to carriers of X-linked disorders, and may help reduce unnecessary invasive procedures for patients who are at low-risk of having an affected child.

3293T

Preimplantation genetic diagnosis leads to successful ongoing pregnancies in complex chromosomal rearrangements carriers. *B.C.F.K. Brunet, J.D. Shen, L.B. Cai, J.Z. Xie, J.Y. Liu, W. Wu.* The State Key Laboratory of Reproductive Medicine, Clinical Center of Reproductive Medicine, First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, PR China.

   BRIEF BACKGROUND: Complex chromosome rearrangements (CCR) is a structural abnormality involving at least three break points on two or more chromosomes. It is a rare event and often related with spontaneous abortions, fetal malformations and infertility. Using next generation sequencing (NGS) for preimplantation genetic diagnosis (PGD) may increase the chance of achieving a balanced euploid pregnancy.   METHOD USED: Five CCR carriers were enrolled in this study: case A: 46, XY, t(1;4;11)(p31;p16;p22), case B: 46, XY, t(3;13;5)(p14;q21;p14), case C: 46, XX, inv(9)(p12;q13), t(13;15)(q14;q24), case D: 46, XX, inv(9)(p12q13), t(7;9)(q22;p22) and case E: 46, XX, t(2;7) (q21; q36), t(2;4) (p10;q10), t(2;4)(q15;q10). After 7 cycles, 85 oocytes were retrieved. Whole genome sequencing (WGS) was performed on 31 biopsied trophectoderm samples on day 5 or 6. Pregnancy was confirmed by fetal heartbeat using ultrasonography.   SUMMARY OF RESULTS: Five embryos were identified by WGS as balanced euploid. Of them, four were transferred resulting in three ongoing pregnancies and one miscarriage. The miscarriage is from case A, 40 days after her single embryo transfer in the first cycle. Chromosomal analysis of the abortus revealed normal karyotype. In the second cycle, the remaining embryos were transferred and her pregnancy is ongoing (over 50 days). There was no transferable embryo for case B after two cycles and for case E after one cycle. In the first cycle of case C, two embryos were identified as transferrable, a single embryo transfer was performed and her pregnancy is ongoing (over 50 days). Case D had a successful embryo transfer for only one euploid embryo, and showed a normal karyotype by amniocentesis at her eighth gestational week. In summary, the implantation rate per transfer is 100% (4 out of 4) and the ongoing pregnancy rate is 75% (3 out of 4). These results show that the use of NGS for CCR cases is promising for PGD.

3294F

An excess of chromosome 4 breakpoint in recurrent pregnancy loss. *R. Frikha, N. Bouayed, T. Rebai. Laboratory of histology, Sfax, Tunisia.*

Recurrent pregnancy loss, as at least more than 2 miscarriages, is a serious problem that affects couples who are trying to conceive. Many factors involved in reproductive process are far to being completely understood. Often, cytogenetic studies have an important role in the assesement of chromosomes abnormalities. For the search of potential loci, which might be revealed by clustering of chromosomal breakpoints, we report two reciprocal translocation involving chromosome 4 among couples with more than 2 early miscarriages and with a particular history of familial reproductive failure. These translocations were detected, among a total of six reciprocal translocations identified in 169 couples. Thus, this overrepresentation of chromosome 4 must reflect a real phenomenon. We suggest that chromosome 4 harbors a large chromosomal domain, the integrity of which is important for normal gametogenesis.

3295W

"They can't find anything wrong with him, yet": Mothers' experiences of parenting an infant with a prenatally diagnosed copy number variant (CNV). *B.A. Bernhardt, S.A. Walser, F.K. Barg, A. Werner-Lin. University of Pennsylvania, Philadelphia, PA.*

Chromosome microarray analysis (CMA) is used widely in prenatal settings. Some CNVs detected using CMA are associated with a variable or uncertain phenotype and/or possible neurocognitive involvement. Little is known about the experiences of parenting an infant following such findings. We conducted semi-structured interviews with 23 mothers of 6-12 month old infants diagnosed prenatally with a potentially pathogenic CNV. Questions elicited perspectives on the child's health and development, ongoing needs, and sharing results with others. Interviews were audiotaped, transcribed, and analyzed for common themes. Although most mothers reported that their infants were developing typically, the majority expressed concern about the child's future development given the CNV. They tried to reassure themselves that their child was unaffected by: comparing him/her to siblings, scrutinizing the child's appearance and behavior, or accepting provider reassurances. Even in the absence of developmental and neurological concerns, most remained vigilant about their child's development, and for some led to enrollment in early intervention or ongoing assessments by specialists. Maternal concern and vigilance were heightened if there was a positive family history of a neurodevelopmental disorder, the child had minor physical abnormalities, or a clinician expressed concern. Even mothers who were not overtly concerned stated they would likely attribute any abnormal behavior or development to the CNV. Half of the interviewees did not share the result with the child's pediatrician, relatives, or others due to fear of stigma, difficulty explaining the CNV, or presumptions that the child was unaffected. A few mothers regretted learning CMA results, yet most felt empowered to identify and address problems early. Interviewees recommended that clinicians provide regular updates about understanding of their child's CNV, access to families with a child with the same CNV, and referrals to psychological counseling as needed. These data reveal a gap between clinical involvement of a prenatally diagnosed CNV and parental interpretation and intervention to address problems that do not (yet) exist. Additional research is needed to address the long-term consequences of returning uncertain CMA results for parent-child bonding, and costs associated with ongoing assessment and/or unnecessary early intervention for typically developing children.

3296T

Profiling sncRNA expression through gestation by RNA sequencing. *E.M. Price[1,2], V. Martinez[3,4], D. Becker-Santos[3,4], I. Manokhina[1,2], W.L. Lam[3,5], W.P. Robinson[1,2].* 1) Medical Genetics, UBC, Vancouver, BC; 2) Child and Family Research Institute, Vancouver, BC; 3) BC Cancer Agency, Vancouver, BC; 4) Integrative Oncology, BCCA, Vancouver, BC; 5) Pathology and Laboratory Medicine, UBC, Vancouver, BC.

Small non-coding RNAs (sncRNA), including miRNAs (21-24 nucleotides long) and piRNAs (26-31 nucleotides long), have been recognized as important modifiers of gene expression. During pregnancy, active secretion and shedding of trophoblast releases placental sncRNAs into maternal blood circulation, which makes these placental sncRNAs attractive non-invasive biomarkers of health and disease. Though differential sncRNA expression has been described in maternal blood and placenta ascertained from complicated pregnancies, rarely are the same candidates reported by different studies. Understanding the biological and technical factors associated with sncRNA expression will be fundamental to developing robust and reproducible clinical tools. Thus in this study, the sncRNA transcriptome of 32 placentas, 6 first trimester (6-11 weeks gestation), 16 second trimester (14-24 weeks gestation) and 10 term (38-40 weeks gestation), was profiled using high-throughput small RNA sequencing. This approach allowed for detection of both piRNA and miRNA sequences by applying population-specific read-length filters. Expression levels were normalized by sequencing depth and transcript length (RPKM) and then log transformed for exploratory and differential expression analyses. Principal component analysis suggested sources of signal variation due to gestational age at delivery, total number of reads and processing time. Using samples with >1 million reads (n=30), 1446 of 2567 (56%) annotated miRNAs were expressed (RPKM >1) in more than one sample and 681 (26%) were expressed in more than 80% of samples of at least one trimester. Using this second set of expressed miRNAs, linear modelling identified 270 differentially expressed miRNAs (FDR <0.05) between trimesters, with most changes observed between the 1st and 3rd trimesters. There was a trend for change in miRNA expression with processing time, though this did not withstand correction for multiple comparisons. Our study suggests that both technical and biological factors should be considered in the study of sncRNAs.

3297F

Deciphering the impact of IL-10 gene polymorphism (-1082A/G, -592A/C and -819T/C) in preterm birth. *M. Pandey, S. Awasthi.* King George`s Medical University, Lucknow, India.

Background/ Objective: IL-10 is a key regulator for diminishing/weakening inflammatory response. Thus the objective of this study is to focus on the association of IL-10 gene and its polymorphisms with Preterm Birth(PTB). The process of spontaneous parturition in both preterm and term delivery is mediated by inflammation in the cervix, membranes, and myometrium. **Study Design:** A case-control study, conducted at Lucknow (India) included a total of 1118 subjects (559 cases and 559 controls). Cases were the mother`s who delivered preterm neonates(i.e. before 37 weeks of gestation) and controls includes mother`s with term delivery (i.e. after 37 weeks) .The demographic, clinical data was also obtained. Venous blood was used for DNA/RNA extraction. The genotyping of IL-10 (-819T/C, -1082G/A,-592C/A) was carried out and IL-10 mRNA levels was also checked in cases and controls. The polymorphisms were analyzed by polymerase chain reaction (PCR) restriction fragment length polymorphism (RFLP). The IL-10mRNA levels were analyzed according to genotype by quantitative real-time PCR (QRT-PCR) in both cases and controls Results: Significant association of IL-10 (-1082G/A) gene polymorphism was found (OR=1.727(1.70-2.5),p=0.006) in cases as compared to controls. The haplotypic analysis determined the association of IL-10 ATA (p=0.002)andATC(p=0.001) with PTB. The IL-10 mRNA levels were detected lower in preterm(mean±S.D.=3.08±2.22) as compared to term(mean±S.D.= 2.63±1.75) (p=0.05). Conclusion: IL-10 marks the protective impact in the inflammatory pathway of PTB. .

3298W

Comparing the clinical yield of carrier screening: Genotyping versus exon sequencing. *K.A. Beauchamp, G.A. Lazarin, H.P. Kang, K. Wong, E.A. Evans, J.D. Goldberg, I.S. Haque.* Counsyl, South San Francisco, CA.

Expanded carrier screening (ECS) identifies carriers of recessive diseases and may be performed using either targeted genotyping (TG) or next generation sequencing (NGS). TG typically provides genotype information on a fixed number of variants (typically below 3000), while NGS detects arbitrary variants within the region of interest, albeit at a somewhat higher cost. The relative efficacy of TG and NGS approaches is incompletely understood and depends on the choice of TG panel, the number of variants interrogated, and the ethnicity profile of the population being studied. Here we report on the de-identified, aggregated ECS results of 70,803 patients who were screened at Counsyl using NGS of up to 110 genes. Screening identified 27,850 individuals as carriers of 36,880 deleterious or likely deleterious mutations (some individuals were carriers for more than one variant). There were 3,760 unique deleterious mutations (many were present in more than one individual). A subset of these variants are detectable by the Counsyl TG panel, allowing direct comparison of the detection rates of NGS and TG approaches. Overall, the TG subset alone would have detected 63.1% of the pathogenic mutation calls identified by NGS, accounting for 69.4% of carrier individuals. However, in ethnicities that are underrepresented in typical TG panel designs (e.g. Asian populations), the TG test detects far fewer carriers (35.5% of mutations, 40.3% of carriers). As an additional comparator, we conceptualized the feasibility of a TG assay that achieves a high detection rate (DR). To achieve 90% DR in the pan-ethnic population requires a panel of 834 mutations. But reaching 95% DR more than doubles that panel size to 1,916 mutations. Even so, DR is not uniform across ethnicities: the 834-site panel only achieves 78% DR in Asian populations. Furthermore, evaluating this panel on patients not used during variant selection suggests additional benefits of NGS on the order of 3-5%. These data demonstrate superior sensitivities of NGS over TG for pan-ethnic ECS.

3299T

Cell based non-invasive prenatal testing using NGS-based copy number assessment of fetal cells. *A. Breman[1], L. Vossaert[1], E. Normand[1], J. Chow[3], L. U'Ren[3], R. Salman[1], S. Qdaisat[1], I. Van den Veyver[1,2], C. Shaw[1], Y. Yang[1], E. Chang[3], J. Stilwell[3], R. Seubert[3], E. Kaldjian[3], A. Beaudet[1].* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 3) RareCyte Inc., Seattle, WA.

Background: Detection of genomic copy number abnormalities from circulating single fetal cells using next generation sequencing (NGS) offers a promising non-invasive alternative for prenatal diagnosis. Towards this goal, we have established a method for performing fetal cell-based, non-invasive prenatal testing (CB-NIPT) during the first trimester. CB-NIPT for prenatal diagnosis has dramatic potential advantages over the currently available cell-free DNA-based tests because it enables analysis of pure fetal DNA. Here we show that we can successfully repeatedly recover individual fetal cells during the first trimester and perform NGS to detect clinically important copy number variants. **Methods:** Fetal cell enrichment was carried out using methods developed by the commercial author organization for blood preservation, density based enrichment, immunostaining, custom high-resolution scanning and analysis, and integrated single-cell picking. Whole genome amplification was performed on recovered single fetal cells and single nucleotide polymorphism-based genotyping studies were carried out for confirmation of fetal origin. NGS on an Illumina platform with approximately 5 million reads per cell (~0.1x haploid genome) was used to generate genome-wide copy number data. **Results:** Our data indicate that we can recover between 2 and 12 fetal cells with a trophoblast phenotype from 20-30 ml of maternal blood in most cases. NGS data from isolated single fetal cells has thus far shown concordance with diagnostic array CGH data from amniotic fluid or chorionic villus samplings in pregnancies affected with trisomy 21, trisomy 18, 47,XXY, and in one case a 2.7 Mb deletion on chromosome 15. **Conclusion:** CB-NIPT has dramatic potential advantages over cell-free DNA-based testing primarily because of the ability to analyze fetal DNA free of contamination by maternal DNA and avoid detecting maternal findings. CB-NIPT has the potential to detect most clinically significant cytogenetic abnormalities and even, in the future, *de novo* deleterious point mutations. Optimization of circulating fetal cell recovery and validation studies on larger numbers of samples from pregnant women are underway to evaluate the clinical validity of this test.

3300F

What Quebec pregnant women think about prenatal testing and their responsibility in the context of prenatal personalized medicine. *G. Lapointe[1], R. Drouin[2,3], C. Bouffard[1].* 1) Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) Univeristé Laval, CHU de Québec, Québec, Canada; 3) Université du Québec à Montréal, Montréal, Canada.

The rapid developments in genomics portends to a future where it will be possible to predict and prevent diseases. The combination of new genomics tools and P4 medicine in the field of prenatal medicine – what we call "prenatal personalized medicine" (PPM) – will produce a large quantity of genomic information on the fetus. PPM encourages parents to actively participate in prenatal care and commits an unprecedented emphasis on the health of the fetus. This responsibility, combined with the explosion of genetic information, the advent of non-invasive prenatal tests (NIPT) and direct-to-consumer genetic testing, exposes PPM to several pitfalls. Our objective is to develop knowledge on the views and needs of pregnant women about current and new prenatal tests and on their responsibility regarding those tests. **METHODOLOGY**: Descriptive qualitative design and thematic analysis of 15 semi-structured interviews with pregnant women without a known risk of genetic disease, and previously briefed on available and forthcoming prenatal tests. **RESULTS**: At first, the majority of participants tended to choose NIPT with fetal cells, since it is a broad test, without risk. Several participants also wished to have an ultrasound, since they considered it important to see the baby. The possibility of undergoing NIPT early during pregnancy is seen as both an advantage and a disadvantage, and most agree that it should be accompanied by a doctor. Pregnant women want to actively become informed during pregnancy in order to increase their knowledge and reduce their concerns. They trust the expertise of their doctors, but find it important to become informed through a variety of other sources (e.g., the internet). They want to be responsible for what they can control during pregnancy (e.g., tobacco use), but feel that it is incorrect to require that pregnant women be responsible for what is beyond their control (e.g., genetic disease). **CONCLUSION:** Participants will opt for a new and safe technology that allows them to learn more about the health of their child. Although, these technological advances seem desirable, the majority of participants agree that these technologies should be used wisely, without putting too much responsibility on the shoulders of parents. .

3301W

Two cases of genetic counseling on additional chromosomes detected by karyotype analysis (G-banding) of amniotic fluid cells. *N. Takeshita, A. Takashima, T. Adachi, M. Nagaoka, M. Ishikawa, I. Sasaki, K. Yokokawa, M. Urita, A. Yokoyama, T. Kinoshita.* Obstetrics and Gynecology, Toho University Sakura Medical Center, Chiba, Japan.

[Introduction] Chromosomal karyotype analysis of amniotic fluid cells is a definitive prenatal genetic test. This report deals with points to be considered in conducting genetic counseling based on our experience in 2 cases in which additional chromosomes were detected by chromosomal karyotype analysis (G-banding) of amniotic fluid cells. **[Case 1]** A 32-year-old , became pregnant after receiving in vitro fertilisation and embryo transfer (IVF-ET) . Ultrasonography (US) examination performed at week 13 of gestation disclosed a fetal cystic hygroma on the neck. The patient was given genetic counseling by a clinical genetics specialist regarding potential causes of abnormalities including chromosomal aneuploidy and structural aberrations, and, at the request of the couple in part, amniocentesis was performed at week 16 of gestation. At week 18, results of the test showed the karyotype to be 46,XX,add(18)(p11.3). The patient gave birth to a female child weighing 3014 g . Based on the infant's countenance and features of her limbs, we suspected the possibility of 18 monosomy. The parents were offered an explanation again postnatally which included the possibility of clarifying the origin of the event, and they requested the newborn's karyotype analysis which was found to be 46,XX-,der(18)t(15;18)(q24;11.3). **[Case 2]** A 41-year-old . She underwent amniocentesis at week 16 of gestation on account of her pregnancy at an advanced maternal age. At week 18, results of the test showed the fetal karyotype to be 46,XY,add(9)(p24). Parental chromosomal analysis and detailed analysis with single nucleotide polymorphism (SNP) microarray procedure to locate the site of addition were also carried out at the parents' request. As a result, the aberrations were found to be arr Xp22p22.13(16,992,941-17,729,022)´2 and Xp21.3(28,791,782-28,903,443)´2, both being thought to be *de novo*. The mother was delivered of a male child weighing 2750 g . No anomaly was noted, nor did she request any postnatal chromosomal examination. **[Discussion]** In the case of an additional chromosome, it is not always easy to clearly evaluate its phenotype due to differences in the originating chromosome and the magnitude of the site of chromosomal addition. If an accurate explanation of the results of testing/observations is not provided to the parents, it may cause great anxiety. This underlines the vital importance of genetic counseling. .

3302T

The establishment of a new leaflet for prenatal diagnosis as an approach to prenatal genetic counseling. *T. Yamada[1,13], K. Sameshima[2,13], H. Sawai[3,13], A. Sekizawa[4,13], S. Nakagomi[5,13], K. Hayata[6,13], Y. Yamanouchi[7,13], Y. Fujii[8,13], H. Miyake[9,13], S. Yamada[10,13], Y. Fukushima[11,13], I. Konishi[12,13].* 1) Department of Obstetrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 2) Department of Pediatrics, Minamikyusyu National Hospital, Aira, Japan; 3) Department of Obstetrics and Gynecology, Hyogo College of Medicine, Nishinomiya, Japan; 4) Department of Obstetrics and Gynecology, Showa University School of Medicine, Tokyo, Japan; 5) University of Yamanashi ;Graduate Faculty of Interdisciplinary Research, Chuo, Japan; 6) Department of Obstetrics and Gynecology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Science, Okayama, Japan; 7) Department of Medical Welfare, Kawasaki University of Medical Welfare, Kurashiki, Japan; 8) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 9) Clinical Genetics Unit, Kyoto University Hospital, Kyoto, Japan; 10) Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 11) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 12) Department of Gynecology and Obstetrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; 13) The Project of Health Research on Children, Youth and Families by Ministry of Health, Labour and Welfare, Japan (2014-2016).

[Background and Purpose] In Japan, "informed choice" has always been a controversial topic, as providing information is sometimes regarded as an act of misleading those with less knowledge. Many Japanese people believe that information on prenatal diagnosis (PD) pressures pregnant women into getting tested; therefore, it is difficult to speak openly about PD. Studies show that enough information is not reaching clients in primary obstetrics clinics because of this misconception. In an attempt to standardize genetic medicine in primary obstetrics clinics in Japan, we drafted a leaflet as a tool that provides balanced information of PD and directs clients to genetic counseling without pressuring them into prenatal testing. The objective of this study was to determine the impression of this leaflet on medical staff, pregnant women, and families of pregnant women and also to improve its contents. [Methods] A questionnaire was administered between November and December 2015 after using this leaflet at 5 institutions of study members throughout Japan. Data regarding the impression of the medical staff and pregnant women and/or family members were collected. This study was funded by Health and Labor Sciences Research Grant (MHLW14428175), and was approved by the institutional review board of each institution. [Results] A total of 51% of medical staff (n = 751) and 46% percent of pregnant women and family members (n = 366) responded to the questionnaire. When asked how they felt after reading the leaflets, 21.2% (119/382) of medical staff answered negatively and 34% (131/382) of them answered favorably. Furthermore, 24.1% (41/170) of pregnant women and family answered negatively and 44.7% (76/170) of them answered favorably. When pregnant women and family members were asked about the change in their feeling of happiness after reading the leaflet, 87.6% (149/170) indicated there was no change and 8.2% (14/170) answered negatively. When asked whether they felt the leaflets recommend PD, 32.5% (124/382) of medical staff and 12.9% (22/170) of pregnant women and family answered yes. [Conclusions] We conclude from this survey that this leaflet provides fair information to pregnant women and family, while medical staff tend to take a cautious attitude. There was a large variety of opinion on this leaflet because of the diversity of attitude and emotion towards PD. This leaflet should be used as a tool to direct pregnant women and family to genetic counseling. .

3303F

Next generation screening: TruSight One "clinical exome" for preconception screening in consanguineous couples. *E.P. Kirk[1,2,3], K. Barlow-Stewart[4], A. Selvanathan[5], S. Josephi-Taylor[6], L. Worgan[5], S. Rajagopalan[5], A. Bittles[7], L. Burnett[3], M. Buckley[3], A. Colley[5], T. Roscioli[1,8].* 1) Dept of Medical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia; 2) School of Women's and Children's Health, University of New South Wales, Randwick, NSW, Australia; 3) SEALS Genetics Laboratory, Randwick, NSW, Australia; 4) Northern Clinical School and School of Public Health, University of Sydney, St Leonards, NSW, Australia; 5) Dept of Clinical Genetics, Liverpool Hospital, Liverpool, NSW, Australia; 6) Dept of Clinical Genetics, Royal North Shore Hospital, St Leonards, NSW, Australia; 7) Centre for Comparative Genomics, Murdoch University, Murdoch, WA, Australia; 8) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia.

Background: Autosomal recessive (AR) disorders are a major cause of morbidity and mortality. Consanguineous couples are at increased risk for having a child with these disorders but to date, risk assessment based on family history and/or ethnicity often added little to that based on empiric observational data. Therefore exome (and in the future, whole genome) sequencing offers the prospect of "the screen for everything", providing information which can inform reproductive choice in this at-risk group. More broadly, it may provide a paradigm for assessment of the practicability and effectiveness of this screening approach in the general population. **Study design:** We used the TruSight One "clinical exome", which includes 4,813 genes linked to human disease, to pilot screen 15 consanguineous couples who were planning a pregnancy. **Results:** Three couples were found to be at risk of having children affected by AR disorders. In two couples, the risk was for disorders with neurodevelopmental phenotypes. Variants were identified in *C5ORF42* (Joubert syndrome/ Orofaciodigital syndrome type VI) and *LAMA1* (Poretti-Boltshauser syndrome). In the third couple, the risk was for glycogen synthase deficiency (*GYS2*) - a metabolic disorder which is treatable but can be difficult to diagnose, and can have severe consequences for an untreated child due to hypoglycemic episodes. Interviews conducted with the couples pre-test identified an intention to utilize any positive findings in their future pregnancy planning. **Conclusion:** Variant interpretation remains a challenge which will limit the sensitivity of this approach. Nonetheless, our results indicate that the benefits of screening are likely to be considerable, and a major advance over a targeted approach.

3304W

Implementing a strategy for massively parallel functional annotation of
HNF1A **missense variants.** *S. Althari[1,2], L.A. Najmi[3,4], A.J. Bennett[1], M. van de Bunt[1,2], P.R. Njolstad[4,5], A.R. Majithia[6,7,8,9], M.I. McCarthy[1,2,10], A.L. Gloyn[1,2,10].* 1) Oxford Centre for Diabetes, Endocrinology and Metabolism, University, of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 4) KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 5) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway; 6) Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 7) Diabetes Research Center, Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 8) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 9) Department of Medicine, Harvard Medical School, Boston, MA, USA; 10) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Loss-of-function variants in *HNF1A* are the most common cause of maturity-onset diabetes of the young (MODY), a Mendelian form of hyperglycaemia where a molecular diagnosis influences treatment and prognosis. Recently, *HNF1A* missense variants have been associated with type 2 diabetes (T2D) predisposition and identified in the genomes of healthy individuals. Sensitive mapping of *HNF1A* alleles to clinical phenotype requires comprehensive understanding of the functional consequences of known and novel sequence variants. To address this clinical challenge, we are implementing a strategy for massively parallel functional screening of every possible missense variant in *HNF1A*. Successful implementation of the approach depends on having robust downstream HNF-1a targets to act as sensitive discriminators of molecular dysfunction. These have not been previously identified for the purpose of high-throughput allelic screening. Therefore, we have conducted a systematic search for HNF-1a regulated transcripts to identify potential markers of trans-activation in a physiologically relevant cell model. On the basis of cytogenetic, transcript and proteomic characterisation, the human liver cell line HepG2 was selected. HepG2 cells were infected with *HNF1A*-harboring lentivirus and exposed to a range of doxycycline (dox) doses (0- 5 μg/ml) over a 48-hr period. Having achieved dose dependent induction of *HNF1A*, with highest expression at 5 μg/ml dox, target identification was performed using qPCR and RNA-seq. The former involved probing of targets based on evidence in the literature of direct regulation by HNF-1a (11 genes), including *TMEM27* (dose-dependent upregulation)*, IGFBP1* (dose responsive downregulation), *SLC2A2* and *PAH* (unaffected). Transcriptome profiling revealed 22 transcripts showing significant differential expression (FDR corrected *P* value<0.01, 1-fold minimum induction) as a function of exogenous *HNF1A* (*P*=1.97e-09). The top two hits, *CUBN* (*P*=1.47e-07) and *CLDN2* (*P*=2.58e-06), have not been previously reported as HNF-1a targets. Preliminary motif analysis revealed several *HNF1A* binding sites 5-10 kb upstream *CUBN* and *CLDN2* (80% match score), with one site per gene overlapping DNase hypersensitivity regions in HepG2 cells. In summary, we have identified HNF-1a regulated transcripts in HepG2 cells including novel target transcripts which are currently being validated *in silico* and *in vitro* prior to their evaluation as markers for high-throughput variant screening.

3305T

Single-cell RNA-seq analysis of human pancreatic islets reveals novel gene signature for different cell types.
M. Garieri[1,2], C. Borel[1], T. Stoeger[3], F. Santoni[1], C. Howald[1,2], E. Falconnet[1], P. Ribaux[1], L. Pelkmans[3], P.A. Halban[1], S.E. Antonarakis[1,2], E.T. Dermitzakis[1,2]. 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland; 3) Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland.

Pancreatic islets consist of several endocrine cell types. Single Cell (SC) transcriptome analyses provide an unprecedented opportunity to discover the heterogeneity of human pancreatic islets and as well the possibility to characterize gene signature for each cell population, which would be difficult to isolate with other methods. We investigate the heterogeneity of this tissue using SC RNA-seq. Dissociated islet cells of 3 donors (n=241, 183 after quality control) were analyzed. We also sequenced sorted-beta SCs (n=60, 44 after quality control) from the same individuals. Clustering analysis on islet SCs identified the following populations of cells: beta (mean *INS* expression 40777 RPKM, representing the 75% of all SCs), alpha (mean *GCG* expression 86443 RPKM, 12%), gamma (mean *PPY* expression 49928 RPKM, 3%), delta (mean *SST* expression 76061 RPKM, 2%) and a small population of exocrine. Out of these we identify 10 ductal and 2 acinar cell. In the sorted beta SCs we observed beta (INS 32708 RPKM, 93%) and delta SCs (SST 135327 RPKM, 7%). Interestingly we observed that insulin (*INS*) is expressed at appreciable levels in all the islets SCs and not just in the subpopulation of beta SCs described above. In addition *GCC* has expression in 93% of all islets SCs at appreciable levels. We controlled for the absence of transcript contamination of *INS* and *GCG* by sequencing 11 empty chambers. We verify the cellularity of the cell types using another technology (branched-DNA single-molecule FISH) allowing for quantification and cellular localization of transcripts. Standard differential expression approaches are not appropriate to identify gene markers for each cell population, while Fisher's exact test allow to identify known and novel genes and transcription factors. For the alpha SCs we identify 36 genes at 0.1% FDR: two transcription factors (*IRX2*, *ARX*), and other markers (i.e. *GC*, *SPOCK3*, *PAPPA2*, *FAP*, *LOXL4*). For the beta SCs we have 95 genes at 0.1% FDR: two transcription factors (*ZNF91*, *DDIT3*), four imprinted genes (*IGF2*, *MEG3*, *DLK1*, *CDKN1C*) and other markers (*HADH*, *RBP4*, *PCSK1*). At 0.1% FDR, we identify 80 genes for exocrine cells and 88 genes for endocrine cells with an enrichment in *SCG* gene family, precursors for biologically active peptides. We are using these gene signatures to further characterize what appears to be a continuity of cell identity among cells in the pancreatic islets. M.G and C.B. and the labs of SEA and ETD contributed equally to the work.

3306F

Modulation of gut microbiota by GLP-1 receptor agonist in the high-fat diet mice. *H. Li, Y. Ran, J. Jin, X. Yang, T. Zhou, Z. Deng.* Dept. Biotechnology, Jinan University, Guangzhou, Guangdong, China.

Accumulating evidence suggests that the gut microbiota is an important factor in mediating the development of obesity-related metabolic disorders, including type 2 diabetes. Glucagon-like peptide-1 (GLP-1) receptor agonists are indicated for treatment of type II diabetes since they mimic the actions of native GLP-1 on pancreatic islet cells, stimulating insulin release, while inhibiting glucagon release, in a glucose-dependent manner, as well as, involve communication with the central and peripheral nervous system, slowing gastric empty, reducing food intake, and producing weight loss. High-fat diet has been shown to cause shifts in the diversity of dominant gut bacteria. In this study, We measured shifts in cecal bacterial communities in mice fed high-fat (HF) diet+PBS or high-fat (HF) diet+ GLP-1receptor agonist for 8 weeks at the level of the diversity and taxa distribution by high-throughput 16S ribosomal RNA gene sequencing. The diversity of gut microbiota was significantly changed by GLP-1receptor agonist treatments. And the HF diet caused an increase in body weight over the 8-week-long feeding trial, whereas GLP-1receptor agonist slowed the increase in body weight.

3307W

The role of *CYP21A2* gene in Premature Pubarche (PP). *M. Soveizi[1,2], A. Rabbani[3,4], A. Setoodeh[3,4], F. Sayarifard[3,4], N. Mahdieh[2], B. Rabbani[2,4].* 1) Azad University, North Tehran Branch, Tehran, Iran, Tehran, Iran; 2) 2. Genetic Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran; 3) Children's Hospital Center, Tehran University of Medical Center; 4) Growth and Development Research, Tehran University of Medical Center.

AbstractPremature pubarche (PP) is considered to be a benign phenomenon; it follows the precocious appearance of pubic hair without other signs of puberty or virilization which diagnosed after exclusion of non-classical (NC) congenital adrenal hyperplasia (CAH). The aim of this study was to assess the relationship between the CYP21A2 genotype and children with sings of premature pubarche. Coding and non-coding regions of the CYP21A2 gene were directly sequenced. Our result showed that, in one of the case pathogenic mutation located in exon 7 at the position c.754G>T (Val252Leu) was observed. This means that only 10% of our patients showed CYP21A2 gene mutations. This indicates that there is relation to PP as discussed in literature. We are planning to sequence whole exome to find out the causal genes of PP phenotype. We need to expand the population for more accurate result and analyze other related genes.

3308T

Obesity in adults with 22q11.2 deletion syndrome. *A.S. Bassett[1,2,3,4,5], S. Voll[2], E. Boot[1,2,3], N. Butcher[2], S. Cooper[3], T. Heung[2], E.W.C. Chow[1,2], C. Silversides[3,5].* 1) Dept Psychiatry, CAMH, RS, University of Toronto, Toronto, ON, Canada; 2) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 3) The Dalglish Family 22q Clinic, Department of Psychiatry, University Health Network, Toronto, Ontario, Canada; 4) Toronto General Research Institute, University Health Network, Toronto, Ontario, Canada; 5) Division of Cardiology, Department of Medicine, University Health Network, Toronto, Ontario, Canada.

Purpose: To characterize the prevalence of and contributing factors to adult obesity in the most common recurrent copy number variation (CNV), 22q11.2 deletion, given that other rare CNVs are known to have obesity phenotypes. **Methods:** For 207 adults with 22q11.2 deletion syndrome (22q11.2DS), we used available height and weight measurements to calculate body mass index (BMI) and recorded associated factors that could play a role in obesity. We used maximum BMI per subject and logistic regression to test a model predicting obesity class. **Results:** The prevalence of obesity (BMI≥30) in 22q11.2DS (n=90, 43.5%; at median age 26.7 years) was significantly greater than for Canadian norms (OR 2.30, 95%CI=1.74-3.02, p<0.0001), even after excluding individuals with a history of antipsychotic use. The regression model was significant (*P*<0.0001). Psychotropic medication use and age, but not sex or presence of intellectual disability, were associated with higher obesity level. Ten (4.8%) individuals were diagnosed with type 2 diabetes at median age 39.5 years; prevalence was higher in those with obesity (*P*<0.01). **Conclusion:** The results suggest that adult obesity is related to the 22q11.2 deletion. The findings expand the potential genetic causes of obesity and have important implications for management of 22q11.2DS.

3309F

Mutation screen of *SLC35D3* in Chinese Han patients with metabolic syndrome. *Z. Zhang.* Qingdao University, Qingdao, China.

Metabolic syndrome (MetS) is a severe health problem that is characterized by Obesity, hyperglycemia,;hyperlipoidemia,;hypertension and other metabolic disorders. Studies suggest that large numbers of genes disorder cause MetS. We screened 363 Chinese Han patients with MetS and 217 unaffected individuals by sequencing the two exons and adjacent exon/intron boundaries together with 1 kb untranslated sequence upstream of the start codon of the *SLC35D3* gene. Two variants of SLC35D3 leading to the frame-shift of the coding sequence were found in two unrelated patients. In patient #1 (Male, Age: 55, BMI: 26.1, waist circumference: 109 cm, blood pressure: 135/85 mmHg, TG: 4.23 mmol/L, Chol: 5.28 mmol/L, Gluc: 4.4 mmol/L), a heterozygous ΔK404 was identified. In patient #2 (Male, Age: 51, BMI: 27.1, waist circumference: 100 cm, blood pressure: 120/80 mmHg, TG: 2.52 mmol/L, Chol: 5.94 mmol/L, Gluc: 5.2 mmol/L), a heterozygous insL201 was identified. In summary, we have identified 2 mutations of the *SLC35D3* genes in 363 MetS patients.

3310W

Chromosome 22 presents a higher density of CNVs among Brazilian patients with rare phenotypes. *C. Milani[1], J.G Damasceno[1], T.V.M.M Costa[1], F.A.R. Madia[1], E.A. Zanardo[1], A.T. Dias[1], M.M. Montenegro[1], G.M. Novo-Filho[1], F.B. Piazzon[1], D.C. Soares[3], A.M. Nascimento[1], M. Rocha[1,2], L.R.R.V Souza[1], M.S. Grassi[3], A.F.L. Carvalho[5], D. Bertola[3], M.I. Melaragno[4], C.A. Kim[3], M.M.S. Carneiro-Sampaio[3], L.D. Kulikowski[1,2].* 1) Laboratório de Citogenomica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo, Brasil; 2) Centro de Reprodução Humana e Genética, Departamento de Saúde Coletiva, Faculdade de Medicina do ABC, Santo André, Brasil; 3) Unidade de Genética, Departamento de Pediatria, Instituto da Criança, HC/FMUSP, São Paulo, Brasil; 4) Departamento de Morfologia e Genetica da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil; 5) Departamento de Biologia Geral, Laboratorio de Genetica Humana e Mutagenese, Universidade Federal da Bahia, Salvador Brasil.

A novel pattern of copy number variations was defined in a cohort of Brazilian patients from Cytogenomic Laboratory of FMUSP, identifying common CNVs and LOH correlated to rare phenotypes. **Materials and Methods:** A total of ninety eight (98) patients were analyzed using using SNP-array (Human-CytoSNP-12 BeadChip and CytoSNP-850K BeadChip) to map the presence of Copy Number Variations (CNVs) and Loss of Heterozygosity (LOH). We use bioinformatics techniques in order to construct a genomic profile of Brazilian patients. Thirty Brazilian patients with a spectrum of 22q11.2 phenotypic manifestations were further studied in order to stablish better chromosome breakpoint definition. **Results:** We identified 13 deletions, 14 duplications and nine regions with loss of heterozygosity. Also, our results showed atypical deletions with two hotspots: 22q11.21 and 22q13.32. Pathogenic CNVs associated with lupus erythematosus, polyarthritis, psoriasis, rheumatoid arthritis, sepsis, bronchitis and pneumonia and several different immunological features were found. The clinical evidence suggests that the immunodeficiency can be seen in most of the patients with chromosome 22q genomic alterations. Immunological phenotype may vary widely between the patients and the deletion size does not appear to be responsible for this variation. **Conclusion:** Cytogenomic investigation allowed a better chromosome breakpoint definition in a cohort of Brazilian patients suggesting that identification of deletion and duplication syndromes in 22q region enables a biological interpretation of the immunologic abnormalities. **Grants:** FAPESP 14/50489-9, CNPq 09/53105-9 and FINEP-CT INFRA 0160/12 SP8.

3311T

Genome-wide copy number variation analysis identifies leukocyte-specific protein 1 deletion vatiant for rheumatoid arthritis susceptibility. *S. Jung[1], S. Hwang[2], H. Jo[1], W. Kim[2], Y. Chung[1].* 1) Integrated Research Center for Genome Polymorphism, The Catholic University of Korea, Seoul, Korea; 2) POSTECH-Catholic Biomedical Engineering Institute, The Catholic University of Korea, Seoul, Korea.

Several copy number variations (CNVs) have been found to be associated with rheumatoid arthritis (RA). However, it remains unclear how they affect immune dysfunction and autoimmune diseases. We screened RA-associated CNVs across the whole genome and identified a novel leukocyte-specific protein 1 (LSP1) deletion variant for RA susceptibility located in 11p15.5. We replicated that the copy number of LSP1 gene is significantly lower in patients with RA, which correlates positively with LSP1 protein expression levels. Differentially expressed genes in Lsp1-deficient primary T cells represent cell motility and immune and cytokine responses. Functional assays revealed that LSP1, induced by T-cell receptor activation, negatively regulates T-cell migration by reducing ERK activation in vitro. Loss of Lsp1 promotes T-cell migration into antigen-instilled tissues and draining lymph nodes in mice with T-cell–dependent chronic inflammation. Moreover, patients with RA show diminished expression of LSP1 in peripheral T cells with increased migratory capacity, suggesting that the defect in LSP1 signaling lowers the threshold for T-cell activation. To our knowledge, our work is the first to demonstrate how CNVs result in immune dysfunction and a disease phenotype. Particularly, our data highlight the importance of LSP1 CNVs and LSP1 insufficiency in the pathogenesis of RA and provide previously unidentified insights into the mechanisms underlying T-cell migration toward the inflamed synovium in RA.

3312F

Accurate Rh blood-typing from next generation sequencing data. *M.M. Wheeler[1], H. Huston[2], K.W. Lannert[2], T. Shaffer[3], S. Fletcher[2], J.G. Underwood[1], M. Delaney[2,4], D.A. Nickerson[1], J.M. Johnsen[2,5].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Bloodworks NW, Seattle WA; 3) Resolution Bioscience, 2023 120th Ave NE Bellevue, WA; 4) Laboratory Medicine, University of Washington School of Medicine,Seattle, WA; 5) Division of Hematology, University of Washington School of Medicine, Seattle, WA.

Antibody-based blood typing is one of the mainstays of precision medicine. However, newer genomic technologies have the potential to transform transfusion medicine by delivering high resolution DNA variants in clinically relevant blood group genes. Genetic variation is of particular importance in the highly immunogenic Rh blood group system, where genetic variants in the *RHD* and *RHCE* genes can alter blood group protein or antigen expression. Characterization of Rh remains challenging because the *RHD* and *RHCE* genes exhibit diverse types of genetic variation, including single nucleotide variants (SNVs), whole gene deletions, duplications, and complex rearrangements. Genotyping of the RH locus must accurately identify all forms of genetic variation in a systematic and unbiased manner. To achieve this, we developed a next generation sequencing (NGS) approach and applied this to 1135 blood donors, previously phenotyped by serology for Rh (D and C antigens). For analysis, we assessed SNVs using the GATK HaplotypeCaller pipeline and assessed structural variants (SVs) using a customized read-depth based approach. Overall, our analyses show high concordance between NGS Rh predicted antigens and serology (D antigen: 99.6% and C antigen: 99.0%). Moreover, SNV analysis detected 10 single nucleotide variants indicative of rare *RHD* alleles, as well as a novel frameshift *RHD* variant in an individual typed as D- by serology. All other D- individuals exhibited a known *RHD* whole gene deletion, while all serology C+ individuals contained a known *RHD*-to-*RHCE* gene conversion event. Finally, SV analyses identified 22 individuals with larger recombination events indicative of *RHD*-*RHCE* hybrid alleles; a subset of these hybrid alleles were validated by quantitative multiplex PCR of short fluorescent fragment analysis. Our results demonstrate that genotype-based, high resolution blood-typing of Rh is possible and can accurately detect known and novel clinically-relevant variation at the *RH* locus.

3313W

Clinical presentation and genotype-phenotype correlation analysis of *de novo* **large deletion including human long non-coding RNA** *TERC*. *G. Aubert[1,2], M. McKinnon[3,4], J. Davis[3,5], P. Eydoux[3,6], S. Vercauteren[2,3,6].* 1) Terry Fox Laroratory BC Cancer Agency , Vancouver, BC, Canada; 2) Repeat Diagnostics, North Vancouver, BC, Canada; 3) Child and Family Research Institute, Vancouver, Canada; 4) Department of Medical genetics, University of British Columbia, Vancouver, Canada; 5) Department of Hematology and Oncology University of British Columbia, Vancouver, Canada; 6) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada.

Dyskeratosis Congenita (DC) is a complex trait marrow failure syndrome characterized by pancytopenia and classically with skin and nail abnormalities and oral leukoplakia. The hallmark of this disease is very short telomere length in peripheral blood leukocytes. Telomere biology gene haploinsufficiency is one of the major disease mechanisms, directly implicating gene dosage in disease pathogenesis and severity. Abnormalities in 11 different genes have been shown to be causative in this rare disorder, including the long non-coding telomerase RNA *TERC*. Here we describe the clinical case of a patient who presented with abnormal bruising at 22 months. Review of clinical information revealed the patient displayed *in utero* growth restriction and was born premature with pancytopenia and macrocytosis. Peri-natal care included platelet transfusion but DC was not suspected at the time. Hemato-pathology investigation at 22 months revealed a hypocellular and fibrotic marrow with macrocytosis, abnormal megakaryocytes and elevated hemoglobin-F levels suggesting an inherited marrow failure syndrome. Peripheral blood leucocyte telomere length measurements were found to be very short, consistent with diagnostic findings in patients with DC. Cytogenetic studies included a positive chromosome breakage assay with detection of radial chromosomes suggestive of genomic instability. Karyotyping on peripheral blood lymphocytes revealed a large single copy deletion on chromosome 3q, however no chromosome fusion or translocation events were noted. Chromosomal array analysis confirmed the *de novo* 3.8Mb deletion includes the full *TERC* gene locus, as well as 30 other genes. Possible anticipation or inheritance of short telomeres were excluded as contributing factors since both parents had leucocyte telomere length within normal range for their age. Hematological symptoms were corrected by successful reduced intensity hematopoietic stem cell transplantation using a matched unrelated cord blood stem cell source. The patient is currently doing well more than 2 years later. Laboratory and genetic findings pertinent to this patient case of early telomere biology disease onset will be discussed in relation to genotype-phenotype and will be compared to other previously published cases with *TERC* haploinsufficiency. The importance of considering copy number variation analysis in the era of clinical next generation exome sequencing for the investigation of rare disorders will be highlighted.

3314T

Next generation sequencing for an asymptomatic case of Myelodysplastic syndromes (7qdel) from Pakistan. *S. Shahid, N. Irfan, S. Ahmed, M. Nadeem, T. Shamsi.* National Institute of blood diseases and bone marrow transplantation, Karachi , Sindh, Pakistan.

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal hematologic malignancies characterized by impaired hematopoietic differentiation. Several studies on the molecular pathogenesis of MDS have identified disease causing alleles in patients but still their pathological contributions are not completely understood. Recent whole genome and targeted gene studies have found novel somatic mutations in MDS patients with pathological importance. The aim of this study was the mutational profiling of MDS patient to identify genetic marker in the genes involved in the regulation of histone function (*EZH2*, *ASXL1*, and *UTX*) and DNA methylation (*DNMT3A*, *IDH1/IDH2*, and *TET2*) that can be used in diseases prognosis and as well in deciding therapeutic options. Here we report a case of a 57 year old man referred to our institute, National Institute of Blood Diseases and Bone Marrow Transplantation (NIBD), Karachi, Pakistan. This patient was diagnosed as MDS and sub categorized as refractory cytopenia with multilineage dysplasia (RCMD) based on the morphological and cytogenetic analyses (46, XY, del (7q) (15). Patient was screened through next generation sequencing using myeloid sequencing panel of 54 genes comprising tumor suppressor genes and oncogenic hotspots. Data analysis was performed on instrument MiSeq reporter software. We identified a nonsense heterozygous mutation (c.3115C>T, p. Q1039X) in exon 13 of *ASXL1* gene. In addition, missense variant c.86 C>G, p.P29R (rs 12498609) in *TET2* gene was identified but this variant lies in non-conserved region and so not regarded as true missense mutation. Mutations in *ASXL1* gene is commonly found in advance stages of MDS and are associated with poor prognosis with overall inferior survival. But our study finds the better overall survival with good prognosis which is in contrast to reported cases. Hence screening of large number of patients is required to understand the mechanisms causing MDS pathogenesis. .

3315F

Deep targeted sequencing reveals potential causal alleles at SLE risk loci. *P. Raj[1], R. Song[1], B.E. Wakeland[1], K. Viswanathan[1], C. Arana[1], C. Liang[1], B. Zhang[1], F. Carr-Johnson[1], J.A. Kelly[2], B.R. Lauwerys[3], N. Olsen[4], C.K. Garcia[5], C. Wise[6], S.K. Nath[2], J.A. James[2], C.O. Jacobs[7], B.P. Tsao[8], D.R. Karp[9], Q.Z. Li[1], P.M. Gaffney[2], E.K. Wakeland[1].* 1) Immunology, UT Southwestern Medical Center, Dallas, TX; 2) Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, OK; 3) Pôle de pathologies rhumatismales, Institut de Recherche Expérimentale et Clinique,Brussels, Belgium; 4) Division of Rheumatology, Department of Medicine, Penn State Medical School, PA, USA; 5) Eugene McDermott Center for Human Growth & Development, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75335 USA; 6) Eugene McDermott Center for Human Growth & Development, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75335 USA; 7) Department of Medicine, University of Southern California, Los Angeles, CA; 8) Department of Medicine, Medical University of South Carolina; 9) Rheumatic Diseases Division, University of Texas Southwestern Medical Center.

Systemic lupus erythematosus (SLE) is an autoimmune disease caused by loss of humoral immune tolerance leading to the production of autoantibodies to a spectrum of self-antigens. Genetic predisposition is key for SLE susceptibility, however little is known about the magnitude and functional properties of causal genetic variants. We used targeted population sequencing approach to comprehensively characterize genetic variability at 28 risk loci for SLE in a panel of 1349 Caucasian SLE cases and controls. Overall, vast majority of SLE associated variants were discovered in non-coding region of genome. We analyzed 16 common risk loci in detail and formed haplotypes based on SLE associated potentially regulatory variants in strong LD with GWAS tag SNPs. We identified 88 potentially causal alleles embedded in these regulatory haplotypes. Based on ENCODE data these alleles are highly likely to impact binding of important transcription factors and regulate transcription of cis-genes (cis-eQTL) in human monocytes, dendritic cells (DCs) and macrophages. About 30 percent of these alleles were less common in population (MAF below 0.15), as a result, were rarely analyzed in SNP array based investigations in past. We developed a panel of over 100 normal human monocyte derived macrophages and dendritic cells to study the effect of these alleles on gene expression by RNAseq analysis. Such regulatory alleles at HLA locus upregulates transcription and cell surface level of HLA class II molecules in DCs. Similarly, at IRF5 locus, these alleles upregulates expression of IRF5 in Macrophages and LCLs. Based on the analysis of these regulatory alleles at risk loci, we identified transcriptional complexes that are most likely to be affected by these genetic changes which results in altered transcription of local genes and development of disease endophenotypes.

3316W

Shared and specific features of the individual transcriptomic response in sepsis due to faecal peritonitis. *K.L. Burnham[1], E.E. Davenport[1], J. Radhakrishnan[1], E. Svoren-Jabalera[2], P. Humburg[1], P. Hutton[3], A.C. Gordon[4], C. Garrard[3], A.V.S. Hill[1], C.J. Hinds[5], J.C. Knight[1].* 1) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 2) Queen Mary University of London, Queen's Hospital University Hospital Trust, London, United Kingdom; 3) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford, United Kingdom; 4) Section of Anaesthetics, Pain Medicine and Intensive Care, Imperial College, London, United Kingdom; 5) William Harvey Research Institute, Barts and the London School of Medicine, Queen Mary University, London, United Kingdom.

   Sepsis remains a major area of unmet clinical need, with mortality rates >30%. Although conventionally considered a unified disease with a common pathway to organ failure and death, substantial heterogeneity is seen, which has limited efforts to understand pathophysiology and improve therapeutic strategies. This study aimed to use transcriptomic profiling to investigate variation between and within patients with sepsis due to faecal peritonitis (FP). We present data for 147 samples from 117 adult FP patients recruited to the Genomic Advances in Sepsis study. Detailed phenotypic information was recorded and serial samples taken over five days following admission to intensive care units (ICUs). Genome-wide gene expression profiling of total leukocyte RNA purified using the Ambion LeukoLOCK Total RNA Isolation System was performed using Illumina HumanHT-12v4 Expression BeadChip arrays. An unsupervised clustering approach revealed sepsis response signature (SRS) subgroups within the dataset. The SRS1 group is associated with a higher 14 day mortality (p=0.0096) and shows enrichment of pathways relating to T cell exhaustion, cell death, and endotoxin tolerance. Serial sampling allowed resolution of the SRS groups over time, indicating that SRS1 occurs early and 54% of patients remain in the same SRS group over the study period. We additionally identified temporal changes in gene expression from onset of FP involving leukocyte activation, phagocytosis, and NK cell and IL-3 signalling. Comparisons with pneumonia patients (n=126) and non-septic controls (n=10) demonstrated a shared septic response across aetiologies, with EIF2 signalling the most enriched canonical pathway. We also validated differential gene expression relating to source of infection involving IFN signalling and viral infection, showing that while there is significant overlap in the sepsis transcriptomic response some specificity is seen. Our findings demonstrate that shared and specific processes modulate the individual transcriptomic response to sepsis. The presence of generic sepsis signatures informative for the underlying response state highlights the opportunity for patient stratification and precision medicine in sepsis.

3317T

Role of a long non-coding RNA in prednisolone resistance, cell proliferation and migration in childhood acute lymphoblastic leukemia. *M. Ouimet, M. Lajoie, C. Richer, R. Gioia, S. Drouin, D. Sinnett.* Division of Hematology-Oncology, CHU Sainte-Justine, Montreal, Quebec, Canada.

   Acute lymphoblastic leukemia (ALL) is the most common cancer among children. Survival rates of pediatric ALL have improved enormously over the past decades. However, 20% of B-cell precursor ALL (BCP-ALL) relapse during or shortly after completion of treatment and still constitute one of the main cause of disease-related death in pediatrics. We need to better understand the molecular determinants of treatment outcome. Long non-coding RNAs (lncRNAs) are important regulators of gene expression that play multiple roles in the life cycle of genes, from transcription to mRNA splicing, RNA decay, and translation. Defects in cell differentiation and uncontrolled proliferation are a hallmark of leukemia. Thus lncRNAs involved in gene expression control during hematopoiesis, if deregulated, might contribute to the development and the progression of leukemia. To test this hypothesis, we examined lncRNA expression in pre-B cell childhood ALL (cALL). We performed RNAseq analysis of 57 newly diagnosed cALL and 3 CD19+ control from human cord blood. We assessed whether lncRNAs overexpressed in cALL have an impact on cell proliferation, apoptosis and migration. We observed that silencing of one particular lncRNA (*RP11-137H2.4*) in pre-B ALL cell lines inhibited leukemic survival and abrogated cell migration. Furthermore, we showed that the silencing of *RP11-137H2.4* renders pre-B ALL cells more sensitive to prednisolone, one of the spearhead drug used in multi-drug treatment of pediatric ALL. This is an important finding considering that i*n vivo* and *in vitro* response to prednisolone predicts long-term clinical outcome. Moreover, a disproportionate number of relapsed ALL patients acquire prednisolone resistance as compared to other anti-leukemic agents. These findings suggest a role for the lncRNA *RP11-137H2.4* in the pre-B cell ALL survival, proliferation and migration and might constitute a new target to address prednisone resistance and increase current survival rates. .

3318F

Identification of copy number variations (CNVs) associated with congenital diaphragmatic hernia (CDH). *Q. Zhu[1], E. Cerveira[1], M. Romanovitch[1], J. Kwon[1], F.A. High[2], M. Longoni[2], C. Coletti[2], C. Zhang[1], C. Lee[1], P.K. Donahoe[2].* 1) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA.

   Objectives: CDH is a common and severe birth defect. The incidence of CDH is approximately 1 in 2,500 newborns. More than 80 percent of the CDH patients are not associated with any known genetic syndromes and the underlying molecular mechanisms remain largely unknown. Copy number variations (CNVs) have been suggested to play important roles in the evolution and the pathogenesis of a variety of human diseases. The objective of this study is to identify CNVs associated with CDH. **Methods:** Array comparative genomic hybridization (aCGH) was performed for detection of CNVs in this study. We designed a high-resolution custom microarray (Agilent 8x60K) consisting of 181 target regions representing CNVs or genomic regions generated from extensive transcriptome analyses of normal gene expression in mouse embryonic diaphragms, bio-informatics analysis of protein-protein interactions through key signaling pathways, genes within chromosome deletions detected by CNV analyses in our patient cohort, genes known to cause diaphragm defects in knockout mouse models and human patients, and targets identified from linkage analyses. We performed aCGH on 170 CDH patient samples including 36 trio samples and 590 normal control samples. A panel of CNV candidates were validated by droplet digital PCR (ddPCR). Fisher's exact test was used for statistical analysis.   **Results:** In this study, we identified 67 copy number gains and 207 losses from the 170 CDH patient samples, as well as 192 gains and 434 losses from the 590 normal control samples. Cumulatively, the number of observed CNV losses is highly significant (P=3.8x10^{-5}, OR 1.65, 95% CI 1.29-2.11, Fisher's exact test). Among these CNV calls, 22 gains and 28 losses were observed in the CDH patients only. We also identified several de novo CNV calls from the 36 trio samples. We selected a panel of CNV candidates (8 gains and losses) for validation. Importantly, the panel included one novel gain specifically detected in three of the CDH patient samples (P=0.01by Fisher's exact test) and validated by ddPCR.   **Conclusion:** We have successfully developed and utilized a high-resolution custom array for identification of CNVs from the CDH patient samples. In addition, we have developed a highly sensitive and quantitative ddPCR assay for validation of the CNV candidates. This study has identified novel CNVs associated with CDH and has the potential for further discovery.

3319W

***De novo* assembly and RNA-seq shows season-dependent expression and editing in black bear kidneys.** *A. Srivastava[1], V.K. Sarsani[1], S.M. Sheehan[1], R.L. Seger[2], M.E. Barter[1], C. Lindqvist[3], L.C. Brody[4], J.C. Mullikin[5], R. Korstanje[1].* 1) The Jackson Laboratory, Bar Harbor, Maine; 2) Department of Animal and Veterinary Sciences, University of Maine, Orono, Maine; 3) Department of Biological Sciences, University at Buffalo, Buffalo, New York; 4) Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland; 5) Comparative Genomics Analysis Unit, National Human Genome Research Institute, NIH, Bethesda, Maryland.

  Kidney disease affects more than 26 million Americans. Several mammals have the ability of regenerating the kidney after damage, and although humans likely have the same genes as those species, this does not occur. Therefore, a logical approach is to study the kidney in mammals that have this ability and compare this with human, with the goal of identifying genes involved in the recovery of the damaged kidney. One of those species is the American black bear (*Ursus americanus*). Previous studies in bears suggest unique features in the kidney that allow them to endure lower functioning during and recovery soon after hibernation. These processes are likely in part instilled in the genome sequence and gene expression patterns unique to the bear. To understand the molecular mechanisms underlying these processes, we have performed the whole genome de novo assembly of the black bear with over 100x coverage using Illumina paired-end, Illumina mate-pair, and low-pass PacBio single molecule sequencing. Hybrid assembly of genome yielded 113,759 contigs/scaffolds with N50 size of 190KB and went under comprehensive annotation. Further, we have sequenced RNA from Maine Black Bear and perform the comparison of renal gene expression between spring and fall samples. Our work identified the genes that showed significant expression difference that might play important roles in changes in physiology during hibernation and recovery. Moreover, from the RNA-seq data, we also found evidence of season-dependent differences in RNA-editing.

3320T

Long interspersed nuclear element-1 (LINE1)-mediated 8q13 microdeletion detected in a Japanese case with mesomelia-synostoses syndrome. *I. Imoto, T. Naruto, T. Kohmoto, M. Watanabe, S. Ujiro, K. Masuda.* Tokushima University, Tokushima, Japan.

Mesomelia-synostoses syndrome (MSS, MIM #600383) is a rare autosomal-dominant a syndromal osteochondrodysplasia characterized by mesomelic limb shortening, acral synostoses, and multiple congenital malformations due to a microdeletion at 8q13. To date five unrelated patients have been reported worldwide, and the sizes of reported deletions vary from 582 kb to 738 kb but invariably emcompass two contiguous genes: *SULF1* and *SLCO5A1*. At least two cases, it was reported that deletions do not result from nonallelic homologous recombination (NAHR). Here, we report the first case with MSS caused by a long interspersed nuclear element-1 (LINE1, L1)-mediated 591 kb deletion at 8q13 containing those two genes. An 11-year-old undiagnosed Japanese girl was the second child of non-consanguineous Japanese parents. She had multiple congenital anoanomalies, such as blepharophimosis, blepharoptosis, micrognathia, low-set ear, hypoplastic calcaneus, bilateral ectopic ureters, and left hydronephrosis. Throughout childhood, the limited extension of elbow joints, the bowing and shortening of the forearms, and the ulnar deviation of the hands had not significantly progressed. Symmetrical deformity of the legs resulted in genua valga and proximal dislocation of tibiae and fibulae had not significantly worsened, although malformed ankle joint and toes adversely affected the patient's gait. Targeted exome sequence using next-generation sequencing detected heterozygous deletion of *SULF1*, and subsequently performed array-based copy-number analysis detected a novel 591 kb deletion at 8q13 containing *SULF1* and *SLCO5A1*. Clinical features of this patient were mostly matched with those of MSS, which was considered to be caused by 8q13 microdeletion. Based on the results of array-based analysis, we performed long range-PCR to identify the deletion junction, and sequencing of junctional fragments revealed fusion of the two retrotransposons, L1PA2 and L1PA5, suggesting that the deletion is caused by recombination between L1 elements. Our case highlights the importance of NAHR using L1 repetitive sequences as one of mechanisms causing the 8q13 microdeletion, enclosing *SULF1* and *SLCO5A1*.

3321F

Single nucleotide polymorphism -829C→T of *DHFR* gene in Mexican patients with rheumatoid arthritis. *M.R. Hernández-Preciado[1,3], M.C. Morán-Moguel.[2], I.P. Dávalos-Rodríguez.[2,3], M.A. Aceves-Aceves.[2], I.M. Salazar-Dávalos.[2], M. Romero-Morán.[2], J.P. Valdovinos-Maravilla.[4], C.M. Enríquez.[4], L. González-López.[1,4], J.I. Gámez-Nava.[1,4], M. Salazar-Páramo.[1,4].* 1) Doctorado en Farmacología, Universidad de Guadalajara , Guadalajara, Jalisco, Mexico; 2) Doctorado en Genética Humana, Instituto de Genética Humana, Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara; Guadalajara, Jalisco, México; 3) Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México; 4) Unidad Médica de Alta Especialidad, Hospital de Especialidades, Centro Médico Nacional de Occidente. Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México.

Background: Rheumatoid arthritis (RA) is a chronic systemic inflammatory disease with a prevalence of 1% worldwide. Therapy of RA is based on the use of disease-modifying antirheumatic drugs (DMARDs). Methotrexate (MTX) is recommended as the first DMARD for most patients with RA. MTX is a potent inhibitor of the dihydrofolate reductase enzyme (DHFR). Changes in the levels of DHFR and consequently in the sensitivity to MTX could be due to single nucleotide polymorphisms (SNPs), the SNP -829C→T located at 223 nucleotide downstream from the stop codon between the first and second polyadenylation sites in the 3'UTR of the *DHFR* gene, has been involved in the expression of the enzyme. **Aim:** The objective of this study was to determine the association between the presence of SNP -829CT with response to treatment with MTX among patients with RA. **Methods:** Genomic DNA was obtained from 34 Mexican patients with RA (ACR 1987). From these patients, 16 had a good response to treatment with MTX (responders: evaluated by DAS 28 ≤ 3.2); 18 were considered as a treatment failure with MTX (non responders: DAS 28> 3.2). *DHFR* gene was amplified by polymerase chain reaction (PCR). PCR products were subjected to electrophoresis and analyzed on polyacrylamide gels stained with silver nitrate. The amplified product (269 bp) was digested with the *TspRI* enzyme. The C allele showed three fragments (203, 45 and 21 bp), the T allele has two fragments (248 and 21 bp). The analysis statistic was for direct count of genotypes and alleles, X^2 was applied for comparing groups. Statistical significance was considered p<0.05. **Results:** The genotype frequencies (GF) of SNP -829C→T were for responders patients: 9 CC (0.5), 7 CT (0.39), and 2 TT (0.11), while for non responder patients were 3 CC (0.19), 11 CT (0.68) and 2 TT (0.13). Allelic frequencies (AF) were for responder patients: 17 C (0.53) and 15 T (0.47), while for non responder patients were: 25 C (0.69) and 11 T (0.31). No significant difference were observed when comparing GF and AF in responders vs non responder patients (p>0.05). **Conclusion:** In this preliminary approach we found no significant association between the SNP -829CT and the response to treatment with MTX.

3322W

Correction of *IKBKAP* exon 20 splicing by splice switching oligonucleotides. *G.H. Bruun[1], T.K. Doktor[1], A.R. Krainer[2], T. Bruun[1], B.S. Andresen[1].* 1) Biochemistry and Molecular Biology (BMB), University of Southern Denmark, Odense, Odense M, Denmark; 2) Cold Spring Harbor Laboratory, NY, USA.

Correct splicing of exons with weak splice sites depends on a tight balance between positive and negative splicing regulatory elements (SREs). A large fraction of disease-causing mutations disrupt mRNA splicing by decreasing splice site strength or disrupting/creating SREs. We used iCLIP (individual-nucleotide resolution crosslink and immunoprecipitation) to create an in vivo binding map for the splicing regulatory protein hnRNP A1, which usually binds negative SREs to inhibit exon inclusion. HeLa cells with inducible expression of T7-tagged hnRNP A1 were UV irradiated generating irreversible crosslinks between RNA and RNA binding proteins allowing stringent purification of the bound RNA. The iCLIP libraries were subject to next-generation sequencing. We correlated the hnRNP A1 binding map with the hnRNP A1-regulated exons identified by hnRNP A1 knockdown and RNA-seq, and observed that the region immediately downstream of 5' splice sites is important for hnRNP A1-mediated exon repression. We hypothesized that exons with weak splice sites, which are skipped due to a tipped balance between positive and negative SREs, could be reactivated by blocking iCLIP identified hnRNP A1 binding negative SREs using splice switching oligonucleotides (SSOs). We demonstrated the validity of this approach to activate a *MTRR* pseudoexon and the alternative exon 3 in *SKA2* by SSO-mediated blocking of iCLIP-identified hnRNPA1 binding sites located immediately downstream of the 5' splice sites. Interestingly, our iCLIP map revealed a hnRNP A1 binding site in intron 20 close to the 5' splice site of *IKBKAP* exon 20. Skipping of *IKBKAP* exon 20 causes familial dysautonomia (FD). FD is a recessive disease, affecting the sensory and autonomic nervous system and is frequent in Ashkenazi Jews. Almost all patients with FD are homozygous for a IVS20+6T>C mutation in *IKBKAP*, which decreases 5' splice site strength, causing exon 20 skipping. The iCLIP identified hnRNP A1 binding site was confirmed by in vitro studies. Thus, we designed an SSO blocking the hnRNP A1 binding site in intron 20. Transfection of this SSO into FD patient fibroblasts fully restores *IKBKAP* exon 20 splicing. This points to a future SSO-based therapeutic possibility in FD and illustrates that the iCLIP generated hnRNP A1 binding map can be used to identify potential targets for SSO-based therapy.

3323T

Alternative splicing in *FMR1* premutation carriers. *F. Tassone[1,2], R. Olaby[1], H. Tang[1], L. Hickey[3], E. Tseng[3].* 1) Biochemistry and Molecular Medicine, UC Davis, Sacramento, CA, 95817, USA; 2) MIND Institute, UC Davis, Sacramento, CA, 95817, USA; 3) Pacific Biosciences, Inc., Menlo Park, CA, 94025, USA.

Over 40% of males and ~16% of female carriers of a *FMR1* premutation allele (55-200 CGG repeats) are at risk for developing Fragile X-associated Tremor/Ataxia Syndrome (FXTAS), an adult onset neurodegenerative disorder while, about 20% of female carriers will develop Fragile X-associated Primary Ovarian Insufficiency (FXPOI), in addition to a number of adult-onset clinical problems (*FMR1* associated disorders). Marked elevation in *FMR1* mRNA levels have been observed with premutation alleles and the resulting RNA toxicity is believed to be the leading molecular mechanism proposed for these disorders. The *FMR1* gene, as many housekeeping genes, undergoes alternative splicing. Using long-read isoform sequencing (SMRT) and qRT-PCR we have recently reported that, although the relative abundance of all *FMR1* mRNA isoforms is significantly increased in the premutation group compared to controls, there is a disproportionate increase, relative to the overall increase in mRNA, in the abundance of isoforms spliced at both exons 12 and 14. In total, we confirmed the existence of 16 out of 24 predicted isoforms in our samples. However, it is unknown, which isoforms, when overexpressed, may contribute to the premutation pathology. To address this question we have further defined the transcriptional *FMR1* isoforms distribution pattern in different tissues, including heart, muscle, brain and testis derived from FXTAS premutation carriers and age-matched controls. Preliminary data indicates the presence of a transcriptional signature of the *FMR1* gene, which clusters more by individual than by tissue type. We identified additional isoforms than the 16 reported in our previous study, including a group with particular splice patterns that were observed only in premutations but not in controls. Our findings suggest that the characterization of expression levels of the different *FMR1* isoforms is fundamental for understanding the regulation of the *FMR1* gene as well as for elucidating the mechanism(s) by which "toxic gain of function" of the *FMR1* mRNA may play a role in FXTAS and/or in the other *FMR1*-associated conditions. In addition to the elevated levels of *FMR1* isoforms, the altered abundance/ratio of the corresponding FMRP isomers may affect the overall function of FMRP in premutations. .

3324F

Single Molecule, Real-Time (SMRT) Sequencing of expanded spinocerebellar ataxia type 10 (SCA10) pentanucleotide repeat alleles directly from genomic DNA. *T. Ashizawa[1,2,3], Y.C. Tsai[6], B. Schüle[7], T.A. Clark[6], W.G. Farmerie[4,5], K.N. McFarland[2,3].* 1) Department of Neurology, Houston Methodist Research Institute, Houston, TX; 2) Department of Neurology, College of Medicine, University of Florida, Gainesville, FL; 3) McKnight Brain Institute, University of Florida, Gainesville, FL; 4) Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL; 5) Genetics Institute, University of Florida, Gainesville, FL; 6) Pacific Biosciences, Menlo Park, CA; 7) Parkinson's Institute and Clinical Center, Sunnyvale, CA.

   Background: SCA10 is an autosomal dominant cerebellar ataxia caused by a large (up to 21kb) expansion of an intronic ATTCT repeat in *ATXN10*. Using SMRT Sequencing technology, we have reported contiguous sequences of PCR-amplified expanded SCA10 alleles. However, many SCA10 expansion alleles are too large for PCR amplification, and *in vitro* amplification may introduce sequence artifacts. **Purpose:** To obtain contiguous sequences of SCA10 alleles directly from the genomic DNA by the SMRT Sequencing technology, without PCR- or cloning amplification. **Methods:** Native SMRTbell libraries were enriched using a novel CRISPR-Cas9 approach. DNA molecules were cleaved adjacent to the SCA10 repeat, followed by ligation of a capture adapter and subsequent hybridization to magbeads. The enriched DNA was subjected to SMRT Sequencing. **Results:** The enriched DNA from representative SCA10 patients showed three types of the expansion motifs: type A alleles are mostly pure $(ATTCT)_{n1}$ repeat, type B alleles consist of $(ATTCT)_{n2}$-$(ATTCC)_{n3}$, and type C alleles have $(ATTCT)_{n4}$-$(ATCCT)_{n5}$-$(ATCCC)_{n6}$, where each of n1-n6 varies between circular consensus sequencing (CCS) reads of single individuals and between different family members, while the allele type does not change within each SCA10 family. The 5' repeat structure of these samples are consistent with the corresponding repeat-primed PCR data. The presence of $(ATCCT)_{n5}$ has been highly correlated with epilepsy phenotype of SCA10. All expansion alleles share a rare haplotype (<3% in the general population) and are found exclusively in populations with the Native American or Han Chinese background. **Discussions:** Since each CCS read represents the SCA10 repeat sequence from a single cell, the SMRT Sequencing can serve as a powerful assay for studies of repeat instability. The sequence variations in the entire span of expanded alleles allow us to critically examine genotype-phenotype correlations and repeat-dependent pathogenicity. We will discuss applications in future studies, including (1) elucidating the mechanism of SCA10 repeat instability using cells differentiated from patient-derived iPS cells, (2) investigating the role of type A alleles in development of the epilepsy phenotype, and (3) studying evolution of SCA10 repeat expansion alleles during the human migration. **Conclusions:** The SMRT technology has a potential to sequence large repeat regions directly from the genomic DNA.

3325W

Characterization of NAHR-mediated type-1 *NF1* deletions with breakpoints not located within the known NAHR hotspots PRS1 and PRS2. *H. Kehrer-Sawatzki[1], M. Hillmer[1], V.-F. Mautner[2], L. Messiaen[3], D.N. Cooper[4], A. Summerer[1].* 1) University of Ulm, Institute of Human Genetics, Ulm, Germany; 2) University Hospital Hamburg Eppendorf, Department of Neurology, Hamburg, Germany; 3) University of Alabama at Birmingham, Medical Genomics Laboratory, Department of Genetics, Birmingham, AL, USA; 4) Cardiff University, Institute of Medical Genetics, School of Medicine, Cardiff, UK.

   Neurofibromatosis type 1 (NF1) is a hereditary cancer syndrome with an incidence of 1 in 3000. In 5 % of all NF1 patients, large deletions encompassing the *NF1* gene and its flanking regions at 17q11.2 are the cause of the disease. The majority (80%) of all large *NF1* deletions are of type-1, encompass 1.4-Mb and are caused by nonallelic homologous recombination (NAHR). The breakpoints of type-1 deletions are located in the low-copy repeats NF1-REPa and NF1-REPc which exhibit high sequence homology to one another. Previous studies indicated that type-1 deletion breakpoints cluster within the paralogous recombination sites PRS1 and PRS2. These NAHR hotspots harbour the strand exchange regions (SERs) of all the type-1 *NF1* deletions reported so far. The respective SERs are of different length, spanning from 20-bp to 700-bp. The extent of the SERs are determined according to the distance between the paralogous sequence variants (PSVs) that flank the SERs. Within the SERs, NF1-REPa and NF1-REPc exhibit sequence identity. The NAHR-associated crossover sites are located within the SERs. In our present study, we investigated 250 patients with type-1 *NF1* deletions by means of multiplex ligation-dependent probe amplification (MLPA) and long-range PCRs to detect breakpoints located within PRS1 and PRS2. According to these analyses, 10% of the type-1 deletions do not exhibit breakpoints located with the NAHR hotspots PRS1 and PRS2. To identify the breakpoint location of these 25 type-1 *NF1* deletions not mediated by NAHR within PRS1 or PRS2, we performed high density array CGH. Subsequently breakpoint-spanning PCRs were designed in order to identify the respective SERs. These analyses suggested that the majority of the deletion breakpoints are located within the 14-kb region that separates PRS1 from PRS2. In the next step of our analysis we investigated the breakpoint-flanking regions of all type-1 *NF1* deletions with regard to the presence of sequence motifs that have the capability to adopt non-B DNA sequence conformations. Our findings imply that non-B DNA forming sequence motifs are overrepresented in the breakpoint-flanking regions of all type-1 deletions, those mediated by NAHR within the known hotspots PRS1 and PRS2 as well as those with breakpoints located outside these hotspots. We hypothesize that non-B DNA sequence motifs contribute to the occurrence of DNA double strand breaks that initiate the NAHR events underlying type-1 *NF1* deletions.

3326T

Evaluation of the contribution of complex structural variation at the 17q21.31 locus to the risk of progressive supranuclear palsy. *A. Huang[1], J. Chen[1], A. Karydas[2], B. Miller[2], A. Boxer[2], G. Coppola[1].* 1) Departments of Psychiatry and Neurology, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA; 2) Department of Neurology, University of California San Francisco, San Francisco, CA.

Progressive supranuclear palsy (PSP) is a devastating neurodegenerative disorder that affects movement, gait, breathing, balance, and cognitive function. It is a rare disorder that affects approximately three in 100,000 people. Genome-wide association studies have confirmed that, similar to other neurodegenerative disorders, genetic variation at a common inversion polymorphism located on 17q21.31 are among the greatest known contributors to disease risk. This 900 kbp inversion is divided into two main haplotypes groups, deemed H1 and H2. The absence of recombination is a hallmark consequence of the inversion polymorphism that distinguishes these two haplotype clades, the latter of which is exclusive to European populations. This stratification has generated some distinctive consequences with regard to disease risk. While the H2 structural haplotype predisposes European populations to 17q21.31 microdeletion syndrome, it has also shown to be highly protective against PSP. With several types of genetic data from over 380 individuals affected with PSP and an additional 180 control individuals, we carried out a wide variety of analysis in order to better characterize the complex structural variation within 17q21.31 and evaluate its contribution to PSP. First, we used dense, single nucleotide polymorphism (SNP) microarrays (Illumina Omni 2.5) to genetically determine sample ethnicity and divide our sample by H1/H2 carrier status. Raw Intensity data from these SNP arrays was then used to roughly delineate highly copy-number variable regions (CNVRs) within the 17q21.31 locus. The precise breakpoints of these CNVRs were further refined with discordant read-pair analysis of deep (30X) whole-genome sequencing (WGS) data in all individuals. Finally, read-depth analysis was used to type these multi-allelic CNVRs. Although within this limited dataset, we find no significant association overall copy number to PSP among any of the CNVRs analyzed, we describe the utilization of our WGS data as a reference for the construction of an imputation panel for common structural diversity in the region. This deeply-sequenced dataset thus serves as an invaluable resource for further examination of this structurally complex region. Evaluation in an expanded sample, utilizing the imputation of CNVRs, rather than cost-prohibitive WGS, to dissect the genetic association of structural variation within 17q21.31 to disease risk, are currently underway.

3327F

Modeling ring chromosomes: Ring structure affects gene expression both from the ring and throughout the genome. *L. Conlin, R. Rajagopalan, D. McEldrew, J. Mills, N. Spinner.* Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA.

Ring chromosomes 14 (r(14)) and 20 (r(20)) are associated with well-characterized clinical syndromes that include seizures. To understand the effect of the ring chromosomes on gene expression, we carried out RNA-sequencing. In patients with r(20) mosaicism, we previously showed that the rings have no deletions or duplications, allowing comparison of normal and ring containing cells with no additional abnormalities. For each r(20) patient cell line, normal (46,XX) and r(20) (46,XX,r(20)) fibroblast cell lines were generated. For r(14), mosaicism is atypical, so RNA-sequencing was performed on 7 r(14) patients and 9 control cell lines. Four hundred and thirty five genes were differentially expressed (DEG) between ring chromosome 20 and isogenic normal samples. Of these DEGs, 335 were upregulated and 100 were down-regulated in the r(20) samples. Sixteen of these genes were located on chromosome 20. Gene ontology enrichment analysis of DEGs indicated altered transcription for a number of cellular phenotypes including cell movement, morphology, structure and assembly, growth, proliferation and cell cycle processes, which were consistent with functional studies. Allele-specific expression analysis of the r(20) samples showed decreased expression from the ring. Our data shows that ring chromosomes affect cellular transcription and function, even in the absence of associated copy number alterations. Additionally, we set out to create iPS cells for r(14) and r(20) to develop an *in vitro* neural developmental model to study the seizures and other associated brain abnormalities. IPS cells were created from fibroblasts and characterized for pluripotency markers. Multiple samples were studied for both r(14) and r(20), and in all cases the cells lost the ring chromosomes with the appearance of 2 normal chromosomes. We carried out fluorescence *in situ* hybridization for polymorphic CNVs and genome-wide SNP arrays to characterize these "normal" cells. In all cases, iPSCs resulted in uniparental isodisomy (UPD) for chromosome 14 or 20, as has been previously reported for ring chromosomes 13 and 17. We are therefore pursuing alternate strategies to create cellular models for ring chromosomes focused on generation of neural cells through direct-conversion of fibroblasts to neural progenitor or post-mitotic neurons. These cells will be used to to study gene expression profiles in disease-specific cell types and compared to patient fibroblast expression data.

3328W

Expression and proteomic analyses of KIF1A/25B in hereditary sensory and autonomic neuropathies type II. *S. Mohtashami[1,2], J.F. Schmouth[1,2], P.A. Dion[1,2], G.A. Rouleau[1,2].* 1) Experimental Medicine, McGill University , Montreal, Quebec, Canada; 2) Montreal Neurological Institute and Hospital, Montreal, Quebec, Canada.

Expression and proteomic analyses of KIF1A/25B in hereditary sensory and autonomic neuropathies type II *S. Mohtashami[1,2], J. F. Schmouth[1,2], P. A. Dion[1,2], G. A. Rouleau[1,2]* 1) McGill University, Montreal, QC, Canada; 2) Montreal Neurological Institute and Hospital, Montreal, QC, Canada.   Hereditary sensory and autonomic neuropathies form a group of genetic disorders characterized by variable sensory and autonomic dysfunctions. HSAN type II (HSANII) is a debilitating subtype manifesting in early childhood with distal numbness and loss of pain, temperature and touch. Our laboratory has reported truncating mutations in a nervous-tissue-specific exon (*HSN2*) of the *WNK1* gene. The *WNK1* isoform containing the alternatively spliced exon (*HSN2*) is referred to as the *WNK1/HSN2* isoform. Interestingly the protein region encoded by the alternatively spliced exon was found to interact with a particular isoform of another HSANII causative gene, *KIF1A*. The HSANII-causing *KIF1A* isoform is referred to as *KIF1A/25B* since disease-causing mutations were exclusively found in the alternative exon "*25B*". The expression profile of KIF1A/25B across the nervous system is determined by performing WB immunodetection using tissues from wild-type mice and a rabbit antiserum. The function of the protein encoded by exon 25B is assessed through a profiling of its interacting partners by performing co-immunoprecipitation for both full-length KIF1A/25B protein and the protein that correspond to exon 25B. Positive interactions are confirmed using liquid chromatography–mass spectrometry. We hypothesize that KIF1A/25B is the transit system through which WNK1/HSN2 traffics within the cells and offers the two proteins an opportunity to interact with other cellular elements relevant to the sensory and nociceptive aspects of HSANII. This study answers fundamental questions regarding the molecular pathophysiology of HSANII, though our findings have an impact on the understanding of other neuropathies and of normal neuronal processing.

3329T

Urea concentration is increased in the brains of Huntington's disease patients and in the striatum of the prodromal OVT73 sheep model. *R.G. Snell[1,2], R.R Handley[1,2], S. Patassini[1,2,4], I. Fourie[1], C.S Bawden[3], S.R Rudiger[3], P. Verma[3], G.J Cooper[4], R.D Unwin[4], H.J Waldvogel[2], J.F Gusella[5], M.E MacDonald[5], R.L.M Faull[2], S.J Reid[1,2].* 1) School of Biological Sciences, The University of Auckland, Auckland, New Zealand; 2) Centre for Brain Research, The University of Auckland, Auckland; 3) Molecular Biology and Reproductive Technology Laboratories, South Australian Research and Development, Adelaide; 4) Centre for Advanced Discovery and Experimental Therapeutics (CADET), Central Manchester University Manchester; 5) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston.

  Huntington's disease (HD) is an autosomal dominant inherited neurodegenerative disorder with onset generally in mid life. The mutation that causes HD is an expanded CAG repeat coding for glutamine in exon 1 of the HTT gene (1). Despite the gene being isolated over 20 years ago there is still no clear understanding of the biochemical processes leading to cell dysfunction and cell loss, processes that are initiated well before the onset of symptoms. We have recently reported the observation of an increase in urea distributed throughout post-mortem HD brains (2). Disruption of the urea cycle through mutation or acute bacterial toxicity can result in neuronal cell death and is therefore a potential pathogenic mechanism in HD. In order to investigate this phenomena further we have extended our human post mortem brain studies and have measured urea concentrations in brain and peripheral tissues from our prodromal transgenic sheep model. Confirming our preliminary observation we have seen an increase (3-4 fold) increased in HD brain samples from a large cohort (30 cases, 18 controls). We have also seen this increase in a subset of pathological grade 0-1 HD cases where there is minimal cell loss. Importantly, we find an increase in urea specifically within striatum from our prodromal HD sheep model in which cell loss is not a feature. Taken together this suggests that the increase in urea is unlikely to be a direct consequence of secondary catabolism. These observations are mechanistically supported by our observation of an up-regulation of the urea transporter transcript (SLC14A1) transcript throughout the human and sheep brains possibly as a physiological process to clear the urea.   1. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's-disease chromosomes. Cell 1993, 72 (6), 971-983. 2. Patassini, S.; Begley, P.; Reid, S. J.; Xu, J.; Church, S. J.; Curtis, M.; Dragunow, M.; Waldvogel, H. J.; Unwin, R. D.; Snell, R. G.; et al. Identification of elevated urea as a severe, ubiquitous metabolic defect in the brain of patients with Huntington's disease. Biochem Biophys Res Com 2016 468(1–2): 161-6.

3330F

Role of the Prader-Willi Syndrome protein MAGEL2 in intracellular pathways regulating the leptin receptor. *T.M. Wijesuriya[1], K.V. Carias[1], K. Fagan[1], L. De Ceuninck[2], J. Tavernir[2], R. Wevrick[1].* 1) Department of Medical Genetics, University of Alberta, 8–16 Medical Sciences Building, Edmonton, AB, Canada T6G 2H7; 2) Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium.

Background: Children with Prader-Willi syndrome have neonatal feeding difficulties, developmental delay and excessive appetite. Loss of MAGEL2 causes a neurodevelopmental disorder (Schaaf-Yang syndrome) and may contribute to obesity in children with Prader-Willi Syndrome who lack MAGEL2 and other genes. MAGEL2 is essential in neurons that sense levels of the adipose tissue-derived hormone leptin. The MAGEL2 protein is important for recycling or degradation of proteins in the brain and interacts with and modifies the activity of E3 ubiquitin ligases. RNF41 is a E3 ubiquitin ligase that associates with a ubiquitin-specific protease (USP8). Together with USP8, RNF41 regulates the recycling of the leptin receptor by targeting it either for degradation or for recycling to the cell membrane. We hypothesized that MAGEL2 normally regulates the interaction between RNF41 and USP8, and that loss of this regulation could impair leptin response pathways in the brain in children with PWS. Methods: Human U2OS cells were transfected with recombinant constructs encoding epitope tagged versions of MAGEL2 and wild type and mutant forms of RNF41. Immunofluorescence was used to visualize the co-localization of different forms of RNF41 with MAGEL2 in intracellular compartments. We expressed recombinant MAGEL2, RNF41 and USP8 in combinations in human U2OS cells and examined the relative abundance of each protein in the presence or absence of the other components of the complex. Results: We identified interactions among components of the RNF41-USP8 complex that depended either on the activity of the RING domain of RNF41 (RNF-SQ mutant form) or the binding domain of RNF41 (RNF-AE mutant form). Co-expression of MAGEL2 with components of the RNF41-USP8 complex modified the abundance of proteins in the complex. Preliminary results suggest that MAGEL2 modifies the ability of RNF41 to auto-ubiquitinate or to ubiquitinate MAGEL2. Co-expression of MAGEL2 also modified the intracellular localization of components of the RNF41-USP8 complex. We also measured levels of endogenous RNF41 and USP8 level in brain tissues from Magel2 knockout mice and compared these levels to those found in tissues from wild type littermates. Conclusion: Our results suggest that MAGEL2 could modify the activity of the RNF41-USP8 ubiquitination complex in leptin sensing neurons, providing a possible mechanism for dysregulation of leptin sensing in neurons in children with Prader-Willi syndrome.

3331W

Integrated gene co-expression network analysis reveals genotype-phenotype correlations in Williams syndrome. *R. Kimura[1], K. Tomiwa[2], T. Awaya[3], T. Kato[3], M. Nakata[1], Y. Funabiki[4], T. Heike[3], M. Hagiwara[1].* 1) Dept Anatomy and Developmental Biology, Kyoto University Graduate School of Medicine, Sakyo-ku Kyoto, Japan; 2) Todaiji Medical and Educational Center, Nara, Japan; 3) Dept Pediatrics, Kyoto University Graduate School of Medicine, Sakyo-ku Kyoto, Japan; 4) Dep Human Coexistence, Kyoto University Graduate School of Human and Environmental Studies, Sakyo-ku Kyoto, Japan.

Williams Syndrome (WS) is caused by a 7q11.23 heterozygous deletion containing 26-28 genes. WS is characterized by various symptoms including hyper sociability. However, it remains unclear a biologically relevant signature about global dysregulations caused by deletion at 7q11.23. In contrast, Autism Spectrum Disorder (ASD) is a characterized by impairments in social communication and presence of restricted/repetitive patterns of behavior. Therefore, WS and ASD have been described as "opposite" disorders in terms of their social behavior. A better understanding of the differences between these two disorders could provide insights into genotype-phenotype relationships. In this study, we performed integrated gene co-expression network analysis to identify genes and pathways related to neuropsychiatric phenotypes. A total of 154 subjects, including 94 discovery groups (32 WS, 32 ASD and 30 Controls) and 60 independent replication groups (26 WS, 17 ASD and 17 Controls) participated in this study. To evaluate behavioral problems, we used the Child Behavior Checklist (CBCL) or the Adult Behavior Checklist (ABCL) and the Social Responsiveness Scale-2 (SRS-2). In addition to standard differential expression analysis using peripheral blood, we applied weighted gene co-expression network analysis (WGCNA) to identify genes and pathways related to specific phenotypes in WS. We identified several differentially expressed genes in WS compared to both ASD and Controls. Using WGCNA, we identified 7 modules associated with WS and 1 module associated with ASD. Interestingly, 3 modules, which were up regulated in WS, were significantly associated with neuropsychiatric phenotypes. Our findings strongly support the hypothesis the dysregulated transcriptomic network at the genome-wide level is associated with the variety of neuropsychiatric phenotypes in WS.

3332T

Kctd13 deficiency in mice causes recognition memory and sociability deficits. *T. Arbogast[1], E. Oh[1], R. Rodriguiz[2], W. Wetsel[2], N. Katsanis[1], C. Golzio[1].* 1) 1Center for Human Disease Modeling, Duke University, 300N Duke Street, Durham NC 27701; 2) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC 27710.

The 16p11.2 BP4-BP5 deletion and duplication syndromes are associated with developmental delay and autism spectrum disorders, with a reciprocal effect on head circumference (i.e. micro/macrocephaly) and body mass index (i.e. obesity/underweight). We and others have generated mouse models of the 16p11.2 CNV (copy number variant) and these models recapitulated some of the phenotypes observed in humans including activity and short-term memory alterations (Arbogast, et al. 2016). By performing functional testing of the 16p11.2 CNV in zebrafish, we also found that overexpression of the human *KCTD13* transcript in embryos induces microcephaly, whereas knockdown of endogenous *kctd13* leads to macrocephaly, recapitulating the mirrored phenotype seen in 16p11.2 CNV carriers (Golzio, et al. 2012). To evaluate further the candidacy of *Kctd13*, we generated a ubiquitous knockout of *Kctd13* in the mouse and asked whether loss of *Kctd13* leads to behavioral deficits. First, we observed normal weight and activity in *Kctd13* mutant mice in both homozygotes (Hom) and heterozygotes (Het) indicating that *Kctd13* is not the driver of the BMI defects associated with 16p11.2 CNV. Second, we subjected the *Kctd13* Hom and Het mice and wildtype littermates to a suite of behavioral tests. We observed a significant deficit in novel object recognition and social interaction for the Het and Hom mice. In parallel, MRI analysis is currently performed to determine whether loss of *Kctd13* also has an impact on brain anatomy including the cortex and hippocampus in the mouse. Taken together, our functional studies suggest that dosage imbalance of *Kctd13* contributes to the intellectual disability and social deficits observed in individuals with the 16p11.2 CNV syndromes. Arbogast, T., A. M. Ouagazzal, C. Chevalier, M. Kopanitsa, N. Afinowi, E. Migliavacca, B. S. Cowling, M. C. Birling, M. F. Champy, A. Reymond and Y. Herault (2016). "Reciprocal Effects on Neurocognitive and Metabolic Phenotypes in Mouse Models of 16p11.2 Deletion and Duplication Syndromes." PLoS Genet **12**(2): e1005709. Golzio, C., J. Willer, M. E. Talkowski, E. C. Oh, Y. Taniguchi, S. Jacquemont, A. Reymond, M. Sun, A. Sawa, J. F. Gusella, A. Kamiya, J. S. Beckmann and N. Katsanis (2012). "KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant." Nature **485**(7398): 363-367.   .

3333F

Deep sequencing reveal variations in somatic cell mosaic mutations between monozygotic twins with discordant psychiatric disease. *Y. Morimoto[1], S. Ono[1], A. Imamura[1], Y. Okazaki[2], A. Kinoshita[3], H. Mishima[3], H. Nakane[4], H. Ozawa[1], K. Yoshiura[3], N. Kurotaki[1].* 1) Department of Neuropsychiatry, Unit of Translation Medicine Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 2) Michino-o Hospital, Nagasaki, Japan; 3) Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 4) Department of Psychiatric Rehabilitation Science, Unit of Rehabilitation Science, University Graduate School of Biomedical Sciences, Nagasaki, Japan.

Neurodevelopmental disorders including schizophrenia, autistic spectrum disorder (ASD), and gender dysphoria (GD) are thought to be common, multifactorial diseases. Genetic factors as well as environmental factors including life events are involved in their pathogenesis. A number of family and twin studies have shown that genetic factors play an important role in the onset of various neurodevelopmental disorders. Phenotypically discordant MZ twins are interesting resources for genetic studies in diseases with a high concordance rate, and can make twin studies helpful in identifying the causative genes for heritable diseases. Identifying the molecular genetic differences between discordant MZ twins can help to elucidate the molecular mechanisms that underlie phenotypic discordance. To date, only a few studies have successfully detected genetic differences such as repeat length, single nucleotide variants, and copy number variants between discordant MZ twins. A number of genetic studies for neurodevelopmental disorders have been reported; however, because of genetic heterogeneity, their pathogenesis is still unclear. To identify causative genes for neurodevelopmental disorders, we performed whole exome sequencing (WES) and deep sequencing in 5 pairs of MZ twins with discordant neurodevelopmental disorders and 1 healthy control MZ twin. As a result, we identified 3 discordant variants confirmed by deep sequencing after analysis by personalized next generation sequencing in GD discordant twin. Three mutations in FBXO38, SMOC2 and TDRP, were detected with low allele frequency of mutant alleles on deep sequencing, suggesting that these loci are mosaic due to somatic mutations in a developmental stage. In silico analysis showed that the effects of all 3 mutations were "damaging". Specifically, testis development-related protein (*TDRP*) gene was found to be important in spermatogenesis and male infertility and may be related to maleness. In this context, *TDRP* could be a GD-related gene. In conclusion, we analyzed discordant variants in MZ twin pairs and 1 healthy control MZ twin using WES, and detected 3 somatic cell mosaic mutations in the GD twins by deep sequencing.It might be important to consider the presence of somatic cell mosaic mutations when confirming data obtained from WES. .

3334W

Human endogenous retroviruses and schizophrenia. *F. Macciardi[1], G. Guffanti[2], J. Fallon[1].* 1) Dept of Psychiatry & Human Behavior, Univ California, Irvine, Irvine, CA; 2) McLean Hospital, Department of Psychiatry, Harvard Medical School.

The Human Endogenous Retroviruses (HERVs) are the most represented elements of the Long Terminal Repeat class (LTR) of Transposable Elements (TEs), and make up 8% of the nuclear DNA. Intact HERVs maintain the structure of a retrovirus, but the vast majority are truncated or just Solo-LTRs. Yet, HERVs can be functionally expressed, acting primarily as enhancers or promoters of neighboring genes, in various tissues, as in the brain, with a putative functional role in diseases, like schizophrenia. The majority of HERVs studies estimated the combined expression of all insertions of an element (eg all copies of HERVs belonging to -K family), rather than pinpointing the exact genomic instance of a specific HERV (eg HERVK-1), limiting our understanding of their functional role. ENCODE RNA-seq studies revealed a widespread pattern of expression of different HERVs families in twelve different cell lines, and through application of an *ad hoc* mapping pipeline, demonstrated the feasibility of profiling single HERVs at their specific chromosomal locations. These results suggest that expression of specific HERVs insertions can be uncovered and their regulatory effect on neighboring genes assessed. We combined existing and newly developed analytical strategies for transcriptome assembly and annotation into a pipeline for HERV detection from RNA-seq data, and applied it to > 20 billion RNA-seq reads from 20 individual DLPFC expression profiles. We used TETranscript to evaluate the amount of expression for the hg38 reported 516 endogenous retroviruses (HERVs; #=481) and Long Terminal Repeats (LTR; #=35) families, obtaining expression signals for 97.7% of them. Then, we further analyzed our data with Trinity to map the individual elements of each HERV and LTR locus. We unequivocally mapped 28,432 expressed HERVs / LTRs, considering those RNAseq transcripts that overlap with at least 80% with an annotated HERV/LTR, confirming the sequencing by blastn. For example, we easily mapped the HERVKC4-int element within intron 9 of the C4A locus with high specificity and sensitivity, i.e., the HERVKC4 transcript is independent from exons and isoforms of C4A itself. Our preliminary results show that (1) HERVs and LTRs are expressed at detectable levels in postmortem brains, (2) we can unequivocally map individual elements and (3) reliably estimate their differential expression between cases and controls.

3335T

Mosaic small supernumerary marker chromosome potentially generated through a rolling circle mechanism. *S. Gu[1], C.M.B. Carvalho[1], B. Yuan[1], J. TCW[2,3], S. McCarthy[4], D. Malhotra[5], J. Sebat[6], U. Rudolph[7,8], K.J. Brennand[9], D.L. Levy[8,10], J.R. Lupski[1,11,12,13].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Departments of Neuroscience, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029; 3) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029; 4) Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY 11724; 5) F. Hoffmann-La Roche Ltd, CH-4070 Basel, Switzerland; 6) University of California San Diego, Department of Psychiatry and Department of Cellular and Molecular Medicine, University Of California, San Diego, La Jolla, CA 92093; 7) Laboratory of Genetic Neuropharmacology, McLean Hospital, Belmont, MA 02478; 8) Department of Psychiatry, Harvard Medical School, Boston, MA 02215; 9) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029; 10) Psychology Research Laboratory, McLean Hospital, Belmont, MA 02478; 11) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; 12) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 13) Texas Children's Hospital, Houston, TX 77030.

With a detection frequency of 0.044% in newborns, small supernumerary marker chromosomes (sSMC) derived from every chromosome have been observed. However, little is known about the mechanisms underlying their formation. In a proband diagnosed with schizo-affective disorder, we identified a complex copy-number variant (CNV) at 9p24.1p23 by array comparative genomic hybridization (aCGH). Custom high-density aCGH further revealed the pattern of the CNV to be triplication-duplication-triplication-normal-duplication (TRP-DUP-TRP-NML-DUP). Chromosome analysis and fluorescence *in situ* hybridization (FISH) in leucocyte cells revealed that the CNV was due to the presence of a sSMC stably inherited from his mother, who has a diagnosis of bipolar disorder with psychotic features. SNP array results indicated that this *de novo* complex genomic rearrangement (CGR) in the mother was likely generated through a mitotic interchromosomal event in the proband's maternal grandmother. The proband and his mother showed different levels of mosaicism for the CNV in all tissues studied (leucocytes, skin fibroblasts, lymphoblastoid cell lines, and induced pluripotent stem cells). High-density aCGH and quantification of the CNV by droplet digital PCR (ddPCR) consistently demonstrated fewer cells carrying the CNV in the mother than her son in both fibroblasts (75% versus 89%) and lymphoblastoid cell lines (0% versus 90%), potentially influencing their distinct clinical phenotypes. By long-range PCR and Sanger sequencing of the PCR products, we fine-mapped two breakpoint junctions (Jct1 and Jct2) of this complex CNV to nucleotide resolution, revealing a templated insertion of 399bp in Jct1 copied from a more proximal region at 9p23. Surprisingly, quantification of Jct1 and Jct2 by ddPCR indicated the presence of two copies of Jct1 versus one copy of Jct2, consistent with a rolling circle mechanism underlying formation of the triplication event. In summary, we propose that replication-based mechanisms have a role in the formation of sSMC with the CGR pattern potentially generated through a rolling circle. In addition, distinct levels of mosaicism in individuals carrying the same sSMC may have pleiotropic effects on psychiatric phenotypes.

3336F

Multi-level integrated exome analyses converge to a coherent system of molecular mechanisms on autism. *W. Luo[1], C. Zhang[2], C. Brouwer[1].* 1) Department of Bioinformatics and Genomics, UNC Charlotte, Charlotte, NC; 2) Department of Systems Biology, Columbia University, New York, New York.

Autism spectrum disorder (ASD) is a range of complex genetic diseases. In recent years, multiple whole exome/genome sequencing studies identified thousands of rare mutations and firmly established their roles in ASD. Because these variants rarely recur, major challenges remain as to: 1) evaluate the disease association of individual variants; 2) replicate independent studies or verify the results systematically. A coherent and systematic understanding of autism biology has not been achieved. To address these challenges, we devised a novel integrated analysis across multiple levels, i.e. variant, gene and pathway levels. This multi-level approach has major advantages over the classical one-level approach. First, it produces more informative, systematic, holistic genetic understanding. Second, multiple-level/angle screenings of the same data reaches more robust conclusions. Third, it provides more sophisticated and relevant classification and prioritization of rare mutations, which makes powerful analyses possible with these rare events. We applied this approach to two large scale whole exome studies and identified hundreds of potential causal mutations. We quantified and identified substantial consistence both within and between studies (Table 1), revealing a sequential convergence from variant, gene to pathway level. We deeply dissected ASD genetic association, and built a novel and more inclusive gene+pathway dual-hit model which could be generalizable to CNV or GWAS data. We reconstructed novel, replicable, systematic and multiscale molecular mechanisms for ASD (Table 1, Fig. 1-2). They provide solid and actionable molecular roadmaps for the development of effective and personalized ASD diagnostics and therapeutics. Note this multi-level integrated analysis is generally useful for other complex diseases and genomic studies. Note the full detail of this work is described in our pre-print at: http://biorxiv.org/content/early/2016/05/11/052878.

3337W

Telomere length variation in preschool children: Its relatedness with exposure to chronic stress, maternal depression, and cortisol reactivity. *A. Thakur, M. Kleiber, H. Sheikh, E. Hayden, S.M. Singh.* Biology, The University of Western Ontario, London, Ontario, Canada.

Major depressive disorders (MDD) often co-occur with stress. A number of recent reports have implied that stress-associated depression may be related to telomere length (TL). Shorter TL has been reported in the never-depressed daughters of women with a history of depression (familial risk for depression). Further, shorter telomeres have been reported to be associated with greater cortisol reactivity to stress. Such observations have been used to argue that TL may represent a biomarker for the risk for developing depression (Gotlib et al 2015). Such results and conclusions are not without exception and the relationship between TL, stress and depression, although intriguing, remains to be adequately elucidated. This research represents a follow-up on this controversy with an improved experimental design. It uses a large sample size (N=384) of mostly Caucasian young children (3-4 years) from South-western Ontario. Lifetime history of depression of the mothers was assessed using Structured Clinical Interviews for the DSM-IV and the UCLA Life Stress interview was used to assess chronic stress. All children were subjected to a defined play stress followed by assessment of their salivary cortisol and telomere length by monochrome multiplex qPCR following Cawthon (2009). Relative TL was measured as a ratio of telomere product (T) relative to the single copy albumin (S) gene product for each sample. The mean TL in our cohort is 2.42 (SD=1.98). This variability in TL has allowed us to assess any association of TL with stress related parameters. We found that, there was no correlation between TL of children and the exposure to chronic stress (p=0.319) or maternal depression (p=0.40). These results follow a number of reports in the literature, some used in a recent meta-analysis (Ridout et al 2016). It argues that this effect if exists, is relatively small. Further, we found a positive correlation between TL and increase in cortisol reactivity to the play stress (r= 0.181, p=0.000) and ground cortisol reactivity levels (r= 0.166, p=0.001). These results do not support the observation by Gotlib et al. (2015), who found a negative relationship between TL and cortisol in 97, 10-14 year old girls. Also, they do not support the implication that TL is related to stressful environment in very young children. Longitudinal studies are required on these children to offer a better insight on the relationship between TL and stressful environment and children's responses to stress.

3338T

Single cell allele specific expression (ASE) in T21 and common triso-mies: A novel approach to understand Down syndrome and gene dosage effects in aneuploidies. *G. Stamoulis[1], F. Santoni[2], P. Makrythanasis[2], A. Letourneau[1], M. Guiponni[2], M. Garieri[1], N. Panousis[1], E. Falconnet[1], P. Ribaux[1], C. Borel[1], S.E. Antonarakis[1,2,3].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Geneva University Hospitals-HUG, Service of Genetic Medicine, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, Geneva, Switzerland.

Trisomy 21(T21), is a model disorder of altered gene expression. Previous studies on aneuploidies were conducted on cell-populations or tissues. In our lab, in order to reduce the interindividual genetic variability, we have previously used a pair of monozygotic-twins discordant for T21 to study the global dysregulation of gene expression, (*Nature:*508;345-350;2014). However studies on gene and allelic-expression of single-cells(SC) may provide a different perspective on the biological and cellular impact of aneuploidy and in particular on the understanding of the fundamental mechanisms of gene dosage imbalance. We estimated the allele-specific-expression(ASE) from RNAseq of ~1000 single-cells in different aneuploidies:352 SC fibroblasts (172 Normal-179 T21 cells) from the pair of monozygotic-twins discordant for T21, 166 from a mosaic-T21, 176 mosaic-T18, 151 mosaic-T8, and 146 SC-fibro-blasts from mosaic-T13.In the monozygotic-twins, a considerable number of heterozygous sites genomewide were monoallelically expressed on a single cell level (Normal:73.5%-564,668 sites, and T21:78.7%-549,799 sites). We detected an increased monoallelic expression for chr21 sites in T21 cells with respect to Normal (Normal:63,3%-5,009 observations,T21:72.8%-6,456 observations). We classified chr21 genes in three classes based on the level of the aggregate monoallelic-expression of their corresponding heterozygous sites (9-monoallelic,29-intermediate,2-biallelic). Notably,monoallelic genes do not present the expected gene dosage imbalance (1.5 fold change) at single cell level as measured in bulk whole tissue studies.Similar results were also observed in the other aneuploidies. Here we demonstrate that, for monoallelic genes, the altered gene dosage observed in whole tissue studies induced by the aneuploid chromosome,is partially due to the higher fraction of cells expressing the gene on the supernumerary chromosome. This difference in the fraction of expressing cells could perturb the spatial and temporal development of the embryo and contribute to the development and variability of phenotypes in aneuploidies. We show that there is a larger fraction of trisomic single cells, compared to normal single cells, expressing genes on the extra chromosome contributing to the gene dosage imbalance mechanism.This study provides a new fundamental understanding of gene dosage effects in aneuploidies by exploring the gene dosage imbalance mechanism on a single cell level.

3339F

Delineating the impact of rare copy number variations on cognition in schizophrenia. *C. Lowther[1], J. Wei[2], G. Costain[1], D.J. Stavropoulos[3], C.R. Marshall[2], A.C. Lionel[2], S.W. Scherer[4], A.S. Bassett[5].* 1) University of Toronto, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada.

Background: The risk for developing schizophrenia (SZ) is 3-5% higher in individuals with an intellectual disability (ID) than in the general population. Individuals with comorbid ID are often excluded from genetic studies of SZ, leaving a gap in our understanding of the underlying genetic architecture of SZ-ID. Objectives: The purpose of this study was to 1) determine the yield of rare copy number variations (CNVs) in a cohort of adults with SZ-ID and 2) determine if the prevalence of rare CNVs is higher in SZ-ID compared to SZ-only and controls. Methods: 482 adults of European ancestry with SZ were ascertained from four psychiatric outpatient clinics in Eastern Canada. We used a broad definition of premorbid ID (FSIQ <85) to assign individuals to the SZ group with impaired cognition (n=160; SZ-ID). Those with FSIQ ≥ 85 were assigned to the SZ-only group (n=322). Using rigorous methods to detect genome-wide CNVs from high resolution microarrays we examined the burden of rare (<0.1%) CNVs >25 kb between these two groups and n=319 population-based controls of European ancestry. Results: The prevalence of subjects with one or more rare autosomal CNVs in the SZ-ID group was significantly higher than in SZ-only and controls across three CNV size cut-offs: >25 kb, >100 kb and >500 kb. The largest effect size for SZ-ID vs. SZ-only, and SZ-ID vs. control, comparisons was for two or more rare deletions >100 kb overlapping an exon (OR=9.28, CI 95% 1.14-246.99, p=0.0084), and one or more deletions >500 kb overlapping a gene (OR=5.21, CI 95% 2.17-13.91, p<0.0001), respectively. Compiling all SZ cases together revealed a significant negative correlation between FSIQ and the number of genes overlapped by a rare CNV >25 kb (Pearson r = -0.25, p=0.004). Identification of genes overlapped by a rare CNV in two or more SZ-ID cases and no SZ-only cases or controls suggests transcriptional regulation (*ZNF804A*, *LBH*), cortical formation (*ARHGAP11B*), synaptic signaling (*NRXN1*) and microtubule formation (*TUBB2A*, *TUBB2B*, *KIF25*) may represent novel etiological mechanisms associated with SZ-ID. Conclusions: To our knowledge this is the first study to use high resolution microarrays to examine the genetic architecture of SZ-ID. These results suggest there is an overall increased burden of genic deletions in SZ-ID compared to SZ-only. Examination of the genes overlapped by these CNVs may help identify novel pathways that are more specific to SZ-ID.

3340W

Low-copy repeats mediate genomic rearrangements of chromosome 22q11.2 in five probands from Brazilian public health service. *L.B. Minasi[1,2], A.J.C. Leite[1,2], I.P. Pinto[1,3], D.M.C. Cunha[1], C.L. Ribeiro[1,3], C.C. da Silva[1,2,3,4,5], A.D. da Cruz[1,2,3,4].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Eseffego, Brazil.

The instability of 22q11 has been demonstrated by the high frequency of pathological rearrangements of this region. This instability is attributed to the presence of several large paralogous low copy repeats (LCRs), each containing a complex modular structure and a high degree of sequence identity. Non-allelic homologous recombination between these LCRs causes a spectrum of recurrent rearrangements, including deletions and duplications. In total, the region contains eight LCRs, known as LCR22-A to LCR22-H. Herein, we performed an analysis of duplicated genomic sequences including known LCRs surrounding the proximal 22q11.2 locus identified in five probandos who carried deletions and duplication. Chromosomal Microarray Analysis using Affymetrix GeneChip® CytoScanHD™ array showed at 22q11.2 region two cases with *de novo* deletions of 2.88Mb and 0.27Mb, one case with paternal inherited deletion of 0.75Mb, and two cases with maternal inherited of 0.34Mb and 0.79Mb. Using the Segmental Duplication track of the Human Genome Build 36.1 comparing 3 times of the CNV size surrounding the proximal 22q11.2 locus (chr22: 18,640,000—25,080,000) against itself, we recognized LCR-A to LCR-D flanking 2.88Mb deletion and LCR-B to LCR-D flanking 0.75 Mb deletion. For the 0.27Mb deletion, 0.34Mb and 0.69Mb duplications breakpoints we only identified LCR-H within the 22q11.2 region. Deletion at 22q11.2 was usually sporadic and was reported to be inherited in 6 to 28% of patients with DiGeorge Syndrome. In general, duplications are more likely to be inherited than their reciprocal deletions, with approximately 93% of small duplications. The most common deletion occurs between the two largest blocks, LCR22-A to LCR22-D, and results in a 3Mb deletion. We identified two groups of LCR22 that are flanking the genomic rearrangements from in two probands. These finding suggest that the occurrence of genomic rearrangements might be mediated by NAHR between of the LCRs, increasing the susceptibility to the generation of CNVs. On the order hand, the three other cases showed LCRs within the reported deletion and duplications breakpoints, thus it is not possible to speculate the detailed mechanistic in relation to how this deletion and duplication initially appeared in these patients. The recognition of LCRs provides important insights related to the role of genomic architecture in chromosomal rearrangements, chromosome evolution, and in human disease. .

3341T

Acadian variant of Fanconi syndrome is caused by mitochondrial respiratory chain complex I deficiency due to a non-coding mutation in complex I assembly factor NDUFAF6. *L. Piherova[1], H. Hartmannova[1], K. Tauchmannova[2], K. Kidd[3], P.D. Acott[4], J.F.S. Crocker[4], Y. Oussedik[5], M. Mallet[6], K. Hodanova[1], V. Stranecky[1], A. Pristoupilova[1], V. Baresova[1], I. Jedlickova[1], M. Zivna[1], V. Robins[3], M. Vrbacky[2], P. Pecina[2], V. Kaplanova[2], J. Houstek[2], T. Mracek[2], Y. Thibeault[7], A. Bleyer[3], S. Kmoch[1].* 1) Institute for Inherited Metabolic Disorders, Charles University, Prague 2, Czech Republic; 2) Institute of Physiology of the Czech Academy of Sciences, Prague, Czech Republic; 3) Section on Nephrology, Wake Forest School of Medicine, Medical Center Blvd., Winston-Salem, NC, USA; 4) Section of Pediatric Nephrology, Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia; IWK Health Center, Halifax, Nova Scotia, Canada; 5) Department of Pathology, Dr. Georges L. Dumont University Hospital Centre, Moncton, New Brunswick, Moncton, New Brunswick, Canada; 6) Section of Pulmonology, Dr. Georges L. Dumont University Hospital Centre, Moncton, New Brunswick, Canada; 7) Section of Nephrology, Dr. Georges L. Dumont University Hospital Centre, Moncton, New Brunswick, Canada.

The Acadian variant of Fanconi Syndrome refers to a specific condition characterized by generalized proximal tubular dysfunction from birth, slowly progressive chronic kidney disease and pulmonary interstitial fibrosis. This condition is found only in Acadians, a founder population in Nova Scotia, Canada. The genetic and molecular basis of this disease is unknown. We carried out whole exome and genome sequencing and found that nine affected individuals were homozygous for the ultra-rare noncoding variant chr8:96046914 T>C; rs575462405, whereas thirteen healthy siblings were either heterozygotes or lacked the mutant allele. This variant is located in intron 2 of *NDUFAF6* (NM_152416.3; c.298-768 T>C), 37 base pairs upstream from an alternative splicing variant in *NDUFAF6* (chr8:96046951 A>G (c.298-731 A>G); rs74395342. *NDUFAF6* encodes NADH:ubiquinone oxidoreductase complex assembly factor 6, also known as C8ORF38. We found that rs575462405 - either alone or in combination with rs74395342 - affects splicing and synthesis of NDUFAF6 isoforms. Affected kidney and lung showed specific loss of the mitochondria-located NDUFAF6 isoform and ultrastructural characteristics of mitochondrial dysfunction. Accordingly, affected tissues had defects in mitochondrial respiration and complex I biogenesis that were corrected with NDUFAF6 cDNA transfection. Our results demonstrate that the Acadian variant of Fanconi Syndrome results from mitochondrial respiratory chain complex I deficiency. This information may be used in the diagnosis and prevention of this disease in individuals and families of Acadian descent and broadens the spectrum of the clinical presentation of mitochondrial diseases, respiratory chain defects, and defects of complex I specifically.

3342F

CLEC16A constrains NK cell cytotoxicity and protects against auto-immune susceptibility by regulating mitophagy and immune-mediated cell destruction. *R. Pandey[1], M. Bakay[1], S. Yoeun[1], A. Yermakova[2], J.D. Roizen[3], J.K. Kushner[4], J.S. Orange[2], H. Hakonarson[1,5].* 1) Centre for Applied Genomics, Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Section of Immunology, Allergy, and Rheumatology, Department of Pediatric Medicine, Texas Children's Hospital, Houston, TX; 3) Division of Endocrinology and Diabetes Children's Hospital of Philadelphia, Philadelphia, PA; 4) Section of Pediatric Diabetes and Endocrinology, Department of Pediatric Medicine, Endocrine-Metabolism, Texas Children's Hospital, Houston, TX; 5) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

The C-type lectin-like domain family 16A (CLEC16A) is implicated in a majority of known autoimmune diseases via a previously unknown mechanism. Genetic variants in the 16p13 locus harboring the CLEC16A gene are associated with the susceptibility to T1DM. However, the role of CLEC16A and of CLEC16A variations in the physiology and pathogenesis of T1DM remains poorly understood. We hypothesize that the molecular link between CLEC16A and autophagy/mitophagy is abnormal in T1DM, and that subjects with T1DM have accelerated mitophagy that a) increases autoimmunity and b) exhausts the energy sources of the insulin-producing β cells resulting in the expression of clinical diabetes. **Results:** We generated Clec16a inducible knockdown (KD) mice and CLEC16A overexpressing Natural Killer (NK) cell line to examine the mechanism underlying the link between NK CLEC16A expression alteration and autoimmune susceptibility. Clec16a KD mice showed increased NK cell cytotoxicity, increased expression of NK activating receptors and accelerated mitophagy in splenic immune cells. Treatment with autophagy inhibitors Bafilomycin and Wortmannin offset the effects of Clec16a knockdown. We show that CLEC16A/Nrdp1 pathway regulates mitochondria integrity and function and that uncontrolled loss of CLEC16A leads to increased mitophagy and intrinsic pathway of cell death. KD of CLEC16A in human NK cells doubled their cytotoxicity, whereas CLEC16A overexpression led to reductions in cytotoxicity, expression of NK cell activating receptors via autophagy, DC maturation and IFN-γ release. Our mechanistic studies reveal that CLEC16A restrains NK cell functions by modulating expression of NK receptors via C Vps-HOPS complex, CART and autophagy. Thus, owing to its role in autophagy, Clec16a impacts the function of NK cells and the risk of autoimmunity. **Conclusions:** Our results establish a functional link between CLEC16A and mitophagy/autophagy in the context of autoimmunity which explains the immune dysregulation in Clec16a KD mice and may explain the risk of autoimmunity in humans with specific CLEC16A risk variants. Thus, drugs modulating mitophagy/autophagy may be specifically effective as treatment in individuals with associated CLEC16A risk variants.

3343W

The combinative effect of copy number variation (CNV) in 47,XXY, a sex chromosome variation (SCV). *C. Keen[1], T. Sadeghin[1,2], C. Samango-Sprouse[3,4].* 1) The Focus Foundation, Davidsonville, MD; 2) Neurodevelopmental Diagnostic Center, Crofton, MD; 3) Pediatrics, George Washington University, Washington, DC; 4) Molecular Genetics, Florida International University, Miami, FL.

Introduction: 47,XXY is a SCV associated with low fertility, tall stature, and language-based learning disorders. The prevalence of CNV, and subsequent effects on phenotypic variation, is not clearly understood. We report the case of two male siblings (SL and SB) of the same maternal and paternal origin. SL presents with normal IQ, reading disorder and has 8q22.3 duplication (516kb) suggesting trisomy in this location. SB presents with the identical CNV, 47,XXY, severe speech delay and intact IQ. Statement of purpose: Investigate the neurodevelopmental profile of male siblings with the same CNV, and one sibling with 47,XXY additionally. Methods: Chromosomal Microarray (CMA) was completed on both siblings and mother. Father is deceased, therefore CMA could not be completed and origin of the CNV is unknown. Neurodevelopmental assessments probed domains of language, cognition, and motor. Results: SL developmental milestones were within normal limits and no parental concerns. At 9 years, SL demonstrated average IQ, severe reading disorder with above-average performance in mathematics. On the Woodcock Johnson Reading Mastery, he had standard scores (SS) in word identification (SS=67), attack (SS=73), and comprehension (SS=64). He had SS=55 in passage comprehension and SS=57 in reading comprehension. SB had delayed ambulation skills, short stature, and weak visual motor skills. 47,XXY was identified after chromosomal analysis ordered at 23 months because of developmental delay. At 31 months of age SB had increased head circumference (75th percentile) disproportionate to weight (5th percentile) or height (25th percentile). Bayley Scales of Infant Development showed composite scores of 88 in motor and 95 in cognition. Expressive language was severely delayed, and he had trouble imitating motor and speech actions. SB's profile was characteristic of Childhood Apraxia of Speech (CAS), with more intact auditory comprehension (SS=76) than expressive communication (SS=56) on the Preschool Language Scale-5. Discussion: SB & SL have neurodevelopmental profiles indicative of motor planning deficits with language-based disorders. SB had CAS, which is not characteristic of 47,XXY but may be evidence of further compromise by the CNV. SL has reading disorder which is language-based but no signs of CAS. The comparative complexity of SB's neurodevelopmental profile suggests possible combinative effects of CNV and SCV and variable effects neurodevelopmentally.

3344T

Gene dosage effects in Down syndrome revealed by proteome-wide expression and turnover analysis. *Y. Liu[1], C. Borel[2], L. Li[3], T. Mueller[1], P. Boersema[4], P. Germain[5], G. Testa[5], A. Beyer[3], S. Antonarakis[2], R. Aebersold[1,6].* 1) Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Zurich, Switzerland, Switzerland; 2) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva; 3) University of Cologne, Cologne, Germany; 4) Institute of Biochemistry, Department of Biology, ETH Zurich, Zurich, Switzerland; 5) Department of Experimental Oncology, European Institute of Oncology, Milan, Italy; 6) Faculty of Science, University of Zurich, 8057 Zurich, Switzerland.

   Statement of purpose Down Syndrome (DS) is caused by the whole or partial trisomy of chromosome 21 (Trisomy 21, T21). However, how T21 impacts human functional proteome remains unclear. Previous proteomic studies suggested the gene copy number alterations (CNAs) only weakly impact on protein expression. We herein hypothesize that proteome-wide turnover analysis would hold a special promise in revealing the gene dosage effects of T21.   **Methods** We investigated the effect of the extra Chr21 at the levels of transcript quantity, proteome quantity and protein turnover rate. We analyzed the primary fetal skin fibroblasts derived from a pair of monozygotic twins discordant for T21, which uniquely allowed us to characterize the proteome changes due to T21 without the noise of genomic variability. To validate, we also analyzed the fibroblasts from 11 unrelated T21 individuals and 11 controls. We applied the cutting-edge SWATH mass spectrometry to reproducibly and accurately measure the proteomes. We further quantified protein degradation rates through a pulse SILAC experiment (pSILAC) up to 24 hours.   **Results and Discussion** We quantified 4056 unique proteins for expression and ~2200 proteins by pSILAC experiment for protein turnover rates in both normal and T21 twins. The T21/normal fold-change correlation between transcript and protein levels was extremely low, indicating substantial post-transcriptional regulation and buffering effects in T21. Overall, the protein degradation was faster in trisomy cells than the controls. Remarkably, those Chr21 encoded proteins that are members of heteromeric protein complexes were largely exempt from responding to CNAs, primarily through accelerated protein degradation. Moreover, we found that both mitochondrial and cytosolic ribosomal proteomes were degraded heavily in T21, but different degree of translational regulation shaped their final, divergent expression levels. Combining twin samples and unrelated individuals, we found that organelle specific proteomic variation between individuals may contribute to some variable T21 phenotypes.   **Conclusion** Our data suggests that protein specific degradation presents a primary mechanism of proteome remodeling in response to T21 and thus reinforces the understanding of "gene dosage imbalance" in Down Syndrome.

3345F

Chromosome therapy: Correction of chromosomal aberrations by inducing ring chromosomes in induced pluripotent stem cells (iPSCs). *T. Kim, K. Plona, S. Carney, A. Wynshaw-Boris.* Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH.

   Approximately 1 in 500 newborns are born with chromosomal abnormalities that include trisomies, translocations, large deletions and duplications. There is currently no therapeutic approach for correcting such chromosomal aberrations *in vivo* or *in vitro*. We have attempted to utilize a novel pathway for aberrant ring chromosome rescue that we recently uncovered to correct large-scale chromosomal aberrations. When we produced induced pluripotent stem cell (iPSC) models from patients with either a ring chromosome 17 or 13, we found that the ring chromosomes were eliminated and replaced by a duplicated normal copy of chromosome 17 or 13 via uniparental isodisomy (Bershteyn *et al.*, 2014, Nature 507:99). We hypothesized that a chromosome with a large aberration could be corrected by inducing the aberrant chromosome into a ring chromosome in iPSCs, which would then be eliminated and replaced by a normal chromosome which we termed "Chromosome Therapy." To test this, we induced chromosome 17 ring formation in control iPSC lines via a Cre/loxP approach. LoxP sites were inserted by gene editing at sites near each end of chromosome 17 in iPSC clones containing a terminal deletion of 17p. We used different color fluorescent proteins (Cyan and Venus) to confirm proper insertion of the sites into the 17p and 17q arms of chromosome 17 with the combination of antibiotic selection. Once this was accomplished, we infected these cells with Cre recombinase to induce ring formation of chromosome 17. We designed fluorescence *in situ* hybridization (FISH) probes to segregate the clones containing the specific targeted-chromosome 17 from the non-specifically targeted clones. To visualize cells that form a ring chromosome, we designed and cloned the partial tandem dimer Tomato (tdTomato) along with the loxP sites such that the cell will express tdTomato upon formation of a ring chromosome due to Cre-mediated recombination. Individual tdTomato-positive clones were identified, and we are isolating sufficient numbers of cells for karyotype and molecular analysis. Loss of tdTomato from these ring-containing clones will suggest the loss of the ring and duplication of the normal chromosome 17 by uniparental isodisomy. A similar approach is being tested in iPSCs from patients with Down syndrome (trisomy 21). Therefore, we have created a specialized system to test whether "chromosome therapy" for correcting large chromosomal aberrations by the induction of ring chromosomes is possible.

3346W

Chromatin looping and endogenous L1-endonucleases may drive local shattering during chromothripsis. *L. Nazaryan-Petersen[1], B. Bertelsen[2], M. Bak[1], L. Jønson[3], N. Tommerup[1], D.C. Hancks[4], Z. Tümer[2].* 1) Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Copenhagen N, Denmark; 2) Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Glostrup 2600, Denmark; 3) Center for Genomic Medicine, Copenhagen University Hospital, Rigshospitalet, Copenhagen O. 2100, Denmark; 4) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

Chromothripsis is a phenomenon where multiple localized double-stranded DNA breaks result in complex genomic rearrangements. Although the DNA-repair mechanisms involved in chromothripsis have been described, the mechanisms driving the localized "shattering" process remain unclear. High-throughput sequence analysis of a familial germline chromothripsis revealed an inserted SVA$_E$ retrotransposon associated with a 110 kb deletion displaying hallmarks of L1-mediated retrotransposition. Our analysis suggests that the SVA$_E$ insertion did not occur prior to or after, but concurrent with the chromothripsis event. We also observed L1-endonuclease potential target-sites in other breakpoints. In addition, we found four *Alu* elements flanking the 110 kb deletion and associated with an inversion. We suggest that chromatin looping mediated by homologous *Alu* elements may have brought distal DNA regions into close proximity facilitating DNA cleavage by catalytically active L1-endonuclease. Our data provide the first evidence that active and inactive human retrotransposons can serve as endogenous mutagens driving chromothripsis in the germline.

3347T

Complex *cis*-interaction is responsible for the craniofacial and neuroanatomical defects of the 4p16.1 copy number variant. *C. Golzio, G. Hayot, C. Bonnet, N. Katsanis.* Center for Human Disease Modeling, Duke Univ Medical Center, Durham, NC.

Copy number variants (CNVs) are frequent lesions involved in both rare and complex disorders that often involve the dosage misregulation of numerous genes. We have shown previously how the use of structural surrogate phenotypes in zebrafish embryos and genomic studies can dissect a single contributory locus to these phenotypes, which raised the possibility that this approach might be useful in dissecting systematically CNVs. Here we demonstrate the utility of this paradigm by dissecting a CNV on 4p16.1 CNV, deletions and duplications of which causes a syndromic form of autism with associated macrocephaly and facial dysmorphia. The CNV encompasses 5 genes: *DRD5*, *WDR1*, *ZNF518B*, *CLNK*, and *SLC2A9*. Analysis of existing cases in Decipher indicated that the 4p16.1 deletion was associated with micrognathia (small jaw) and microcephaly whereas the duplication was associated with an abnormal facial shape and macrocephaly. We thus sought to determine the contribution of these five genes to brain and face development. To mimic the duplication, we expressed each of the five human transcripts in zebrafish embryos. We found discrete drivers for the two major anatomical features tested. First, overexpression of either *SLC2A9* and *ZNF518B* were sufficient to induce macrocephaly. However, scoring for the possible drivers of the craniofacial defect neither gene induced appreciable pathology. In contrast, expression of either of two other genes, *CLNK* and *WDR1,* led to macrognatia (*CLNK*) and abnormal U-shaped of the Meckel's cartilage (*WDR1*). Finally, we asked whether the same transcripts might be relevant to the deletion by inducing deletions in each of *WDR1* and *CLNK* orthologs by CRISPR/Cas9; in contrast to the duplication experiment, only the loss of *Wdr1* led to micrognatia. Taken together these data suggest that the craniofacial and neuroanatomical phenotypes are due to the dose imbalance of several genes present in the 4p16.1 following a *cis*-interaction complex model rather than the effect of a major gene driver. Moreover, our data intimate that the phenotypes of deletion and duplication carriers might be caused by the combinatorial dosage imbalance of different genes, as opposed to a simper reciprocal model.

3348F

***De novo* sequence and assembly of the human primary cell line WI-38 genome.** *M. Roy[1], P. Peluso[2], C.S. Chin[2], G. Concepcion[2], A. Hastie[3], I. Soifer[1], J. Lee[3], N. Fong[1], M. Chan[1], A. Ireland[1], A. Levinson[1], D. Botstein[1].* 1) Calico Life Sciences LLC, South San Francisco, Ca; 2) Pacific Biosciences, 1380 Willow Rd., Menlo Park, CA 94025; 3) BioNano Genomics, 9640 Towne Centre Drive, Ste. 100, San Diego, CA 92121.

A robust method for *de novo* sequencing and assembly of individual human genomes is an important prerequisite for achieving truly personalized medicine. Currently most whole genome sequencing projects rely on technology that produces "short reads." Attempts at *de novo* assembly of short reads result in highly fragmented assemblies, so short reads are aligned to a standard reference genome that itself still contains gaps and misassemblies. Complex genomic structural variants (SVs) can affect human health (e.g., autism, MHC regions) but alignment of short reads against a reference genome introduces bias towards the reference genome's SVs, and obscures SV differences between haplotypes. Here we illustrate how long reads enable the determination of reference-unbiased sequence using long scaffolds of sufficient contiguity to determine haplotypes. We describe sequencing and *de novo* genome assembly of WI-38. This diploid untransformed human fibroblast line is of considerable medical and scientific importance; it is used in the production of human vaccines (e.g. MMR, varicella, etc.) and in 1965 it was used to demonstrate the Hayflick limit of cell division. **Collapsed Assembly:** WI-38 was sequenced to a depth of 80X on the PacBio RSII using P6-C4 chemistry. *De novo* assembly was accomplished with the Falcon assembler, yielding 2.92 Gb on 2,910 contigs. Assembly contigs had an N50 of 24.9 Mb, with the longest covering 109.1 Mb. Assembled contigs were scaffolded by optical mapping on the BioNano Genomics Irys platform using Nt.BspQI and Nb.BssSI nickases. Assembly statistics improved significantly with the mapping data, producing a genome of only 119 scaffolds with an N50 of 84.9 Mb. The longest scaffold appears to cover completely Chr8 with 143.9 Mb, spanning the centromere. Other large scaffolds included 138.9 Mb on Chr4, 133.6 Mb on Chr2, 110.7 Mb on Chr6, 108.6 Mb on Chr5, and 101.7 Mb on Chr3. **Phased Assembly:** PacBio contigs were unzipped with a diploid-aware FALCON assembler, and haplotype-resolved with FALCON-Unzip. After unzipping, the assembly had fewer contigs (1,952) and a longer N50 (26.4 Mb). Phasing produced 1.94 Gb of haplotigs (phased contigs), indicating 33% of the genome occurs in regions of limited heterozygosity. Haplotigs had an N50 of 0.37 Mb with the longest haplotig covering 2.52 Mb. Further haplotyping efforts will include adding the mapping data and refining the assembly via improved haplotype-aware Falcon and optical mapping algorithms.

3349W

Somatic mutations accumulate in stem cells of human skeletal muscle with aging. *I. Franco[1], K. Olsson[2], A. Johansson[3], P. Lundin[4], C. Bosia[5], T. Gustafsson[2], H. Fischer[2], M. Eriksson[1].* 1) Biosciences and nutrition, Center for Innovative Medicine, Karolinska Institute, Huddinge, Sweden; 2) Dept of Laboratory Medicine, Div of Clinical Physiology, Karolinska Institute, Sweden; 3) Science for Life Laboratories, Dept of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; 4) Science for Life Laboratories, Dept of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; 5) Human Genetics Foundation, Torino, Italy.

Tissue aging is characterized by progressive loss of function and renewal. In skeletal muscle, progressive decline in mass, activity and repair are related to the functional impairment of the resident stem cell population, the satellite cells. Aged satellite cells show reduced proliferative capacity and high levels of senescence. One trigger for senescence is DNA damage, which suggests that mutations might accumulate in the genome of aged satellite cells and compromise their function. In this study, we have analyzed the occurrence of somatic mutations in single satellite cells from young (age 21 and 24, n=2) and old (age 64-78, n=3) healthy individuals. Cells isolated from fresh *m. vastus lateralis* biopsies were selected for expression of the satellite cell marker CD56 and single-cell plated via fluorescence activated cell sorting (FACS). Single cells were clonally expanded to amplify DNA and tested for myogenic commitment by staining for desmin. Five clones/individual were whole genome sequenced obtaining a 15x coverage of 86.6%-90% of the genome. For each clone, comparison to muscle and blood sequences of the same individual was used to call somatic variants. To allow for the identification of variants present in the single cell originally isolated from the tissue, only variants displaying 0.35-0.65 alternate allele frequency were included. On average, 652±41 somatic single nucleotide variants (SNVs) and 38.6±3.4 somatic indels/genome were identified in young individuals. Old individuals showed a two-fold higher number of somatic variants/genome (1314±93 SNVs and 87.3±16 indels, young vs old SNVs p=5.2x10[-6], indels p=1.8x10[-8]), indicating that adult satellite cells accumulate 15 somatic mutations/year. 1.1% of mutations were located in exons and deleteriously impacted protein coding genes known to be expressed in satellite cells at a rate of 5 genes/aged cell. Our results provide the first evidence of an age-related increase in somatic mutations in human satellite cells and suggest that higher mutation burden in the stem cell compartment can contribute to muscle aging.

3350T

A novel high throughput pooled multiplex assay for relative telomere length measurement. *F. Jasmine, J. Shinkle, M. Sabarinathan, H. Ahsan, B.L. Pierce, M.G. Kibriya.* Public Health Sciences, University of Chicago, Chicago, IL.

   Background: Relative Telomere Length (RTL) is a potential biomarker of aging and cancer. We recently developed a non-PCR, probe based RTL assay using QuantiGene plex chemistry on Luminex platform, where probes for Telomere (T) and reference gene (R) for a given DNA sample were tested in a single well (multiplexing T and R). Here we describe, for the first time, a method of pooling multiple samples in the same well to increase the throughput and cost-effectiveness. **Material and Methods**: We used four different micro-beads (T1, T2, T3 and T4) for the same T probe and four different micro-beads (R1, R2, R3 and R4) for the same R probe. Each pair of probe set (T1R1, T2R2, T3R3 and T4R4) was used for a single plate to hybridize on to DNA. Corresponding wells of four such plates were pooled in a single plate for all the subsequent steps for signal amplification, attaching detection dye, wash and final detection of signal intensity on a Luminex 200. Quantification of the signal intensity was done by 8-point standard curve generated from 5-PL algorithm. RTL was calculated as T/R. We used blood DNA samples from 60 independent individuals and repeated in multiple batches to test the precision. Coefficient of variation (CV) for each sample was calculated from the replicates as standard deviation/mean and expressed as percentage. **Results:** By using four different bead sets, we could process a total of 256 independent samples (four 96-well plates each containing 64 samples including controls) at a time in one run. The precision of the assay was checked by the CV. The geometric mean of CV for all samples was 8.04%, arithmetic mean of CV was 8.41% and the median CV was 8.3%. There was no statistical significant difference of RTL between batches (p=0.09). More than 64% of the variation in the RTL data could be explained by sample-to-sample variation; only 0.02% variation was due to batch-to-batch variation and 1.2% variation was explained by bead-to-bead variation. Previously we have validated the Luminex results by comparing to TRF by Southern blot. Our data provides proof of principle for effectively increasing the throughput of RTL Luminex assay. Theoretically, we can further increase the throughput to 1280 samples per run by pooling up to 20 bead sets. **Conclusion:** We describe a novel pooled multiplex assay with currently known highest throughput for average RTL. The assay has reasonable precision and can be used cost-effectively for large-scale studies.

3351F

High throughput single-molecule mapping links subtelomeric variants and long-range haplotypes with high-resolution telomere length analysis. *H. Riethman[1], E. Young[2], S. Pastor[2], K. Lassahn[1], R. Rajagopalan[2], J. McCaffrey[2], J. Sibert[2], A. Mak[3], P-Y. Kwok[3], M. Xiao[2,4].* 1) Medical Diagnostic and Translational Sciences, Old Dominion University, Norfolk, VA; 2) Drexel University, School of Biomedical Engineering, Philadelphia, PA; 3) Cardiovascular Research Institute, University of California, San Francisco; 4) Institute of Molecular Medicine and Infectious Disease, School of Medicine, Drexel University, Philadelphia, PA.

   Accurate maps and DNA sequences for human subtelomere regions, along with detailed knowledge of subtelomere variation and long-range telomere-terminal haplotypes in individuals, are critical for understanding telomere function and its roles in human biology. Here, we use a highly automated whole genome mapping technology in nano-channel arrays to analyze large terminal human chromosome segments extending from chromosome-specific subtelomere sequences through subtelomeric repeat regions to terminal (TTAGGG)n repeat tracts. We establish detailed maps for subtelomere gap regions in the human reference sequence, detect many new large subtelomeric variants, and demonstrate the feasibility of long-range haplotyping through segmentally duplicated subtelomere regions. These features make the method a uniquely valuable new tool for improving the quality of genome assemblies in complex DNA regions. Based on single molecule mapping of telomere-terminal DNA fragments combined with quantitative analysis of physically linked (TTAGGG)n tracts specifically labeled in vitro using CRISPR-Cas9 directed methods, we estimate individual telomere lengths linked to distinguishable telomeric haplotypes; this single-telomere genotyping method opens the door to new high-resolution analyses of telomere length regulation and function in human health and disease.

3352W

Analysis of chitinases from nonpathogenic bacterium *Listeria innocua* .
S. Honda, S. Wakita, Y. Sugahara, F. Oyama, M. Sakaguchi. Department of Chemistry and Life Science, kogakuin University, Hachioji, Tokyo, Japan.

Listeria monocytogenes is causes a human listeriosis, a disease that affects immunosusceptible population such as pregnant women and the elderly. *L. monocytogenes* has two chitinase genes *(lmo0105* and *lmo1883)*, whose genes product have been identified as an important virulence factor associated with enhanced pathogenicity, and enhances the virulence of *L. monocytogenes* through suppression of the host innate immunity. *Listeria innocua* is a nonhemolytic and nonpathogenic bacterium. It is proposed that *L. innocua* and *L. monocytogenes* have evolved from a common ancestor, and the differential pathogenicity is due predominantly to the loss of virulence genes in *L. innocua*. *L. innocua* has putative chitinase genes *(lin0153* and *lin1996)* similar to *L. monocytgenes*. To identify the functional property of the putative chitinase genes, we expressed the putative chitinase genes *(lin0153* and *lin1996)* in *Escherichia coli*. The purified recombinant gene products (LIN0153 and LIN1996) exhibited chitinolytic activity for artificial and natural substrates. LIN0153 and LIN1996 showed optimum catalytic activity under neutral and acidic conditions at 50 °C, respectively, and were stable over broad pH (4-11) and temperature (4-40 °C) ranges. Both enzymes produced mainly dimers from colloidal chitin as a substrate. However, LIN0153 and LIN1996 could hydrolyze oligomeric substrates with processive exo- and nonprocessive endo-manner, respectively, and showed different reactivity toward oligomeric substrates. Thus, we identified two genes *(lin0153* and *lin1996)* encoding functional chitinases in *L. innocua*.

3353T

Protein A-mouse Chit1-V5-His expressed in *Escherichia coli* **possesses chitinase functions comparable to CHO-expressed protein.**
M. Kimura[1], S. Wakita[1], K. Ishikawa[1], K. Sekine[1], S. Yoshikawa[1], A. Sato[1], K. Okawa[1], A. Kashimura[1], D. Yamanaka[2], N. Ohno[2], M. Sakaguchi[1], Y. Sugahara[1], F. Oyama[1]. 1) Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan; 2) Laboratory for Immunopharmacology of Microbial Products, School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, Hachioji, Tokyo, Japan.

Chitinases hydrolyze the β-1-4 glycosidic bonds of chitin, which is a major structural component of fungi, crustaceans and insects. Although mammals do not produce chitin, they express two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase). Chit1 levels are 1000-fold-elevated in the plasma of patients with Gaucher disease, an autosomal recessive lysosomal storage disorder. Chit1 have been shown to be closely associated with another diseases such as chronic obstructive pulmonary disease, Alzheimer disease and cystic fibrosis. However, the contribution of Chit1 to the pathophysiology of these diseases and pathophysiological conditions remains to be determined. In this study, we first expressed mouse mature Chit1 fused with V5 and (His)6 tags at the C-terminus (Chit1-V5-His) in the cytoplasm of *Escherichia coli* and found that most of the expressed protein was insoluble. In contrast, Chit1 tagged with Protein A at the N-terminus and V5-His at the C-terminus, was expressed in the periplasmic space of *E. coli* as a soluble protein and successfully purified. We produced mouse Chit1 as a recombinant fusion of Protein A-Chit1-V5-His, which was comparable to CHO cells-expressed Chit1. By our technique, it is possible to produce large amounts of active recombinant enzyme to study its pathophysiological roles as well as for potential therapeutic testing.

3354F

Loss and recovery of chitinolytic activity by amino acid substitutions in human acidic mammalian chitinase. *K. Okawa, M. Ohno, A. Kashimura, Y. Kobayashi, M. Sakaguchi, Y. Sugahara, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, tokyo, Japan.

Acidic mammalian chitinase (CHIA,;AMCase) is implicated in asthma and other allergic inflammations. Human and mouse AMCase share 82% sequence identity and 86% sequence similarity. Despite these structural similarities, these two homologues significantly differ with respect to their enzymatic characters at enzymatic activities. Multiple AMCase variants have been identified based on single nucleotide polymorphisms (SNPs) in humans and associations of certain AMCase polymorphisms and haplotypes with bronchial asthma have been reported. We reported the chitinolytic activity of the recombinant human AMCase was significantly lower than that of the mouse counterpart. By creating mouse-human chimeric AMCase protein we found that the presence of the N-terminal region of human AMCase containing conserved active site residues reduced the enzymatic activity of the molecule. Knowledge on genetic regulation of chitinolytic activity of AMCase is very limited. Here we report that the AMCase activity can be manipulated by naturally occurring amino acid substitution encoded by SNPs in human AMCase. Amino acid substitutions encoded by SNPs at the N-terminal region in human AMCase had distinct effects on the chitinolytic activity and we were able to reactivate the human AMCase, which were conserved in the mouse homologue.

3355W

Inactivation of mouse AMCase by introducing amino acid substitutions encoded by single-nucleotide polymorphisms in human AMCase gene. *F. Oyama, K. Okawa, M. Ohno, A. Kashimura, Y. Kobayashi, M. Sakaguchi, Y. Sugahara.* Dept Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan.

Acidic mammalian chitinase (AMCase) has attracted considerable attention due to its increased expression under specific pathological conditions. Elevated or reduced AMCase levels have been reported in numerous diseases such as asthma, allergic inflammation, ocular allergy, dry eye syndrome, stomach cancer, adenoid hypertrophy, conjunctivitis, neuromyelitis, gastritis or nasal polyp formation. However, the contribution of AMCase to the pathophysiology of these diseases remains to be determined. We have shown that the chitinolytic activity of the recombinant human AMCase was significantly lower than that of the mouse counterpart. It has been known that AMCase activity is altered by naturally occurring amino acid substitutions encoded by single-nucleotide polymorphisms (SNPs). In this study, we attempted to dissect the effect of amino acid substitutions on the chitinolytic activity of mouse AMCase. For abolition of the mouse AMCase activity, introduction of amino acid substitution associated with human SNPs was sufficient, indicating that the residue is essential for chitinolytic activity of the enzyme.

3356T

Gene expression analysis of mammalian chitinase transcripts in pig tissues. *E. Tabata, M. Ohno, M. Sakaguchi, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo 192-0015, Japan.

Chitin is the second abundant polysaccharide in nature. It is an integral component of the fungal cell walls, the exoskeletons of crustaceans and insects, and the microfilarial sheaths of parasites. Although mammals do not produce chitin and its synthase, genes encoding chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase) and their translation products have been found in both human and mouse. Chit1 levels are elevated in the plasma of Gaucher disease and the bronchoalveolar lavage fluid of smokers with chronic obstructive pulmonary disease (COPD). AMCase expression and activity is upregulated in allergic airway responses in mouse models of asthma. However, the expression levels of chitinases in the other mammals have remained unknown. Pig has been known as a biomedical model due to its similarities to human in physiologically and anatomically which can bridge the gaps between mouse and human. Here, we quantified expression levels of the chitinases and reference genes in pig tissues on the same scale using quantitative real-time RT-PCR system. Both chitinases were expressed in all pig tissues. AMCase mRNA was predominantly expressed in stomach. Chit1 mRNA was also highly expressed in eyes and liver, but their expression levels were lower than the levels of AMCase. Our study suggests that AMCase functions as a digestive enzyme that breaks down chitin and as a part of the host defense against chitin-containing pathogens in pig stomach tissues.

3357F

Functional analysis of acidic mammalian chitinase under physiological conditions. *S. Wakita, M. Kimura, A. Kashimura, M. Sakaguchi, Y. Sugahara, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.

Chitin is the second abundant polysaccharide in nature. It is an integral component of the fungal cell walls, the exoskeletons of crustaceans and insects, and the microfilarial sheaths of parasites. Chitinases are glycosidases that break down glycosidic bonds in chitin. Although mammals do not produce chitin, mice and humans express two active chitinases, chitotriosidase and acidic mammalian chitinase (AMCase). AMCase has attracted considerable attention due to its increased expression under certain pathological conditions related to immune response, for example in an induced asthma mouse model and antigen-induced mouse models of allergic lung inflammation. Furthermore, AMCase has been shown to be involved in eye and stomach diseases. Recently, we reported that AMCase mRNA is synthesized in the mouse stomach at exceptionally high levels, comparable to pepsinogen, suggesting a digestive role of AMCase. Here we performed an extensive analysis of chitin degradation by AMCase not only in strong acidic, but also weak acidic to neutral conditions followed by fluorophore-assisted carbohydrate electrophoresis (FACE), a method based on labeling the reducing ends of oligosaccharides with a fluorophore. We found that AMCase generates $(GlcNAc)_2$ at broad pH range of 2.0~8.0 as well as increasing amount of $(GlcNAc)_3$ with higher pH.

3358W

Systematic characterization of VNTR polymorphisms using deep whole genomes and comprehensive phenotype profiles of 2,255 Estonians. *T. Esko[1,2], M. Kals[3], T. Puurand[4], Estonian Biobank[1].* 1) Estonian Genome Center, University of Tartu, Estonia; 2) Broad Institute, Cambridge, US; 3) Institute of Mathematics and Statistics, University of Tartu, Estonia; 4) Institute of Molecular and Cell Biology, University of Tartu.

Human genome has been systematically characterized since early 2000s but the focus has mainly been on single nucleotide variation (SNV) while the structural variation (like copy number variation) have been less studied. Last year a comprehensive characterization of 1000 Genomes sample for tens of millions of structural variants, but so far systematic studies on tandem repeats, like STRs (short tandem repeats) and VNTRs (variable number tandem repeats), have not yet performed. While VNTRs where one of the first DNA variants to perform locus mapping and they are still widely used in forensics, their role in health and disease is still unknown due to the availability of deep sequenced whole genomes. We have used deep whole-genome sequences (coverage of 30x, Illumina X10PCRfree technology) of 2,255 individuals from Estonian Biobank cohort to systematically study VNTRs. We called the structural variants by k-mer methodology, where the BAM files are indexed for a list of sequence combination for oligomer of 25 nucleotides. We next downloaded all known tandem repeat sequences (N:230,306) from *Tandem Repeat Finder database*. After that we systematically inquired our indexed BAM files and counted the total number sequence combinations for each individual, while estimating the approximate copy number form from the average sequencing depth (enabled by *PCRfree* highly even coverage across the genome). In summary, we identified 68,318 VNTRs which had a repeating element longer than 24bp. On average, the number of repeats per VNTR was 15, while maximum was 3,500 repeats. We validated our approach through Sanger sequencing for several VNTR classes (repeat count accuracy above 75%). Next, we used our Estonian Biobank WGS cohort, which has available comprehensively set of omics profiles (including RNAseq, methylation, MS-LC and NMR metabolomics and a wide set of clinical labs, like example blood cell counts) to systematically estimate the downstream effects of long VNTR polymorphisms on biological processes and cellular pathways. As an example, we identified 240 robust cis-eQTLs for 110 genes from RNAseq from blood (N:566), the strongest eQTL explaining up to 36% of phenotypic variation (*DDT* gene). Further analysis where conducted using electronic health records. In conclusion, we have developed a novel computational framework to call VNTR polymorphisms and through systematic analysis within a phenotypically rich sample have demonstrated a wide range of phenotypic consequences.

3359T

Analysis of nonsense-mediated mRNA decay (NMD) using genomic and transcriptomic data. *K. Inui[1,2], M. Suyama[1].* 1) Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University, Japan; 2) Division of Bioinformatics, Systems Life Sciences, Kyushu University, Japan.

;Nonsense-mediated mRNA decay (NMD) is a surveillance mechanism that degrades mRNAs with premature stop codon to avoid production and accumulation of truncated proteins. Although the mechanism is mainly studied in detail for each gene under specific experimental conditions, there is little analysis focusing on how it works in genome-wide scale in normal individuals. With the progress of personal genomics and the associated transcriptome analysis, now it is possible to analyze NMD in genome-wide scale. ;In this study, we used the data from genomic and the corresponding transcriptomic data from the same individuals (n = 445). First, we counted how many heterozygous stop-gain mutations are occurred in these individuals. On average, a person has about 4 such mutations. The number is comparable to the frequency of nonsense mutations in other studies, such as those reported in many exome analyses. Then we analyzed the corresponding transcriptomic data for the peripheral blood of the same individuals to see if NMD could be observed as allelic imbalance in gene expression, because the transcript from the allele with premature stop codon must be selectively degraded by NMD. Indeed, we observed such allelic imbalance in the expression of the transcripts at the loci with heterozygous stop-gain mutations comparing to those with heterozygous synonymous mutations as a control. We further analyzed positional effect of the premature stop codons on the strength of NMD. As expected, transcripts with premature stop codons in the same exon as the original stop codon and those located close to the last exon junction escape from NMD, which is already reported in many cases. This supports that the observed allelic imbalance can be a real signal for NMD. Interestingly, we found that the transcripts with premature stop codons in the same exon as the translation initiation codon also escape from NMD. These and some additional results obtained from the genome-wide studies of NMD will be presented.

3360F

A systematic analysis of mobile elements' contribution to human transcriptomes. *P. Liang, A. Joshi, W. Tang.* Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada.

Mobile element insertions (MEIs) are a major component of the genome for most higher eukaryotic organisms. In the human genome, MEIs, consist of mainly LINE 1 elements (L1), Alus, SVAs, and endogenous retrovirus (ERVs), and together all MEIs contribute to more than 50% of the genome sequences with over four million copies. MEIs are known to play important roles in genome evolution and gene function through a plethora of mechanisms. In particular, MEIs are known to participate and regulate the transcriptomes via alternation of gene expression and RNA splicing, as well as direct participation in transcriptome. In this study, we performed a systematic analysis of the MEI's participation in human transcriptomes by utilizing the most updated genome sequences and annotation data. Overall, a total of 2,090,354 MEIs (42.8% of all MEIs) locate inside or in the proximity of 44,455 genes/transcripts. While most locate in the introns, 128,761 directly participated in the exons 12,493 protein-coding genes, and 80,070 contributed to 26,984 non-coding transcripts. These MEIs cover all major types of MEIs with Alu/MIR, L1/L2, ERV, hAT-Charlie/TcMar-Tigger/hAT-Tip100/TcMar-Mariner being the major contributors. MEIs also contributed to miRNA target sites for 1,921 miRNAs in 7,183 genes. Among the above, 6,800 MEIs participated in the coding regions (CDS) for 4,768 genes. While Alu and L1 are also the major contributors as in the case for all exons, SVAs seems to have a much higher ratio (at least 5 times higher than other MEI types) for participating in CDS, but not in the untranslated regions. Seven of these MEI-CDS are contributed by MEIs uniquely found in humans, limited only to members from SVA and L1HS and all contributing to rare alternative splice forms, perhaps also gene splitting in two cases. Six of these 7 genes have variations associated with various diseases.   In summary, our results demonstrate that mobile elements have made significant contributions to human transcriptomes via participation in protein coding, alternative splicing, mRNA regulation, and non-coding regulatory RNAs. With most of these events occurring in a lineage- and species-specific fashion, they are partially responsible for primate- and human-specific phenotypes.

3361W

Potential use of *de novo* **CNVs as biomarkers for parental exposure to low doses of ionizing radiation.** *E.O.A. Costa[1,2,6], I.P. Pinto[2,3], M.W. Gonçalves[2], A.S. da Cruz[1,2], C.C. da Silva[1,2,3,4,5], R.W. Pereira[6], A.D. da Cruz[1,2,3,4].* 1) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Goiânia, GO, Brazil; 5) State University of Goias, UnU Goiania Eseffego, Goiânia, Goiás, Brazil; 6) Catholic University of Brasilia, Genomic Sciences and Biotechnology PhD Program, SGAN, 916 Norte, Módulo B, Avenida W5, Bloco C, Sala 205, Brasília, DF, Brazil.

The radiological accident in Goiania in 1987 caused a trail of human contamination, animal, plant and environmental by a radionuclide. The mutagenic effects of ionizing radiation (IR) on the germ cell line are of particular concern because it can persist for generations, leading to the accumulation of additional mutations in the progeny of exposed parents. Recently it was established that Chromosomal Microarray Analysis (CMA) is an important tool for detecting wide spectra of deletions or duplications in the human genome. Using CMA is possible to detect genetic changes in critical regions of the chromosomes, usually not detected with traditional cytogenetic markers. The aim of this study was to establish the frequency of *de novo* CNVs to be used as a germline mutation marker, which could be applied to increase the knowledge on the biological effects of human exposure to ionizing radiation. The exposed group consisted of 12 families with at least one parent directly exposed to ionizing radiation of Caesium-137. A total of 40 subjects were included in this study: 12 parents and 16 offspring that were conceived and born after the accident. All exposed parents had an absorbed dose of ≤0,2 Gy. A group of 8 non-exposed families from Goiânia population was used as controls. The statistical test applied was the Mann-Whitney U. The analysis were performed using the statistical software SPSS 21.0 with a significance level of 5% (p <0.05). The frequencies of CNVs were estimated for loss/generation, gain/generation, and burden/generation, representing 2.2×10^{-5}, 1.0×10^{-5}, and 3.2×10^{-5}, respectively, for the exposed group. For the control population, the frequencies were 1.5×10^{-5}, 6.2×10^{-5}, and 2.2×10^{-5}, respectively. Thus, the frequencies of CNVs showed statistically significant differences between the exposed and control groups for microdeletions ($p < 10^{-3}$, Mann-Whitney U). For microduplications there was no statistically significant difference (p=0,76). Thus, our preliminary data showed that CNVs are induced by low-dose of IR exposure in a human population, while losses were more frequent than gains within the exposed group. Moreover, the progeny of a population accidentally exposed to very low-doses of IR showed ~1.5x more *de novo* CNVs than non-exposed controls. In conclusion, the frequency of germline mutation/generation in CNV is useful to retrospectively study human populations exposed to IR.

3362T

Lower human germline mutation rate in a proband born to accidentally exposed parents. *A.D. da Cruz[1,2,3,4,6], E.O.A Costa[1,2,6], I.P. Pinto[2,3], M.W. Gonçalves[2], A.S. da Cruz[1,2], C.C. da Silva[1,2,3,4,5], R.W. Pereira[6].* 1) Pontifical Catholic University of Goias, Genetics Master's Program, Rua 235, n. 40, Bloco L, Área IV Setor Universitário, Goiânia, GO, Brazil; 2) Pontifical Catholic University of Goiás, Department of Biology, Replicon Research Group; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Rede Centro Oeste de Pós-Graduação, Pesquisa e Inovação, Campus Samambaia, Goiânia, GO, Brazil; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goiás State for Public Health, Goiânia, GO, Brazil; 5) State University of Goiás, UnU Goiania-Esefego, Goiânia, Goiás, Brazil; 6) Catholic University of Brasilia, Genomic Sciences and Biotechnology Graduate Program.

In Goiânia-Brazil, an accident caused human and environmental contamination by Cesium-137. Exposure to IR results in DNA damage, from modified nucleotides to DSB. In this study, we report the results of Chromosomal Microarray (CMA) of a 20yo born to parents accidentally exposed to IR, with absorbed doses (AD) of ≤0,2Gy. The proband's findings were compared to a control group of 8 children from Goiânia whose parents had no history of exposure. At the conception, mother's age was 26yo and father's 24yo. Conception occurred one month after parental exposure. CMA was performed using CytoScan HD. Filters for gain and loss were set at 15 and 8 SNP markers, respectively. *de novo* CNVs were considered if the average marker distance was ≤2.000 kb and size of ³1kb. We identified 10 CNVs, ranging from 1 to 15kb, including 9 deletions and 1 duplication in the proband. In the control population, we observed 238 CNVs, 14 duplications and 224 deletions, with an average of ~30 CNVs/child ranging from 1 to 140 kb. In both proband and controls, deletions were more frequent than duplications. Germline mutation rate of CNVs in the proband was 3.5×10^{-6} while the control group was 2.2×10^{-5}; in the proband, mutation burden showed 6.3x lower than controls. As the parents incurred whole-body exposure, it's possible that germline cells suffered severe DNA damage. It has been estimated that for every 1Gy, 1000 SSB and 40 DSB are produced. DNA damage may have triggered *in vivo* adaptive response, which is known to remains effective *in vivo* for 3 cell cycles in human lymphocytes. DNA repair in gonadal tissues, within the month prior to conception, could be the reason behind the generation of gametes with lower *de novo* mutations, lowering the rate in F1. Human genomic instability (GI) following RI exposure depends on AD, type of irradiated tissue, and cellular genotype. Our results suggest caution in reaching generalized conclusions related with human exposure to low doses of IR. Whether there is a linear or exponential dose response and whether there is a discrete threshold, below which biological consequences are uncommon, remain unanswered. Similarly, we are uncertain of the extent to which factors as LET, dose rate and bystander effect affect the outcome of exposure. Essentially, the extent of human germline mutations arising from exposure to ionizing radiation remains a critical scientific issue and the mechanisms underlying the phenomenon of transgenerational GI remain unresolved.

3363F

Spatial organization of topologically associated domains in individual chromosomes. *S. Wang[1,2], J. Su[1,3], B. Beliveau[4], B. Bintu[1,5], J. Moffitt[1,2], C. Wu[4], X. Zhuang[1,2,5].* 1) Howard Hughes Medical Institute, Cambridge, MA; 2) Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA; 3) Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA; 4) Department of Genetics, Harvard Medical School, Boston, MA; 5) Department of Physics, Harvard University, Cambridge MA.

The spatial organization of chromatin critically impacts genome function. Recent high-throughput chromosome-conformation-capture studies have revealed topologically associated domains (TADs) as a conserved feature of chromatin organization. How TADs are spatially organized in individual chromosomes remains, however, largely unknown. Here we developed an imaging method to directly map the spatial positions of numerous genomic regions along individual chromosomes inside the nucleus and used this method to trace the positions of TADs in four human interphase chromosomes. We observed that chromosome folding deviated from the ideal fractal-globule model at large length scales. Our single-chromosome images revealed that TADs were spatially organized into two physical compartments that were positioned in a highly polarized fashion in individual chromosomes. Moreover, we observed that active and inactive X chromosomes differed from each other and from autosomes in terms of folding and compartmentalization, suggesting that the spatial organization of TADs can be changed in response to regulation.

3364W

Discovery of widespread inherited fusion genes in human populations.
C. Bandlamudi[1,2], V.K. Dhiman[1], B.E. Stranger[1,3], K.P. White[1,2,4,5]. 1) Institute for Genomics & Systems Biology; 2) Committee on Genetics, Genomics & Systems Biology; 3) Department of Medicine; 4) Department of Human Genetics; 5) Department of Ecology and Evolution, University of Chicago, Chicago, IL.

   Recent studies have highlighted the extensive genetic diversity contributed by structural variation (SV) in the human genome. However, the prevalence and contribution of fusion genes, a consequence of SVs, to the genetic variation among humans remains largely unknown. Here, we analyzed more than 9,000 transcriptomes from 524 individuals in the Genotype Tissue Expression project (GTEx) to identify fusion genes in healthy individuals. A median of 15 tissue transcriptomes per individual allowed us to control for various splicing artifacts that typically confound fusion discovery from RNA-seq. We identified 64 fusion genes (59 novel) segregating at overall frequencies ranging from 0.2% (singleton) to 42%. Using whole genome sequencing (WGS) based SV analysis and genomic PCR we show a validation rate of at least 88%. We show that 70% of the fusion genes identified in GTEx are also detected in at least one individual in the TCGA (n=9,360) and HapMap (n=451) cohorts suggesting that substantial proportion of them are inherited rather than de novo. 7/12 fusion genes recurrent in GTEx showed statistically significant difference is fusion frequencies between different populations within multiple cohorts. We are currently investigating if these regions show signals of recent selection. Functional consequences of these events are hard to elucidate. However, using gene expression quantifications, we identified twenty fusion genes that resulted in ectopic expression of the downstream partner gene in at least a median of 3.5 tissues in which the downstream partner gene is typically not expressed. For example, the fusion, TFG-GPR128 (2.4% frequency), resulted in ubiquitous expression of the G-protein coupled receptor domain of highly tissue-specific GPR128 gene. In another example, in one individual, we find the UTR-5 of CASP4 (Caspase 4) fused to the entire protein coding region of CARD18 (Caspase-1 inhibitor) resulting in potential deregulation of caspase pathways due to the production of caspase-1 inhibitor instead of caspase-4. Our findings, for the first time, demonstrate the pervasiveness of fusion genes in human populations and highlight their potential contribution to phenotypic variation.

3365T

Effect of natural variation in copy number on epigenetic patterning and gene expression.
J. Yu, C. Li, S. Iyengar. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

   Purpose. Copy number variation (CNV), the most common type of structural variation, has been associated with many human diseases, but its role in disease pathogenesis is unclear. We hypothesized that CNVs alter gene expression through differences in histone modification patterning and performed a systematic evaluation of *cis* relationships between CNVs and histone marks in two embryonic cell lines, H1 and H9. **Methods**. CNV calls for H1 and H9 were obtained either from ENCODE or with PennCNV on SNP array data from GEO. Only CNV calls confirmed from others sources are kept for analysis. Profiles of 26 consolidated histone marks were obtained from the Roadmap Epigenomics Project for the two cell lines. RNAseq data were obtained from GEO. To evaluate the effect of CNV presence/absence on histone modification, we **1)** compared histone mark intensity between H1 vs. H9 at CNV regions while adjusting for DNase intensity and peak region GC content; **2)** compared the number of histone mark peaks in specific CNV regions in H1 vs. H9, in relation to the rest of the genome; **3)** evaluated the effect of CNVs on the expression of genes that overlap with a CNV region. **Results. 1)** For 13 out of 26 marks, CNVs were significantly associated with the number of peaks after Bonferroni correction. Three of the marks, H2BK120ac (shown to be enriched in transcription start sites and gene bodies), H2BK20ac (shown to be enriched in transcription start sites and gene bodies), and H3K18ac, had decreased number of peaks in duplication regions, while the other 10 had increased number of peaks. Interestingly, among these significantly associated marks, H2BK120ac, H2BK20ac, H3K18ac, H3K14ac, H3K14ac, H3K4me1, and H4K20me1 were significantly associated with decreased peak intensity, suggesting a balancing act to counter duplication. **2)** No significant effect of CNVs on gene expression was identified. **3)** We observed that expression of some gene families was completely shut down, including the olfactory receptor family and defensins. Olfactory receptors are known to be epigenetically regulated. The duplication CNV straddled defensins on chromosome 20 is also a known copy variation region. **Conclusion.** Our study suggested that histone marks significantly associated with CNVs might influence gene expression through changing the number and/or intensity of histone mark peaks, and help counter-balance any effect of CNVs on gene expression.

3366F

Temporal analysis of hippocampal gene networks identifies transcriptional modules related to epileptogenesis in a rat model of febrile seizures. *H. Azevedo, N. Khaled, P. Santos, F.B. Bertonha, C.A. Moreira-Filho.* Pediatrics, University of Sao Paulo, Sao Paulo, Sao Paulo, Brazil.

   The incidence of febrile seizures (FS) during childhood constitutes as a relevant risk factor for the onset of epilepsy, although the molecular alterations induced by FS in the brain remain largely unknown. Experimental hyperthermia-induced seizures have been used in animal models to investigate the consequences of FS, as these animals show histological, behavioral and brain hyperexcitability changes after the initial seizures. The present study aimed at investigating temporal changes in hippocampal gene co-expression networks during the development of rats submitted to hyperthermia-induced seizures. Total RNA was obtained from the ventral hippocampal CA3 region of Wistar rats at four time points after hyperthermic-induced seizures and then subjected to oligonucleotide microarray profiling. A weighted gene co-expression network analysis was applied to investigate modules of co-expressed genes, since these genes may be associated to common biological functions. This analysis consisted of constructing gene co-expression networks, identifying network modules and hubs, performing gene-trait correlations and examining intramodular connectivity changes. The most relevant co-expression modules were then functionally enriched to identify molecular pathways and functions associated to hyperthermic seizures. The results overall showed that after experimental FS the brain undergoes molecular changes linked to developmental pathways, such as those regulated by Wnt, Notch, Hippo and MAP kinases. Interestingly, modules of co-expressed genes related to neurogenesis, neuronal differentiation, axonogenesis and synaptic transmission were activated as early as one day after the hyperthermia episode. Moreover, the appearance of immune-related modules indicates that pro-convulsant events can lead glial cells to release pro-inflammatory mediators, initiating a cascade of events that contribute to the development of seizures. In line with these evidences, gene expression alterations in ion channels and solute carrier transporters suggest that ion imbalance in the neural tissue may also play a role for neuronal hyperexcitability. Therefore, it is possible conclude that hyperthermic seizures induce an intense tissue and synaptic remodeling process that could turn the brain prone to epileptic activity.

3367W

***ATM* heterozygous mutations underlying individual differences in radiosensitivity in human populations.** *S. Matsuura[1], E. Royba[1], T. Miyamoto[1], S.N. Akutsu[1], T. Yamamoto[2], Y. Kudo[3], S. Tashiro[4].* 1) Genetics and Cell Biology, RIRBM, Hiroshima University, Hiroshima, Japan; 2) Department of Mathematics and Life Science, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8530, Japan; 3) Department of Obstetrics and Gynecology, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima 734-8551, Japan; 4) Department of Cellular Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima 734-8553, Japan.

   Ionizing radiation (IR) induces DNA double strand breaks (DSBs), which are an initial step for chromosomal aberrations and cell death to cause acute and/or chronic radiation syndrome. In human cells, DNA repair system recognizes and rejoins DSBs to maintain genomic stability and protect against IR-induced carcinogenesis. It has been suggested that the individual differences in radiosensitivity exist in human populations, and that the nucleotide variants on DNA repair genes might be the genetic determinant for the individual differences. Germline mutations of the *ATM* gene, which encodes a master kinase in DNA damage response, cause a radiation hypersensitive autosomal-recessive disorder Ataxia-Telangiectasia (A-T [MIM 607585]) characterized by neurodegeneration and cancer predisposition. A-T heterozygous carriers exist at around 1 % in human populations and have been suggested to show a high risk of cancer. It is therefore important for radiological protection whether A-T heterozygous carriers are indeed radiosensitive or not. Here we used CRIS-ObRiGaRe (Obligate Ligation-Gated Recombination) method combined CRISPR/Cas9 system with non-homologous end joining (NHEJ)-mediated knock-in technique to generate *ATM* +/- human cultured cell clones in order to evaluate the effect of *ATM* heterozygous mutations on radiosensitivity, independently of the diverse genetic background and confounding factors such as smoking. We also demonstrated that IR-induced micronucleus formation, which is derived from unrepaired DSB-induced chromosomal fragments, in *ATM* +/- cells is enhanced more than that in *ATM* +/+ cells. Taken together, *ATM* gene variants are indeed involved in individual differences in radiosensitivity in human populations.

3368T

Conserved and tissue-specific effects of natural genetic variation on transcript and protein abundance. *S. Munger[1], J. Chick[2], P. Simecek[1], K. Choi[1], E. Huttlin[2], D. Gatti[1], S. Gygi[2], R. Korstanje[1], G. Churchill[1].* 1) The Jackson Laboratory, Bar Harbor, ME; 2) Harvard Medical School, Boston, MA.

Genetic variation can influence protein expression through transcriptional and post-transcriptional mechanisms, and these effects may be conserved across tissues or specific to one. To characterize the shared and tissue-specific effects of natural genetic diversity on the proteome, we combined RNA-seq and multiplexed, quantitative mass spectrometry with a genetically diverse mouse population, the Diversity Outbred (DO) heterogeneous stock. We measured genome-wide transcript and protein abundance in livers and kidneys from 192 DO mice, and mapped quantitative trait loci that influenced transcript (eQTL) and protein (pQTL) expression. We identified nearly 3,000 pQTL in each of the liver and kidney, divided equally between local and distant variants. Local pQTL generally had larger effects on protein abundance, these effects were conferred primarily through transcriptional mechanisms, and half showed conserved protein responses in both tissues. In contrast, distant pQTL influenced protein abundance nearly exclusively through post-transcriptional mechanisms and most were specific to the liver or kidney. We applied mediation analysis and identified a second protein or transcript as the causal mediator for half of the significant distant pQTL. Furthermore, we identified groups of proteins within known pathways that shared coincident subthreshold distant pQTL for which we could identify a single causal protein intermediate from the same pathway, demonstrating the power of integrating ontology and mediation analyses to tease out subtle but real genetic effects from mapping populations with modest sample sizes. Overall, our analysis revealed extensive tissue-specific networks of direct protein-to-protein interactions that act to achieve stoichiometric balance of functionally related enzymes and subunits of multimeric complexes.

3369F

Improved mate-pair sequencing for enhancement of re-sequencing pipeline for improve detection of structural variations in genomes. *S. Lok, S. Walker, S.W. Scherer.* The Hospital for Sick Children, Toronto, Canada.

Whole genome sequencing as it is commonly practiced is actually "re-sequencing", in which next generation DNA sequence reads from a query genome are mapped onto a reference genome and the differences tabulated. Reads unique to the sample, including many variants such as insertions, deletions, translocations, and copy number alterations are typically discarded because they cannot be mapped due to significant differences from the reference genome. To capture a more complete repertoire of genomic variants contributing to disease, there must a greater reliance on personalized genome sequencing (also known as *de novo* sequencing), in which the individual's genome is reconstructed from their own data alone. Currently, the cost of *de novo* sequencing for large patient cohorts is prohibitively expensive and will likely remain so for the foreseeable future. Here we describe the development, use and performance metrics of a new improved chimera-free mate-pair sequencing technology that can be easily integrated into the present re-sequencing pipelines in a cost effective manner. This offers an interim solution to enhance the detection of structural variations until the advent of *de novo* assembly methods appropriate for use in large cohorts. Unique DNA barcode identifiers attached to the ends of individual size-selected target DNAs enable the unambiguous discrimination of true mate-pairs created by intra-molecular ligation from unwanted chimeric mate-pairs created by random inter-molecular ligation during library construction. This simple but powerful step eliminates the major impediment of previous generations of mate-pair sequencing. Using HuRef DNA, we provide sensitivity and specificity benchmarks for INDEL detection, particularly in the 100 to 10,000 bp size range that are under-reported using standard re-sequencing methods. We are also applying this approach in families with autism spectrum disorder presumed to carry mutations, which are still yet to be identified.   .

3370W

Homolog paring: Some speculations and some super-resolution imaging. *T. Wu[1], R.B. McCole[1], J. Erceg[1], G. Nir[1], J. AlHaj Abed[1], B.J. Beliveau[2], H. Sasaki[2], M. Hannan[1], P. Yin[2,3], E.F. Joyce[1,4].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA; 4) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Homolog paring: Some speculations and some super-resolution imaging Conventional wisdom argues that organisms are diploid because "two is better than one": should a chromosome or gene suffer damage, then the homolog of that chromosome or gene will compensate for the loss. The situation with diploidy may be more complex, however. We suspect this because genomes have evolved elaborate mechanisms by which they address homology and, in contradiction to the purported "two is better than one", several such mechanisms cause one of the homologs to be silenced. The silencing can be predetermined, such as in imprinting, or random, such as with random monoallelism and X-inactivation. Either way, it seems that homologs may coordinate their decisions. One way in which somatic homologs communicate may be through direct contact via homolog pairing. Although once thought to be the exclusive property of Drosophila, somatic pairing is now considered a potential feature of many organisms, including humans. Here, we pull from observations of a wide variety of phenomena, including published studies of mutation rates as well as our own work on sequence ultraconservation and homolog pairing and speculate that the state of inter-homolog interactions may reflect a balance of pairing and anti-pairing forces and that the function of pairing may extend far beyond the adage of "an heir and a spare". This presentation will also summarize our most recent work exploring homolog pairing and nuclear organization, in general, using Oligopaints and two methods we have developed for single-molecule super-resolution imaging, called OligoSTORM and OligoDNA-PAINT. This work was supported by awards from the NIH (to CtW, PY, EFJ), NSF (PY), Wyss Institute (PY), W,R. Hearst Fund (RBM), EMBO (JE), and Damon Runyon Foundation (BJB).

3371T

Validation of mutations in iPS cells using droplet digital PCR. *T. Takahashi, M. Nomura, F. Kitaoka, J. Kuwahara, S. Yamanaka, N. Takasu, N. Amano.* Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan.

In order to ensure safety of induced pluripotent stem (iPS) cells and differentiated cells, it is important to check mutations such as single nucleotide variations (SNVs), insertion-deletions (Indels) and copy number variations (CNVs). In our pipeline, SNV/Indels and CNVs are detected using whole genome sequencing, exome sequencing and SNP array. To validate mutation candidates, we use amplicon sequencing. To validate the results of amplicon sequencings further, we used ddPCR in the following two cases. Firstly, once mutations were detected in iPS cells, we want to check whether the same mutations are also in original cells (e.g., fibroblast, peripheral blood mononuclear cell, and cord blood) or not. When we use amplicon sequencing, it is difficult to fix a threshold to remove false positives. To use ddPCR, we designed FAM fluorescent probe on the alternative genome and HEX fluorescent probe on the reference genome. Genomic DNA molecules were partitioned into 15,000 droplets and mutations were detected by the FAM fluorescent signals of each droplet. We successfully detected SNV/Indels (<0.1%) and CNVs in original cells. Secondly, we used ddPCR to distinguish homogeneous iPS cell clones having a specific mutation that were cultured from single cells in one well, and found that it worked well. In this study, one SNV on the X chromosome was checked in 5 clones. These clones were classified into two cases having a specific mutation or not. The ddPCR is a useful tool for validating SNV/Indels and CNVs in iPS cells and differentiated cells. In near future, we will apply ddPCR to validate the inversions and translocations.

3372F

Genome-wide context-dependent conservation in humans at nucleotide resolution. *J. di Iulio, E. Wong, I. Bartha, M. Hicks, J. Yu, J.C. Venter, A. Telenti.* Human Longevity Inc., San Diego, CA.

Two key challenges in the interpretation of human genome sequencing are (i) the large numbers of rare variants and (ii) the fact that the largest proportion of variation occurs in the non-protein coding regions of the genome. We used data from 10,545 deep-sequenced human genomes (mean 30X coverage) to define the probability of variation at any nucleotide position in the human genome before it is observed. This is achieved through the massive alignment of genomic elements from the pool of genomes in order to gain precision and power (number of elements x number of genomes) to estimate human variation probability with nucleotide resolution genome-wide. The model uses all possible heptamers and their variation rates to define regional conservation and diversity on the basis of departure from expectation (context-dependent conservation, CDC). CDC requires no a priori knowledge (such as genomic annotation, variation pathogenicity, etc) and is based solely on observed human variation. It complements other metrics (f.ex. CADD, eigen, DEAPsea) used for the identification of functional elements in the genome. The best prediction is achieved by combining all metrics (median AUC ROC=0.77); however it outperforms other metrics when used alone (median AUC ROC=0.74). We further tested whether non-coding functional loci would be identified on the basis of CDC scores. We describe a >30 fold enrichment in the 1st percentile of CDC scores for pathogenic ClinVar and HGMD variants annotated in non-protein coding regions. As proof of concept, the assessment of manually curated Mendelian-associated trait variants in non-protein coding regions (N=64) confirmed a strong enrichment in the first percentiles of the CDC scores (42% within the 1st percentile; 53% within 5th percentile). These results indicate that CDC scores used alone or in combination with the other existing metrics can reveal functional parts of the genome and support prediction of the pathogenicity of novel variants.

3373W

Assessing the concordance in the prediction of potential cis-regulatory loci from DNAse-seq and ATAC-seq data. *J. Peralta[1], J. Charlesworth[2], J. Blangero[1].* 1) South Texas Diabetes and Obesity Institute, University of Texas at the Rio Grande Valley; 2) Menzies Institute for Medical Research University of Tasmania, Tasmania, AU.

The relevance of non-coding variation for gene expression is difficult to assess. For coding variation the mechanisms of transcription and translation provide a clear path to the interpretation of its biological meaning. In the case of non-coding variation, the elements and mechanisms involved, such as promoters, enhancers, the local chromatin organization or transcription factor binding accessibility and affinity, are more diverse. This diversity increases the complexity of the problem and makes it difficult to establish a conceptual leap from non-coding variation to its biological relevance. One approach that we have been exploring to prioritize non-coding variation is to use a measure of the allele-specific open chromatin accessibility of non-coding heterozygous loci as a proxy of their potential for cis-regulating neighboring genes. Specifically, we use DNAse-I hypersensitivity sites (DHSs) as indicators of open chromatin to assess, within a variance component framework, the contribution to the variance in gene expression phenotypes of a covariance kernel derived from loci at allele-specific DHSs. We have previously applied this approach using DHSs from publicly available GM12878 lymphoblastoid cell line DNAse-seq data from the ENCODE Project, and found signals of potential cis-regulation for ten transcripts in our San Antonio Family Study (SAFS) cohort. Recently an improved method to detect open chromatin, ATAC-seq, has been developed. ATAC-seq is a faster protocol, requires significantly fewer cells, and takes a different approach to the detection of regions of open chromatin: it uses an engineered transposase that inserts the sequencing adapters directly into exposed DNA. We are now applying our method to the publicly available GM12878 ATAC-seq data to evaluate the concordance with our previous DNAse-seq based results.

3374T

Discovery of *de novo* **and rare Alu elements using mobile element scanning.** *J. Feusier, D.J. Witherspoon, W.S. Watkins, L.B. Jorde.* Human Genetics, University of Utah, Salt Lake City, UT.

   Alu elements compose 11% of the human genome, and they influence gene expression, chromosomal recombination, and disease pathogenesis. Identifying rare, polymorphic *Alu* elements will provide additional markers for human population analyses and forensic testing. Furthermore, determining the per-generation *Alu* retrotransposition rate is essential for understanding genomic evolution and as a baseline for somatic retrotransposition rates. However, rare, polymorphic *Alu* elements are difficult to analyze due to the large genomic copy number (~1.1 million elements), the repetitive sequence of the element, and the large sample sizes needed for detection of novel, low-frequency or *de novo* insertions. *Alu*Yb8/9 elements are young subfamilies that contribute ~30% of the recent polymorphic *Alu* retrotransposition events and contain a unique 7bp insertion that is absent in other *Alu* elements. Mobile element scanning (ME-Scan), developed by the authors, is a targeted, high-throughput method that pulls down *Alu* elements with the 7bp insertion prior to sequencing. Because there are only several thousand Yb8/9 copies, 24-sample libraries can be run on a single Illumina lane, making ME-Scan a cost-effective method for studying specific mobile elements in large cohorts of individuals. To identify novel *Alu*Yb8/9 elements, 41 CEPH pedigrees were sequenced using ME-Scan along with 216 individuals from various populations around the world ("population sample"). In the pedigrees, we performed PCR validation of 33 candidate *de novo Alu* insertions and discovered a single *de novo* element. This element was an exact match to the *Alu*Yb8 sequence and was located in an intronic region. Because Yb8/9 is only one active *Alu* subfamily, further sequencing of these pedigrees is required before we can accurately estimate the per-generation *Alu* retrotransposition rate. In the population sample, we discovered and validated 68 rare, polymorphic *Alu* elements by PCR and Sanger sequencing. Interestingly, we identified an *Alu*Yb8 sequence with 3 substitutions shared among 5 elements that has all of the hallmarks of an *Alu* subfamily, which we name *Alu*Yb8c3. We also found an additional ~600 novel, candidate *Alu* elements in the dataset. In conclusion, we identified the first non-disease causing *de novo Alu* insertion event, a new *Alu* subfamily, and identified novel polymorphic *Alu* elements for population analyses.

3375F

Widespread splicing of repetitive elements into mRNA coding regions. *S. Sabunciyan, M.M. Darby, J.T. Leek, B. Langmead, R.H. Yolken.* Johns Hopkins University, Baltimore, US.

   We characterized expression from individual repetitive element loci in the human brain using directional RNA sequencing data. Considering only sequencing reads that map uniquely onto the human genome, we discovered that the overwhelming majority of intronic and exonic repetitive element loci are expressed in the same orientation as the gene in which they reside. Further analysis revealed that repetitive element loci are extensively spliced into coding regions of gene transcripts yielding thousands of novel mRNA variants with altered coding potential. Lower frequency splicing of repetitive element loci into untranslated regions of gene transcripts was also observed. The same pattern of repetitive element loci splicing was detected in the brain tissue of all seven species we examined, ranging from humans to the fruit fly, and was also found in human blood. Repetitive element loci splicing with exons occurs largely at the canonical GT-AG splice junctions and is not restricted to a particular repetitive element class. This type of splicing usually gives rise to a minor splice variant and *in silico* analysis suggests that in certain cases repetitive element loci splicing may introduce a novel open reading frame. Reanalysis of sequencing data performed on polysome associated RNAs in the mouse cerebellum revealed that many repetitive element loci splice variants might be translated into proteins. Our results demonstrate that repetitive element loci expression is more complex than previously envisioned and raise the possibility that repetitive element loci splicing might generate functional protein isoforms.

3376W

Functional somatic mutations are negatively selected during reprogramming of induced pluripotent stem cells but positively selected during subsequent passaging. *M. D'Antonio[1], D.A. Jakubosky[2,3], W.W. Greenwald[4], M. Donovan[3], C. DeBoever[4], E.N. Smith[5], H. Matsui[6], A. Arias[1], H. Li[6], A. D'Antonio-Chronowska[1], K.A. Frazer[1,5,6].* 1) Division of Genome Information Sciences, University of California, San Diego, LA JOLLA, CA; 2) Biomedical Sciences Graduate Program, University of California, San Diego, LA JOLLA, CA; 3) Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA; 4) Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA; 5) Department of Pediatrics and Rady Children's Hospital, University of California San Diego, La Jolla, CA; 6) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA.

To examine the functional effects of somatic mutations in iPSCs we conducted whole genome sequencing of 18 skin fibroblast-derived iPSCs and their matched blood DNA. We detected 46,612 somatic single nucleotide variants (SNVs), 2,776 small indels and 563 copy number variants (CNVs), ranging from 1 kb to 48 Mb. The number of SNVs and indels per sample is highly heterogeneous, ranging from 913 to 7,570 and is likely due to the presence of different numbers of somatic mutations in the cell of origin. Indeed, the eight samples with the highest number of mutations all harbor many CC-TT substitutions, which are known to be caused by UV damage and have been linked with melanoma. To investigate if CC-TT substitutions were under selection during iPSC reprogramming, we analyzed additional 255 blood and 19 skin fibroblast WGSs. These variants are strongly enriched in fibroblasts (335 CC-TT per sample, p-value = 8.7×10^{-11}) and iPSCs derived from fibroblasts (49, p = 4.8×10^{-11}), compared to blood (0.7), suggesting that the CC-TT substitution signature originates in skin fibroblasts. The fact that fibroblasts harbor more mutations than the iPSCs reprogrammed from fibroblasts indicates that fibroblasts with many CC-TT substitutions may be less likely to be successfully reprogrammed into iPSCs. To address functional differences between clonal (present in all cells in an iPSC line) and subclonal (present only in a fraction of the cells) SNVs, we investigated their associations with chromatin organization. By permuting the position of SNVs across the genome, we found that clonal SNVs are strongly enriched in quiescent chromatin (regions not associated with any ChIP-seq or DHS peaks, Z-score = 9.9) and heterochromatin (Z = 9.4), and depleted in enhancers (Z = -2.9), promoters (Z = -4.3) and transcribed regions (Z = -9.2). Conversely, subclonal SNVs are more likely to occur in promoters (Z = 0.5) and less likely in quiescent chromatin (Z = -2.6). These results show that clonal variants, which are likely derived from the cell of origin, tend to occur outside of functional regions, while subclonal mutations, which occur during passaging after reprogramming, are more likely to be present in active chromatin regions. Our data suggest that source tissue cells with high mutational burdens or with mutations that have phenotypic effects are less likely to be successfully reprogrammed; while mutations that arise shortly after reprogramming are likely to have functional effects.

3377T

Impact of 3D chromatin structure on cell type-specific gene expression in iPSC and iPSC-derived cardiomyocytes. *H. Li[1], A. Schmitt[2], P. Benaglio[3], E. Smith[3], N. Nariai[1], M. D'Antonio[1], H. Matsui[1], C. DeBoever[4], B. Ren[1,2], K. Frazer[1,3].* 1) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 2) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA; 3) Department of Pediatrics, University of California, San Diego, La Jolla, CA; 4) Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, CA.

The human genome is organized into 3D chromatin domains that play important regulatory roles in gene expression through long-range interactions. Recent studies have shown that large-scale topological structures are highly conserved across cell types; however, distinct chromatin contacts may exist between different cell types and drive cell type-specific gene expression. This has not yet been investigated due to the resolution of contact maps analyzed to date. In this study, we used high-depth Hi-C data to identify cell type-specific chromatin contacts in induced pluripotent stem cells (iPSCs) and iPSC-derived cardiomyocytes (iPSC-CMs), and determined if they are associated with cell type-specific gene expression. We assessed genome-wide chromatin interactions by performing high-depth Hi-C experiments using iPSC and iPSC-CM lines from 7 individuals. A total of 5.9 billion and 5.5 billion paired-end reads were generated for iPSCs and iPSC-CMs, respectively. Reads were aligned against human reference genome hg19 using BOWTIE2, followed by a stringent filtering process to remove abnormal chimeric pairs, self-ligations, and re-ligations. We built Hi-C contact maps at 10kb resolution, and assessed genome-wide differential contacts (DCs) using diffHic. We also conducted RNA-seq experiments on the same cell lines. A total of 16,217 genes passed quality control and were subjected to differential gene expression analysis using edgeR. We identified 3,762 DCs ($q<0.05$) and 6,101 differentially expressed (DE) genes ($q<0.01$, fold change>2) between iPSC and iPSC-CM. Of the 3,762 DCs, 1,277 (34%) had at least one anchor overlapping the transcription start site (TSS) of one of the 16,217 genes, and 792 (62% of 1,277) overlapped the TSS of one or more DE gene (in total 1,484 DE genes). By permuting the differential state of all 16,217 genes while maintaining the number of genes that were differentially expressed (6,101), we found that the TSS of DE genes were significantly enriched in the DCs ($p<1e-3$). Gene-set enrichment analysis showed that the 1,484 DE genes overlapping DC anchors share general gene ontology terms in common with the 4,617 DE genes that do not overlap DCs; but they are also uniquely enriched for cell-type specific terms including stem cell differentiation, embryonic morphogenesis, and cardiac septum development. These findings suggest higher-order chromatin structure plays an important role in the regulation of cell-type specific gene expression.

3378F

Insights into gene function and tolerance of homozygous variation in a highly endogamous population. *H.C. Martin[1], V.M. Narasimhan[1], K.A. Hunt[2], D. Mason[3], K.J. Karczewski[4], E.R. Maher[5], R.C. Trembath[6], D.G. MacArthur[4], J. Wright[3], J.C. Barrett[1], D.A. van Heel[2], R. Durbin[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK; 2) Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK; 3) Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford BD9 6RJ, UK; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 5) Department of Medical Genetics, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Box 238, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK; 6) Faculty of Life Sciences and Medicine, King's College London, London SE1 1UL, UK.

   Individuals from populations with a high rate of consanguineous marriage are enriched for rare, homozygous loss-of-function (rhLOF) mutations that can be informative about gene function. We sequenced the exomes of 3222 healthy British Pakistanis from Birmingham and the Born in Bradford cohort, and discovered rhLOF mutations in 781 genes, a substantial enrichment compared to discovery rates in outbred populations. We observed a 13.7% depletion of homozygous knockout genotypes, implying selection against deleterious recessive variants. We also observed a smaller but significant depletion of homozygous missense variants predicted to be highly deleterious by CADD. These data allow us to look at the tolerance of particular gene sets to homozygous variation, in a way that is not possible in large outbred cohorts. For example, in known developmental disorder genes that typically act in a dominant loss-of-function manner, we see a significant depletion of missense variation in autozygous regions (p = 0.0002), suggesting an alternative mode of pathogenicity. We found individual gene knockouts significantly associated with lipid and glycaemic traits, which we are currently seeking to replicate in a larger cohort. If validated, these associations could point to potential new drug targets. Finally, we are expanding this study as part of the East London Genes and Health project. We aim ultimately to exome-sequence 25,000 healthy British South Asian adults (enriched for consanguinity) with linked electronic health records. Pilot data on 535 individuals reveals a mean autozygosity of 4.5%, which we expect to increase in future sample selection. Additionally, we are identifying individuals with high autozygosity in the UK Biobank for sequencing. Analysis of the existing Biobank data suggests we should expect a total of ~1,200/500,000 samples with at least 5% autozygosity. We will present preliminary results from these projects, which are expected to greatly expand the catalogue of human knockouts, shedding light on gene essentiality and function, and well as lending greater power to investigate the tolerance of specific gene sets and genic regions to particular types of homozygous variation.

3379W

Population survey of copy number variation in over 85,000 exomes and associations with serum lipids from electronic health records. *E.K. Maxwell[1], C. O'Dushlaine[1], J.S. Packer[1], S.E. McCarthy[1], C. Gonzaga-Jaregui[1], J. Staples[1], W.A. Faucett[2], J.B. Leader[2], I.B. Borecki[1], F.E. Dewey[1], Y.I. Chen[3], J.I. Rotter[3], J.C. Cohen[4], W.K. Chung[5], A.M. Muise[6], E.G. Puffenberger[7], E.J. Reichenberger[8], D. Gaudet[9], H.L. Kirchner[2], M.F. Murray[2], M.D. Ritchie[2], D.J. Carey[2], D.H. Ledbetter[2], C.L. Martin[2], G.D. Yancopoulos[1], A.R. Shuldiner[1], A. Baras[1], O. Gottesman[1], L. Habegger[1], J.G. Reid[1].* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Geisinger Health System, Danville, PA; 3) Los Angeles Biomedical Research Institute, Torrance, CA; 4) University of Texas Southwestern Medical Center, Dallas, TX; 5) Columbia University Medical Center, New York, NY; 6) Hospital for Sick Children, Toronto, Ontario; 7) The Clinic for Special Children, Strasburg, PA; 8) University of Connecticut Health Center, Farmington, CT; 9) ECOGENE-21, Chicoutimi, Québec.

   Copy-number variants (CNVs), a class of structural variation, are relatively underrepresented in genomic catalogs due to difficulties in their ascertainment. We recently developed CLAMMS - a versatile and scalable method that produces high quality exome CNV calls across the allele frequency spectrum and as small as a single exon. We applied CLAMMS to over 85,000 exomes sequenced at the Regeneron Genetics Center (RGC) from multiple collaborations. First, we present a survey of exonic CNVs in 50,726 exomes from the DiscovEHR Study and report associations between both common and rare CNVs and serum lipid traits, demonstrating the feasibility of integrating exome CNVs with EHRs to gain insights into disease. We highlight a novel familial hypercholesterolemia-causing tandem duplication in *LDLR* identified in 29 individuals associated with high LDL cholesterol and coronary artery disease, report on a rare complex structural variant surrounding *HMGCR* (a ~1.5 Mb tandem duplication with a nested ~600 Kb deletion), and show a common *LILRA3* microdeletion that associates with increases in HDL cholesterol and is linked to a previously identified upstream intergenic GWAS hit. We also survey the prevalence of known CNVs in our population, showing that *PMP22* duplications and deletions occur at relatively equal rates, suggesting the deletion is under-diagnosed. Second, we expand our CNV survey to a diverse set of over 85,000 samples, providing a broad view of CNV allele frequencies across populations. We highlight CNVs that show differences in prevalence among populations of distinct ancestry, including *LILRA3* microdeletions in East Asians and duplications in the keratin gene *KRT34* in Africans. We demonstrate that CNV-based principal components analysis can stratify individuals by ancestry concordantly with SNP-based analyses.

3380T

Features of *Alu/Alu* mediated genomic rearrangement. *X. Song[1], C. Beck[1], I. Campbell[1], Z. Coban-Akdemir[1], S. Gu[1], P. Boone[1], C. Shaw[1], J. Lupski[1,2,3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Alu elements belong to a subfamily of short interspersed nuclear elements (SINEs), comprise ~11% of the human genome, and are often found at the breakpoints of genomic rearrangements. Previously, the involvement of repetitive elements in genomic rearrangements was proposed to be due to DNA recombination, because of their large copy number and high level of sequence identity. However, recent studies of copy number variants (CNVs) and complex genomic rearrangements (CGRs) mediated by *Alu* pairs indicate that these events may preferentially utilize replication-based mechanisms with the *Alu* elements facilitating template switching. *Alu/Alu* CNVs can cause Mendelian diseases such as spastic paraplegia, Fanconi anemia, and Von Hippel-Lindau syndrome. To date, however, only a fairly small number of cases (< 200) detailing these events at the molecular level have been published. Many studies are limited by the sequence coverage and mapping difficulty due to the relatively short read length of most genomic sequencing technologies. Therefore, a bioinformatics study on reported *Alu/Alu* rearrangements can potentially illuminate the underlying mechanism(s), further characterize regions prone to genomic instability in the human genome, and may eventually improve molecular diagnosis and clinical treatment. In this study, 187 CNV-*Alu* pairs with experimental evidence supporting their role in deletion formation have been collated from published data; these experimentally determined *Alu/Alu* events were used as positive control to examine molecular features common to the mechanism. For each CNV-*Alu* pair, 1000 non CNV-*Alu* pairs within each generated region were randomly picked as a negative control dataset. Individual features were analyzed for all *Alu* elements and differences between positive and negative control datasets were tested. Statistically, when compared to non CNV-*Alu*s, CNV-*Alu*s tend to be longer in length, have higher GC content, a longer poly-A tail, and higher identity of A box and B box motifs.

3381F

Constructing a map of dispersed duplications in the human genome using whole genome sequencing data from families. *C.W. Whelan[1,2,3], R.E. Handsaker[1,2,3], G. Genovese[1,2,3], M. Lek[1,4], F. Zhao[1,4], J. McElwee[5], M.C. O'Donovan[6,7], D.G. MacArthur[1,4], S.A. McCarroll[1,2,3].* 1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA; 5) Merck Research Laboratories, Merck & Co, Boston, MA, USA; 6) MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK; 7) National Centre for Mental Health, Cardiff University, Cardiff, UK.

A longstanding challenge in the analysis of copy number variations (CNVs) has been to identify duplications that are dispersed – those in which the duplicated segment has inserted at a different genomic locus than its source (as opposed to tandem duplications, in which duplicate copies reside next to one another). Dispersed duplications can hide paralogous gene variants, distort linkage disequilibrium patterns, and make functional analysis of variants difficult, so a map of common dispersed duplications would be useful in many areas of genomics. We developed a way to utilize whole genome sequencing data from quartets (in which both parents and two offspring have been sequenced) to identify and map dispersed duplications. First, we develop an HMM-based method to determine the identity-by-descent (IBD) state of every locus in the genome in sibling pairs from each quartet. We then identify CNVs for which the copy numbers of the duplicated segment in the four family members is incompatible with the inheritance pattern determined by the IBD state between the siblings at that locus. By searching for CNVs that violate IBD inheritance patterns in multiple families, we can identify dispersed duplications. We then use the computed IBD states across the entire genome in every sibling pair in our data set to map the location of the dispersed segments to regions where every quartet's copy numbers and IBD states are compatible. We applied this method to over 200 quartets, identifying and mapping an initial set of over forty common dispersed duplications. We verify predicted dispersed duplications using linked-read data generated with the 10X Genomics technology, finding that the predictions of our approach were correct in 88% (15/17) of currently evaluable examples. By expanding this approach to more than 700 quartets, we are now building a comprehensive map of dispersed duplication polymorphisms that can be used to improve read alignment and variant calling in the affected genomic loci, remove confounding signals from association studies, and increase our understanding of genetic diversity.

3382W

Genome-wide copy number variation analysis and association with facial shape variation in a large cohort of Bantu Africans. *F. Yilmaz[1,2], M. Sorenson[3], D. Astling[2,4], H-C. Yu[2], J. Cole[2,5], S.A. Santorico[3,5,6], R.A. Spritz[2,5], A.E. Hendricks[3,5,6], T.H. Shaikh[2,5].* 1) Integrative and Systems Biology Program; 2) School of Medicine; 3) Mathematical and Statistical Sciences; 4) Department of Biochemistry and Molecular Genetics; 5) Human Medical Genetics and Genomics Program University of Colorado Denver; 6) Biostatistics and Informatics, Colorado School of Public Health.

The face is one of the most distinguishable characteristics of the human body. Family-based analyses of facial measurements have provided evidence that facial shape and morphology have a strong genetic component. However, the specific genetic factors underlying normal facial shape appearance remain largely unknown. Copy number variations (CNVs) are a significant source of normal genetic variation as well as the genetic burden associated with human disease. CNVs have been implicated in the etiology of numerous disorders resulting in multiple birth defects, including craniofacial abnormalities. We hypothesized that some variability in normal facial appearance results from CNVs which affect dosage of genes that modulate normal facial development. To test this hypothesis, we used genotype and phenotype data from ≈3700 Bantu Africans analyzed as part of the FaceBase I initiative. These individuals have been genotyped using high-density SNP-based microarrays and photographed using 3D morphometric cameras, which enable quantification of facial distances and shapes that explain the majority of facial shape variance. CNV data were generated from the raw data files obtained after genotyping with Illumina SNP arrays, using three CNV calling algorithms: PennCNV, DNAcopy and VanillaICE. CNVs were required to be concordant between two or more of the algorithms with an overlap of ≥50%, in order to reduce the number of false-positive calls. Further filtering for probe content, genomic size and segment logRatio was applied to obtain a set of high confidence CNVs. Our preliminary results detected CNVs in several genes previously implicated in craniofacial development and disorders, including *SHH, PAX3, FGFR1, FGFR2* and *IRF6*. We will present association analysis of single variants at a frequency ≥1% and gene-region based variants at frequency ≤5% to 34 quantitative facial phenotypes including 25 linear distance measures between facial landmarks, landmark derived centroid size and allometry (a measure of growth rate), head circumference, the first five Principal Components (PCs) from PCA of the whole face and the first PC from PCA of the midface. In addition to providing a comprehensive analysis of the relationship between CNVs and facial morphology, our study also provides a unique source of information about CNVs in a large African sample.

3383T

Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy. *M. Zamani Esteki[1,7], A. Destouni[1,7], M. Catteeuw[2], O. Tšuiko[1,3], E. Dimitriadou[1], K. Smits[2], A. Kurg[3], A. Salumets[4,5], A. Van Soom[2], T. Voet[1,6,8], J. Vermeesch[1,8].* 1) Centre for Human Genetics, University Hospital Leuven, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium; 2) Department of Obstetrics, Reproduction and Herd Health, Ghent University, Ghent, 9820, Belgium; 3) Institute of Molecular and Cell Biology, Tartu University, Tartu, 51010, Estonia; 4) Competence Centre on Health Technologies, Tartu, 50410, Estonia; 5) Department of Obstetrics and Gynecology, University of Tartu, Tartu, 51014, Estonia; 6) Sanger-EBI Single Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, United Kingdom; 7) These authors are joint first authors and contributed equally to this work; 8) Co-senior authors.

Dramatic genome dynamics, such as chromosome instability, contribute to the remarkable genomic heterogeneity among the blastomeres comprising a single embryo during human preimplantation development. This heterogeneity, when compatible with life, manifests as constitutional mosaicism, chimerism, and mixoploidy in live-born individuals. Chimerism and mixoploidy are defined by the presence of cell lineages with different parental genomes or different ploidy states in a single individual, respectively. Our knowledge of their mechanistic origin results from indirect observations, often when the cell lineages have been subject to rigorous selective pressure during development. Here, we applied haplarithmisis to infer the haplotypes and the copy number of parental genomes in 116 single blastomeres comprising entire preimplantation bovine embryos (n = 23) following in vitro fertilization. We not only demonstrate that chromosome instability is conserved between bovine and human cleavage embryos, but we also discovered that zygotes can spontaneously segregate entire parental genomes into different cell lineages during the first post-zygotic cleavage division. Parental genome segregation was not exclusively triggered by abnormal fertilizations leading to triploid zygotes, but also normally fertilized zygotes can spontaneously segregate entire parental genomes into different cell lineages during cleavage of the zygote. We coin the term "heterogoneic division" to indicate the events leading to noncanonical zygotic cytokinesis, segregating the parental genomes into distinct cell lineages. Persistence of those cell lines during development is a likely cause of chimerism and mixoploidy in mammals.

3384F

Discovery and annotation of duplicate genes using long-read capture and transcript assembly. *M.L. Dougherty[1], J.G. Underwood[1,2], E. Tseng[2], B.J. Nelson[1], O. Penn[1], L. Harshman[1], M.Y. Dennis[3], E.E. Eichler[1,4].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Pacific Biosciences, Menlo Park, CA; 3) University of California, Davis, CA; 4) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

High-quality sequence assembly and accurate gene annotation are key to understanding genetic variation, but are often frustrated by segmental duplications (SDs). SDs comprise 5% of the human genome and contain approximately 1,000 genes in whole or in part. They are an important source of genetic instability associated with both rare and common disease and provide templates for evolutionary innovation. However, our understanding of these genes still lags behind the rest of the genome. Fundamental knowledge, such as gene models, which describe exonic structure and protein-coding potential, are still incomplete, incorrect, or even lacking for many duplicate genes. We developed a method that would yield full-length transcript information while also being able to distinguish between nearly identical gene copies. Our approach uses complementary biotinylated oligonucleotide probes to enrich for duplicate genes from first-strand cDNA. Targeted cDNA is amplified, size fractionated, and sequenced using single-molecule, long-read (PacBio) technology. Long-read sequencing permits us to distinguish between highly identical transcripts and can probe complex splicing patterns. We designed 516 probes to 20 gene families that underwent duplications specifically on the human lineage since divergence from chimpanzee. Sequence analysis of captured cDNA from fetal and adult brain reveals a 251-fold enrichment in the abundance of our targeted genes over non-enriched control. Mean transcript sizes ranged from 1200-2300 bp with transcripts up to 4 kbp identified with high confidence, while we are continuing to target larger size bins. Among the human-specific duplications, we observe new isoforms including novel sites of transcription initiation and termination as well as fusion transcripts with neighboring genes. We also observe previously unannotated open-reading frames (ORFs), indicating potential novel genes that have previously been missed. Our method provides a powerful approach to accurately define full-length transcripts, isoforms, and ORFs for new genes mapping to duplicate regions and will be more broadly applicable to other structurally complex regions of genomes.

3385W

Somatic mutation profiles from blood cells of healthy individuals reveals strong replication-associated mutational asymmetry. *W. Chen, Y. Hong, D. Zhang, C. Zeng.* Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, Beijing, China.

Somatic mutations rather than inherited from germ cells, usually occur due to unrepaired DNA damages and DNA replication errors. This type of mutations may contribute to genotypic and phenotypic heterogeneity, and even jeopardize the human health. Most studies on somatic mutations focused on their association with pathogenesis of diverse diseases. According to current disease development models, somatic mutations take place continually and accumulate prior to disease occurrence. However, the process of the mutation accumulation remains poorly understood. To well characterize somatic mutations, we set up an optimized library preparation protocol and probability based variation calling method, and analyzed blood samples from four healthy individuals and their parents. By sequencing a mutation-free control DNA, we first obtained a priori trinucleotide-specific sequencing error rate, ranging from 0.0001 to 0.017. And the probability of a call as to be a mutation or an error was then calculated accordingly. In all, 493 mutations including 3 de novo germline mutations were detected across the 4 samples with an average of 123 mutations per sample, and the lowest allele fraction was 0.4%. Furthermore, we also generated a simulated mutation profile assuming a consistent mutation rate among different generations during cell division. The allele fraction distribution for real muations showed a bias toward low allele fraction ones comparing to this simulation. One possibility for this phenomenon is the change in mutation rate during the process of cell development. Subsequently, we summarized the percentages for all mutations at each of the 96 mutated trinucleotides for each individual, a significant enrichment of C>T was observed among all individuals, especially the mutated cytosine adjacent to guanine (namely CpG). Additionally, we also found that most of the C>T mutations occurred in the earlier time during replication in comparison with tumor somatic mutations. Meanwhile, C>T mutations was more frequently observed in the lagging-strand template than in the leading-strand template, showing the strand asymmetry for somatic mutations in normal cells. One possible explanation is that the lagging-strand templet may experience a longer exposure as single-stranded DNA which consequently result in a higher rate of cytosine deamination than that occurred in double-stranded DNA.

3386T

Genome-wide effects of engineered chromosomal translocations on transcription and genome architecture. *S. Kim[1], V. Ramani[1], C. Disteche[2], J. Shendure[1,3].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, Seattle, WA.

The spatial organization of the genome within the nucleus is closely inter-twined with gene expression at multiple levels, ranging from chromosome localization to positioning of chromatin loops. However, it remains unclear to what extent each level of spatial organization drives changes in gene expression. Chromosomal translocations have previously been shown to perturb both the spatial organization of the genome and global gene expression, but a more complete understanding of these changes requires examining the effects of translocations in mammalian cells in isolation from other genetic or epigenetic differences. To address this challenge, we used CRISPR/Cas9 to induce translocations between an intron of the gene *HPRT1* and an intergenic region on either chromosome 8 or 21, and then selected for loss of functional HPRT1 to enrich for cells carrying translocations. Using this approach, we generated monoclonal cell lines derived from the near-haploid cell line HAP1, carrying t(X;8)(q26;q23), t(X;8)(q26;q24), or t(X;21)(q26;q21). Existing models of how chromosomal translocations impact gene expression suggest that expression changes would be primarily on the derived chromosomes, which might be localized in a novel region of the nucleus or have a novel chromatin structure near the breakpoints. We tested this hypothesis by both profiling the transcriptional changes in our translocation cell lines using RNA-seq and their changes in genome conformation using Hi-C. We find changes in gene expression spread throughout the genome, but only minimal changes in genome conformation at ~1 Mb resolution. Further, we find that even unmodified monoclonal cell lines derived from the same HAP1 population exhibit high variability in gene expression, suggesting that the variable effects of cell culture on gene expression can exceed those of engineered translocations.

3387F

Characterizing extreme structural rearrangement complexity in the human germline using whole-genome sequencing. *C. Redin[1,2,3], R.L. Collins[1,4], H. Brand[1,2,3], C. Hanscom[1], F. Quintero-Rivera[5], M. Devito[6], M. Talucci[6], R.J. Wapner[6], W.P. Kloosterman[7], M.E. Talkowski[1,2,3].* 1) Center for Human Genetic Research, Massachusets General Hospital, Boston, MA; 2) Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Program in Bioinformatics and Integrative Genomics, Harvard University, Boston, MA; 5) Department of Pathology & Laboratory Medicine, UCLA Clinical Genomic Center, UCLA David Geffen School of Medicine, Los Angeles, CA; 6) Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Columbia University Medical Center, New York, NY; 7) Department of Genetics, Division of Biomedical Genetics, University Medical Center Utrecht, 3508 AB Utrecht, The Netherlands.

Structural variation (SV) dynamically shapes the content and organization of an individual genome. A fraction of the SV landscape is non-canonical, or complex, involving multiple SV signatures compounded into a single variant. These complex SV (cxSV) range in intricacy from CNV-flanked inversions to extraordinarily complex rearrangements originating from chromosome shattering and random reassembly. The latter phenomenon, referred to as chromothripsis, was first reported in cancer but has also been observed in the viable human germline, albeit frequently co-occurring with disease phenotypes. We have performed long-insert whole-genome sequencing (liWGS) on >1000 human genomes from both control individuals and those with neurodevelopmental disorders or congenital anomalies. Throughout those studies, we have observed a diverse landscape of cxSV in the human germline, including extreme forms of largely balanced chromothripsis. To date, we have characterized 14 distinct recurrent subclasses of cxSV that are relatively large (>5 kb), demonstrating that: 1) >85% of all large cxSV involved at least one inverted segment; 2) collectively, cxSV were larger than canonical SV; and 3) complex duplications were over 2X larger but 3X less common than complex deletions. Importantly, the vast majority of cxSV were stably inherited from unaffected parents, and 22.1% of all cxSV appeared in >5% of subjects, evincing that cxSV is a standing class of variation in all human genomes. In three cases we have resolved extreme forms of chromothripsis involving up to 65 breakpoints across four chromosomes and disrupting up to 29 genes in a single individual. While two of these rearrangements were relatively balanced, the third one reported in a subject with developmental delay involved substantial genomic imbalance (>19Mb), a hallmark of chromothripsis in cancer but not previously observed in the germline. We further revealed this rearrangement to be mosaic (mosaicism of 55-58% by liWGS, ~60% by microarray), further suggesting that massively unbalanced chromothriptic rearrangements are unlikely to be tolerated in the human germline. Collectively, these analyses revealed an unprecedented diversity of large and relatively frequent cxSV in the human germline, and extreme examples of chromothripsis. These data suggest that such cxSV are tolerated in the human germline and are preferentially accessible to methods providing long-range information on genome structure.

3388W

High confidence phasing of rare variants over long distances using RNA-sequencing and phASER improves medical and functional genomic studies. *S.E. Castel[1,2], P. Mohammadi[1,2], W.K. Chung[3], Y. Shen[2,4], T. Lappalainen[1,2].* 1) New York Genome Center, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY; 3) Departments of Pediatrics and Medicine, Columbia University, New York, NY; 4) Department of Biomedical Informatics, Columbia University, New York, NY.

Haplotype phasing of genetic variants is important for clinical interpretation of the genome, population genetic analysis, and functional genomic analysis of allelic activity. Here we present phASER, a read-backed approach for phasing variants using RNA-seq reads, which often span multiple exons due to splicing, allowing for phasing of rare and *de novo* variants over long distances. Using data from both population-scale functional genomic and medical genetic studies we show that the inclusion of RNA-seq provides significant advantages over DNA-seq or population based phasing alone. We extensively benchmarked the characteristics of phASER read-backed phasing, including total variants phased, accuracy, and distance using GTEx data from across sequencing assays and tissue types, and found that up to 15% of rare coding variants could be phased using RNA-seq from a single tissue, at distances up to hundreds of kilobases. The incorporation of RNA-seq increased the number of rare coding variants phased by 1.5x for whole exome sequencing (WES) and 2.4x for whole genome sequencing (WGS). Next, we investigated the advantages of read-backed phasing with several types of sequencing reads in resolving potential compound heterozygosity for protein-disrupting variants. In 1000 Genomes individuals the inclusion of Geuvadis RNA-seq from LCLs increased power by 30%, and in a medical genetics study of congenital diaphragmatic hernia, RNA-seq in a tissue of disease relevance increased the number of expressed compound heterozygote alleles identified by 2.6x over WES alone. In addition to phasing, our method allows for the quantification of allelic expression at the haplotypic level as opposed to the variant level. When applied to Geuvadis data, phASER reduced false positive signals of allelic imbalance at 56% of genes tested, while uncovering false negatives at 7%. Finally, using GTEx v7 data, we combined read-back phased 30x WGS genotype calls with RNA-seq data from dozens of tissues to produce the largest panel of high confidence rare variant haplotypes to date. Using this panel we were able to uncover evidence of epistatic interaction between coding variants in the form of the accumulation of rare, putatively deleterious alleles on the same haplotype. Our work demonstrates that phasing using RNA-seq reads and phASER significantly improves haplotype-based analyses by incorporating a widely available data type that has become trivial to produce at population scale.

3389T

Building complete chromosomal haplotypes using single-cell sequencing. *D. Porubsky[1], A.D. Sanders[2], N. van Wietmarschen[1], E. Falconer[2], M. Hills[2], M.R. Bevova[1], D. Spierings[1], V. Guryev[1], P.M. Lansdorp[1,2,3].* 1) ERIBA, University Medical Center Groningen, Groningen, Netherlands; 2) Terry Fox Laboratory, BC Cancer Agency, Vancouver, Canada; 3) Division of Hematology, Department of Medicine, University of British Columbia, Vancouver, Canada.

Haplotypes are fundamental to fully characterize the diploid genome of an individual, yet phasing of genetic variants along the whole length of homologous chromosomes remains challenging. Here we introduce single cell DNA template strand sequencing (Strand-seq) as a novel approach able to phase diploid genomes along the entire length of all chromosomes. Strand-seq (Falconer et al., 2012) is a single cell sequencing technique with a unique ability to retain directionality of sequencing reads, based on the DNA template strand inheritance. This allows us to map every read to a single parental chromosome in a single cell. We demonstrate the power of haplotyping using Strand-Seq by phasing a HapMap individual (NA12878) at high accuracy (concordance 99.3%) without using generational information or statistical inference. With this approach, we mapped all meiotic recombination events in a family trio with high resolution (median range ~14kb). In addition our method allows phasing of larger structural variants like deletions, insertions as well as homologue specific estimate of copy number of segmental duplications. Moreover, balanced rearrangements like inversions which are difficult to detect with current technologies can be reliably phased using Strand-seq (Sanders et al., accepted Genome Research). To our knowledge this is the only sequencing technique able to map and phase heterozygous inversions. Lastly, the single cell resolution of our approach allowed us to observe loss of heterozygosity regions in small number of cells, a significant advantage for studies of heterogeneous cell populations, such as cancer cells. We conclude that Strand-seq is a unique and powerful approach to completely phase individual genomes and map inheritance patterns in families as well as detect haplotype differences in single cells. Falconer, E. et al. DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. Nat. Methods 9, 1107–12 (2012)Sanders, A.D., Hills, M., Porubsky, D., Guryev, V., Falconer, E., Lansdorp. P.M. Characterizing polymorphic inversions in human genomes by single cell sequencing. Accepted, Genome Research; submission number: GENOME2015/201160.

3390F

Impact of rare variants on gene expression. *X. Li[1], Y. Kim[2], J. Davis[3], E. Tsang[3], Z. Zappala[3], F. Damani[2], C. Chiang[4], . GTEx Consortium , I. Hall[4], A. Battle[2], S. Montgomery[1,3].* 1) Department of Pathology, Stanford University, Stanford, CA; 2) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 3) Department of Genetics, Stanford University, Stanford, CA; 4) Genome institute, Washington University, St. Louis, MO.

Due to less selective process, rare variants have generally larger effect sizes than common variants among human population. However their exact functional impact remains largely unexplored hurdled by the requirement of extra-large sample sizes and deep whole genome sequencing. GTEx human tissue RNA-seq samples paired with genome sequencing samples (over 5000 RNA-Seq samples of 44 tissues of 449 donors, 148 individuals also with sequenced genome) provide us with the unique opportunity to unveil a comprehensive landscape of the functional impact of rare variants. In this study we set out to discover expression outliers among different tissues and characterize the roles of many types of rare variants behind those effects. For each of the 44 tissues, we identify 400~1400 outliers in each tissue (at FDR 2~6%).There are significant sharing of outliers among tissues, ranging from 5 to 17% depending on closeness of tissues. Investigating all categories of variants: SNP, indels and structural variants behind those outliers, we estimate that ~50% of the under-expression and ~20% of the over-expression outliers can be explained by rare variants in the genic region (TSS -10kb to TES + 10kb). Severe impact is found at multiple loss-of-function variants, structural variants and indels. For non-coding variants, most striking effect is seen at promoters but no evidence of enrichment is observed at enhancers. Looking at finer scale allele frequencies in 4000 UK10k individuals, we found that rare variants at our outliers are at even lower frequencies than rare variants detected with similar annotation in non-outliers, evidencing the presence of selective pressure on their potentially deleterious effects. Outlier expression is further validated by widespread allele specific expression (ASE) at those genes. We observe much larger ASE effect sizes at outliers than non-outliers, which is further confirmed by excluding mapping biases and blacklisted low-mappability regions. Our outlier findings also have substantial intersections with known GWAS, OMIM, OrphaNet, cancer and cardio gold standard genes, indicates that outlier expression can have functional impacts for various rare and common diseases. In summary, our study builds a detailed catalog of all types of rare variants and their regulatory effects, which forms the basis for prioritizing rare variants for further phenotypic characterization.

3391W

Copy number variation in human embryonic stem cell lines. *R. Handsaker[1,2], F. Merkle[3,4], G. Genovese[1,2], S. Ghosh[1,2,5,6], S. Kashin[1,2], M. Pato[7], C. Pato[7], K. Eggan[2,5,6], S. McCarroll[1,2].* 1) Department of Genetics, Harvard Medical School, Boston, MA, USA; 2) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; 3) Metabolic Research Laboratories and Medical Research Council Metabolic Diseases Unit, Wellcome Trust - MRC Institute of Metabolic Sciences, Cambridge, UK; 4) Wellcome Trust – MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; 5) Harvard Stem Cell Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA; 6) The Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA; 7) SUNY Downstate Medical Center, Brooklyn, NY, USA.

Human stem cells are an important tool for biomedical research with many potential therapeutic and clinical applications. Previous studies have demonstrated that human stem cell lines can harbor significant genetic abnormalities and that these genetic variations can arise in culture, accumulate in the cell line, and undergo positive selection for growth advantage. Most studies of the genetic variation in human stem cells have used karyotyping or SNP arrays to characterize stem cell genomes, limiting the sizes of copy number variants that could be ascertained and the dynamic range of the copy numbers changes that could be inferred, or have investigated only a small number of cell lines. We performed deep (25x) whole genome sequencing on 130 human embryonic stem cell (hESC) lines, 99 lymphoblastoid cell lines (LCL) and 401 human DNA samples derived from whole blood. These data were used to analyze copy number variation, including both germline variation and variation that may have arisen in culture. We ascertained both fixed copy number mutations and mosaic copy number mutations that are present in a subset of the cultured cell population. We find that hESCs exhibit a high burden of large copy number mutations compared to LCLs, although the majority of the hESC lines surveyed are free of large copy number mutations. We identify several apparent hotspots of copy number mutation in hESCs, including further evidence for a previously reported hotspot on chromosome 20q. We find that certain mutational processes appear more than expected in hESCs, including the apparent formation of isochromosomes. These findings will inform both clinical and research applications that rely on the genomic stability and phenotypic readout from stem cell lines.

3392T

Analysis of the human subtelomeric regions to elucidate the real structure and pathogenesis of the related diseases. *Y. Kuroki[1], A. Toyoda[2], H. Noguchi[2], K. Yanagi[1], A. Fujiyama[2], T. Kaname[1].* 1) Genome Medicine, National Center for Child Health and Development, Setagaya-ku, Tokyo, Japan; 2) Center for Information Biology, and Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka, Japan.

Human subtelomeres have unique structure of the genome, which form the transitions between chromosome-specific sequence and arrays of telomere repetitive sequence on the vicinity of chromosomal ends. Such regions show a complex genomic structure. The regions consist of variable mosaics of multi-chromosomal blocks of the nucleotide sequences. The subtelomere-specific repeats are identified as duplicated structures among the most of human subtelomeres. Abnormalities (e.g. deletions) of the subtelomeric region are associated with human diseases such as intellectual impairment and malformation syndrome (known as subteleomere deletion syndrome). Cause of subtelomeric abnormalities and precise relationship between their variations and clinical phenotypes are not fully elucidated. It is known that the latest reference data of human genome (GRCh38/hg38) are the most complete genome assembly in mammalian. However, there are still euchromatic gaps in several subtelomeric regions, due to their complex genomic structures. In fact, gene contents, constitution of the DNA elements, and variation of the copy number of the duplicated sequences in the subtelomeric regions have not been fully analyzed. Precise analyses in the subtelomeric regions are important to understand the subtelomere deletion syndrome. To clarify the precise and exact structure of the human subtelomeres, we isolated Fosmid/Cosmid clones covering the human subtelomeric region and started to analyze their complete structures by NGS analysis. In addition, we also performed the comparative analysis of subtelomeric sequences, which derived from Chimpanzee, Gorilla and Orangutan, for understanding the genome diversity and the evolutionary aspects in Primates. We present an overview of our project and show the current status at the meeting.

3393F

Stochastic gene activity in whole blood highlights more complex tissue identity. *L. Fresard[1], S.B. Montgomery[1,2].* 1) Department of Pathology, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA.

In transcriptome analysis, the number and repertoire of genes that are expressed often defines a tissue's identity. However, it is also well known that such expression can vary between individuals. We sought to quantify this variability and its impact on our understanding of tissue specificity by taking advantage of the availability of a large number of blood transcriptomes from the Depression Genes and Networks cohort. We analyzed blood expression of 909 individuals in order to precisely characterize and categorize gene expression in this tissue. We observe that 8,834 of all annotated protein coding genes (GENCODE v19) are always expressed in blood (43%), with 4,452 never expressed (22%) and 5,063 variably expressed (25%) among individuals. This variable gene activity can have several sources, but are in majority caused by environmental perturbations, genetics or their interaction. Most of the variable genes have expression levels comparable to those of constitutively active blood genes but may not be canonically described as a blood gene because not detected in a sufficient number of samples. We compared our results with standard resources describing tissue specific expression and we show that some of our variably expressed genes are conventionally described as not-expressed in blood. When exploring genes that are described as specific to tissues other than blood, we also see that a part of them are variably expressed in blood. To better describe the variable genes expression in other tissues, we used GTEx data and studied the proportion that is expressed in most individuals in different tissues. As many of the variably expressed genes in blood are constitutively active in other tissues, large studies of blood provide the opportunity to detect additional eQTLs that may be conventionally thought of as tissue-specific. These results highlight the need to refine our definition of tissue specific expression. Further, this work allowed us to assess expression of conventionally non-blood, tissue specific genes that are detectable in blood with numerous samples demonstrating the opportunities of large transcriptome genetic studies in blood.

3394W

Biased and coherent expression patterns of human Y chromosome genes. *A.K. Godfrey[1,2], D.C. Page[1,2,3].* 1) Whitehead Institute, Cambridge, MA; 2) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; 3) Howard Hughes Medical Institute, Whitehead Institute, Cambridge, MA.

The gonadal functions of human Y chromosome genes have long been a focal point of Y-chromosome research, despite the existence of widely transcribed genes in its male-specific region (MSY). To gain a broad and unbiased view of how MSY genes are expressed throuhgout the body, we measured their expression levels in 33 human tissues collected from 158 adult men, using a publicly available RNA-Seq dataset. We deployed methods to accurately estimate the expression levels of genes present in the MSY's abundant repetitive regions, as these were systematically underestimated in previous analyses. MSY genes cluster into two classes based on their expression profiles. Genes of one class demonstrate a high degree of specificity for the testis, with all but one (*SRY*) expressed predominantly in testicular germ cells. Genes of the second class, which include Y-specific regulators of gene activity, exhibit wide but biased expression across the body. These widely expressed MSY genes have correlated expression levels across both tissues and individuals, suggesting that they might function in similar biological processes. They are expressed similarly to their homologs on the X chromosome, yet they show relatively higher expression in the prostate and key endocrine glands, and more depleted expression throughout the brain. The biased and coherent expression patterns of human MSY genes motivate an expanded consideration of the Y chromosome's impact beyond the gonad and in differences between the sexes.

3395T

Genome assembly and full-length transcript profiling of stem cells from great apes. *J.G. Underwood[1,2], M.L. Dougherty[2], E. Tseng[1], O. Penn[2], C.M. Hill[2], D. Gordon[2,3], J. Huddleston[2,3], M.J.P. Chaisson[2], Z.N. Kronenberg[2], K.M. Munson[2], M. Malig[2], A. Raja[2,3], I. Fiddes[4], L.W. Hillier[5], C. Dunn[1], C. Baker[2], J. Armstrong[4], M. Diekhans[4], B. Paten[4], J. Shendure[2,3], R.K. Wilson[5], D. Haussler[4], C.S. Chin[1], A.M. Denli[6], F.H. Gage[6], E.E. Eichler[6].* 1) Pacific Biosciences of California, Menlo Park, CA; 2) Dept. of Genome Sciences, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, University of Washington, Seattle, WA; 4) Genomics Institute, University of California Santa Cruz and Howard Hughes Medical Institute, Santa Cruz, CA; 5) McDonnell Genome Institute, Department of Medicine, Department of Genetics, Washington University School of Medicine, St. Louis, MO; 6) Laboratory of Genetics, Salk Institute for Biological Studies, La Jolla, CA.

Single-molecule, real-time (SMRT) sequencing technology has shown considerable promise in improving human genome assemblies as well as our understanding of genetic variation. Existing great ape genome assemblies contain hundreds of thousands of gaps, concealing genetic differences that separate the great ape species. We previously presented an accurate assembly of the Western lowland gorilla and now we add to that the common chimpanzee genome generated from SMRT sequence data and application of a string-graph *de novo* assembly algorithm (Falcon). The new reference genomes of gorilla and chimpanzee represent a more than 200-fold improvement in contiguity with contig N50s of 9.6 Mbp and 12.7 Mbp, respectively. Incomplete gene annotation of non-human primate genomes is a major impediment to understanding functional differences between humans and closely related primate species, especially during early development. Transcriptome profiling studies of human embryonic stem (ES) cells and induced pluripotent stem (iPS) cells are advantageous because a large number of protein-coding genes are expressed prior to lineage specification and differentiation. Moreover, large genes, genes embedded within duplications, and genes associated with repeats such as LINE and HERVH-associated pluripotency transcripts are difficult to study by short read expression profiling and assembly. To improve annotation and gain fundamental insights into pluripotency differences between humans and other primates, we generated PacBio full-length cDNA data to reflect the polyadenylated RNAs expressed in iPS cells from human, chimpanzee, and gorilla along with rhesus macaque as an outgroup. From each species, we are generating ~ 100,000 transcripts based on long-read sequencing of size-fractionated cDNA from 1 to 8kb. In addition to providing a higher quality annotation of transcriptomes for the great ape genomes, the data are providing novel insights into alternative promoters, non-coding RNA and fixed differences with respect to RNA processing, and novel ORF annotations between humans and great apes. With full-length transcript and gene models in hand, we are reassessing short read RNAseq data in order to quantify the abundance and expression patterns of these new gene models more accurately in other tissues. The full-length transcript data will provide a more complete understanding of genic differences between humans and our nearest phylogenetic neighbors.

3396F

Repeat elements expression profile across different tissues in GTEx samples. *H. Yang[1], S. Mangul[1,2], N. Zaitlen[3], S. Shifman[4].* 1) Department of Comptuer Science, University of California Los Angeles, Los Angeles, CA; 2) Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles,CA; 3) Department of Medicine, University of California, San Francisco, CA; 4) Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jeruslam, Jerusalem, Israel.

Next Generation RNA Sequencing technology enabled researchers to investigate transcriptomics, which is directly correlated to protein expressions, as well as other activities within cells. However, some of the transcribed RNAs are not translated into proteins but rather reincorporated into the genome through a process called retrotranscription, and those sequences that are re-incorporated into the genome through this process is called retrotransposons. Retrotransposons occupy a significant portion of the human genome and are believed to have significant roles in certain diseases, such as Rett's disease or Schizophrenia. We investigated the differential expression of non-long terminal repeat retrotransposons: short and long interspersed elements, denoted as SINE and LINE, respectively, as well as Alu elements. The differential expression of non-LTR retrotransposons is investigated on 3000 RNA-Seq samples across 32 tissues types in Genotype-Tissue Expression (GTEx) dataset. The repeat profile was extracted by mapping reads to the reference of repeat elements on human genome. On average, 7.47% of total RNA-Seq reads are mapped to repeat elements. Of the reads mapped to repeat elements, 32% of repeat element reads are mapped to LINE, while 41% of repeat element reads are mapped to SINE. Of the tissue types, testes showed significantly higher expression of SINE-VNTR-*Alu* (SVA) subclass F compared with the mean expression of SVA-F throughout the samples (p = 2.46 * 10^{-33}), suggesting that there are important biological mechanism that upregulates the SVA element expression. Furthermore, high co-expression of *Alu* elements and LINE L1 elements is detected across tissues (R^2 = 0.7615), signifying that *Alu* expression may depend on the expression of L1 elements.

3397W

Identification and characterization of structural sequence variants in French Canadians. *K.S. Lo[1], C. Low-Kam[1], D. Rhainds[1], M.-P. Dubé[1,2], J.-C. Tardif[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Department of Medicine, Université de Montréal, Montreal, Quebec, Canada.

Structural DNA sequence variants (SV) include deletions, duplications, copy-number variants (CNV), insertions, inversions and translocations that affect 1 kb–3 Mb of DNA sequence. Recent discoveries have highlighted the importance of accurately genotyping complex SV in order to test their role in human diseases. In this project, we applied the Genome STRiP 2.0 (GS2) software (Handsaker et *al*., Nature Genet., 2015) to detect SV in 1,970 French Canadians from Quebec using data from low-pass whole-genome DNA sequencing (~5.6X coverage/participant). In comparison with other available tools, GS2 integrates information from multiple samples to increase confidence in SV genotyping calls.GS2 is computationally intensive, requiring the use of high-performance computing resources (>625 core-days for our dataset). After initial filtering of poor quality variants, 22,001 distinct autosomal CNV were identified, with an average of 701 CNV per individual and a mean minor allele(s) frequency (MAF) of ~1.2%. To assess the quality of the CNV identified, we compared our results with CNV from the 1000 Genomes Project (1000GP) phase 3 dataset identified with GS2. Approximately 60% of the CNV identified in 1000GP are also found in the French-Canadian dataset. About 39% of the CNV identified (mean length of 15kb) overlap with a gene, including 1,236 genes that are completely covered by a CNV. Of specific clinical interest, we found 165 genes annotated in the Online Inheritance in Man (OMIM) database that are completely encompassed within a CNV in French Canadians.To begin the exploration of the role of SV in the French-Canadian population, we tested the association between CNV genotypes and LDL-cholesterol (LDL-C) levels using an additive model in a linear regression framework. Interestingly, the most significant CNV is a multi-allelic variant with MAF=0.9% that overlaps with the first exon of the LDL-C receptor gene *LDLR* (*P*=$2.4x10^{-7}$). This CNV is absent from the 1000GP dataset, and corresponds to the previously described French-Canadian *LDLR* founder mutation identified in patients with familial hypercholesterolemia (Hobbs et *al*., NEJM, 1987). The CNV explained ~1% of the variation in LDL-C and conditional analyses indicate, as expected, that it is independent from the known *LDLR* SNPs previously identified by GWAS. Although association tests with other heart disease risk factors are ongoing, this example already illustrates the importance to assess SV in populations of interest.